



# Cisco Nexus 3548 Switch Architecture

White Paper

September 2012

---

# Contents

<a href="#">What You Will Learn</a> .....	3
<a href="#">Key Features of the Cisco Nexus 3548</a> .....	3
<a href="#">Cisco Nexus 3548 Switch-on-a-Chip Data-Plane Architecture</a> .....	5
<a href="#">Cisco Nexus 3548 Switch-on-a-Chip Packet Forwarding</a> .....	6
<a href="#">Cisco Nexus 3548 Warp Mode</a> .....	7
<a href="#">Cisco Nexus 3548 Warp SPAN</a> .....	7
<a href="#">Cisco Nexus 3548 Buffering</a> .....	7
<a href="#">Cisco Nexus 3548 Quality of Service</a> .....	9
<a href="#">Traffic Classification</a> .....	9
<a href="#">CoS and DSCP Marking</a> .....	10
<a href="#">MTU</a> .....	10
<a href="#">Queuing and Scheduling</a> .....	10
<a href="#">Congestion Management and DCTCP</a> .....	10
<a href="#">Link Bandwidth Management</a> .....	10
<a href="#">Control-Plane Policing</a> .....	11
<a href="#">Cisco Nexus 3548 Hardware Architecture</a> .....	11
<a href="#">Cisco Nexus 3548 Physical Specifications</a> .....	11
<a href="#">Conclusion</a> .....	12
<a href="#">For More Information</a> .....	12

---

## What You Will Learn

The Cisco Nexus<sup>®</sup> 3548 Switch (Figure 1) is a high-performance, high-density, ultra-low-latency Ethernet switch that is a new member of the Cisco Nexus 3000 Series Switches. This compact one-rack-unit (1RU) 1, 10 and 40 Gigabit Ethernet switch provides line-rate Layer 2 and 3 switching. It runs the industry-leading Cisco<sup>®</sup> NX-OS Software operating system, providing customers with features and capabilities that are widely deployed throughout the world. The Cisco Nexus 3548 is optimized for financial co-location deployments that require robust unicast and multicast routing protocol features with ultra-low latency as low as 190 nanoseconds (ns). This document provides an overview of the features and benefits of the Cisco Nexus 3548 and a detailed description of the switch's internal architecture. For software feature availability regarding the functionalities described in this document, please refer to the Release Notes documentation.

**Figure 1.** Cisco Nexus 3548 Switch



## Key Features of the Cisco Nexus 3548

The Cisco Nexus 3548 is the first member of the second generation of Cisco 3000 Series high-performance, high-density, ultra-low-latency Ethernet switches. It provides 48 ports of 10 Gigabit Ethernet ports. The switch offers the following high-value features:

- Nonblocking line-rate performance: All the Ethernet ports in the Cisco Nexus 3548 can handle packet flows at wire speed. All 48 Ethernet ports can be transmitting and receiving packets simultaneously at 10 Gbps without any effect on performance, offering 960 Gbps of bidirectional bandwidth, with more than 714.24 million packets per second (mpps). The Cisco Nexus 3548 can provide true 40 Gbps by bundling series of 4 10 Gbps ports together, providing a 40 Gbps MAC.
- Algo Boost technology: The cut-through switching technology used in the switch-on-chip (SoC) architecture of the Cisco Nexus 3548 enables the product to offer ultra-low latency. The ultra-low latency of the Cisco Nexus 3548 together with a dedicated buffer for each port and a dynamic shared buffer make the Cisco Nexus 3548 the best choice for ultra-latency-sensitive environments.

Algo Boost technology is a set of hardware functions that provide an ultra-low latency as low as 190 nanoseconds (ns), even with features such as Network Address Translation (NAT) in hardware. Algo Boost also provides real-time buffer visibility and monitoring with extensive statistics and histograms. It also provides intelligent traffic mirroring capabilities, with a hardware capability of eight sessions and sampled, filtered, truncated, and Precision Time Protocol (PTP) time-stamped mirrored traffic features. Finally, Algo Boost also provides a Pulse-Per-Second (PPS) port for PTP accuracy.

- NAT: The Cisco Nexus 3548 supports NAT in hardware, with no impact on latency. Three types of NAT are supported:
  - Static Network Address Translation (SNAT) provides the option to configure static mapping between local and global IP addresses and User Datagram Protocol (UDP) and Transmission Control Protocol (TCP) ports.

- Dynamic Network Address Translation (DNAT) allows the user to establish 1 to 1 dynamic mappings between the local and global addresses by describing the local addresses to be translated and the pool of addresses from which to allocate global addresses and then associating the two.
- PAT provides a many to one option to map multiple IPv4 addresses to a fewer number of IPv4 addresses, using different TCP/UDP port numbers.
- Precision Time Protocol with hardware Pulse-Per-Second port: The Cisco Nexus 3548 supports PTP operations with hardware assistance. The forwarding application-specific integrated circuit (ASIC) time-stamps the PTP packets in both the ingress and the egress direction in hardware.

With PTP, the IEEE1588 packet is time-stamped at the ASIC ingress point to record the event message arrival time in hardware at the parser level. The time stamp points to the first bit of the packet (following the start frame delimiter [SFD]). Next, the packet is copied to the CPU with the time stamp and destination port number. After this, the packet traverses the PTP stack. The advanced PTP clock algorithm in the Cisco Nexus 3548 keeps track of all the timing and frequency information and makes necessary adjustments to help ensure accurate time. Finally, the packet is sent out at the egress port and internally marked as a high-priority packet to help ensure priority egress out the switch. The corresponding time stamp for the transmitted packet is available from the First In, First Out (FIFO) transmission time stamp.

- Shared buffer architecture: The Cisco Nexus 3548 has 18 MB of buffer space, including per-port and dynamically allocated shared buffers. A given port can use up to 5 MB of buffer space.
- Separate egress queues for unicast and multicast traffic: The Cisco Nexus 3548 increases the number of egress groups available by supporting eight queues: unicast or multicast, configurable in eight quality-of-service (QoS) groups. Two of these queues are reserved for system control-plane traffic.
- Data Center TCP (DCTCP) is an extension of Explicit Congestion Notification (ECN) and provides further support than Explicit Congestion Notification (ECN) marking. DCTCP provides the ability mechanism to gauge the extent of the congestion. ECN itself is an extension of TCP/IP defined in RFC 3168 that enables end-to-end notification of network congestion before a packet drop occurs, unlike the behavior of traditional TCP, which detects network congestion after a packet drop. When congestion is detected, the TCP sender takes action by controlling the flow of traffic. Dropped packets can sometimes lead to long TCP timeouts and consequent loss of throughput. Instead of dropping a packet, the Cisco Nexus 3548 uses DCTCP to signal impending congestion to the sender, who can then slow the rate of traffic before a packet drop occurs.
- Robust Layer 3 mode: The Cisco Nexus 3548 can operate in Layer 3 mode without any additional hardware. It has a comprehensive Layer 3 feature set that includes full Border Gateway Protocol (BGP) support. Additional features such as Graceful Restart in Open Shortest Path First (OSPF), Enhanced Interior Gateway Routing Protocol (EIGRP), and BGP provide a high-performance Layer 3 platform. For more information, please refer to the Cisco Nexus 3000 Series data sheet.
- Multicast: Protocol-Independent Multicast sparse mode (PIM-SM), PIM source-specific multicast (PIM-SSM), Bidirectional PIM (PIM-Bidir) and Multicast Source Discovery Protocol (MSDP) multicast protocols are supported. The switch forwards multicast traffic at line rate on all 48 ports. Multicast packets are replicated in the hardware without CPU intervention. When a Layer 2 lookup succeeds, the packet is forwarded based on the destination MAC address in the Layer 2 multicast table. Next, if the lookup results point to an entry in the Layer 3 IP multicast table, the packet is replicated at the egress ports and VLANs referred to by the table.

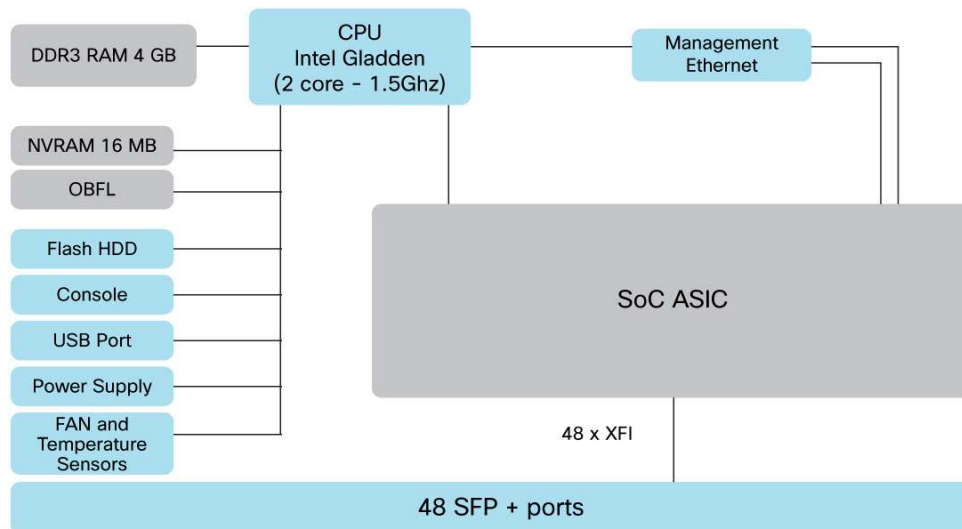
When the switch receives an IP multicast packet with a group address not yet learned by the switch, the link-local unknown multicast packets are flooded in the source VLAN. By default, the Cisco Nexus 3548 distributes (S,G) PIM joins among ECMP paths.

- Access control list (ACL) capabilities: The Cisco Nexus 3548 hardware supports a broad range of ACL fundamental and advanced features. The fundamental ACL features supported include router ACLs (RACLs), VLAN ACLs (VACLs), and port ACLs (PACLs). Advanced ACL features such as Policy-Based Routing (PBR), NAT, Address Resolution Protocol (ARP) inspection, and Dynamic Host Configuration Protocol (DHCP) snooping are also supported by the hardware, with no increase in latency. The hardware supports a total of 4096 ACL ternary content addressable memory (TCAM) entries, which can be applied to the ingress and egress on a physical port. The SoC ASIC also supports the logical operation unit (LOU) mechanism, with 64 LOUs, allowing label sharing among ACL entries and resource optimization. Statistics and debug mechanisms provide extensive information about the runtime ACL configuration used.
- Advanced SPAN capabilities: On many platforms in actual high-performance environments in which ultra-low latency is essential to provide near-real-time information, SPAN creates additional latency since the network device needs to duplicate the packets. The Cisco Nexus 3548 Series is equipped with a high-performance implementation of SPAN that enables traffic replication without additional latency for the destination SPAN port, while capable to perform line rate SPAN. Furthermore, it also provides the capability to configure up to 8 bidirectional sessions which can be sampled, filtered, truncated, or even PTP time-stamped to provide precision time information to remote capture utility tracking the real time the packets had arrived to the Cisco Nexus 3548 switch.

### Cisco Nexus 3548 Switch-on-a-Chip Data-Plane Architecture

The Cisco Nexus 3548 has a unique SoC design. The Cisco SoC has many features built into the hardware, including Algo Boost technology and a unique optimization of the logic for ultra-low-latency performance regardless of the features enabled. It is a single-ASIC architecture that connects to all the Ethernet ports (Figure 2).

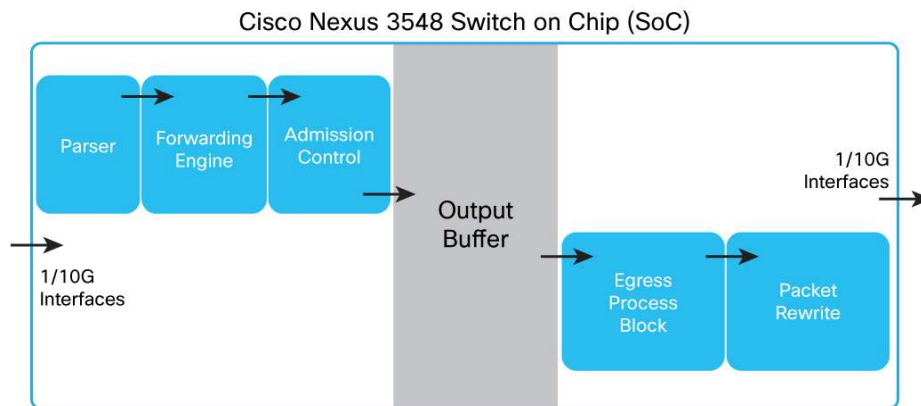
**Figure 2.** Cisco Nexus 3548 Data Plane and Switch-on-a-Chip Architecture



### Cisco Nexus 3548 Switch-on-a-Chip Packet Forwarding

The ingress flow component is responsible for most of the switch features, such as VLAN assignment and Layer 2 and 3 table lookups, buffer allocation, and time stamping. Essentially, it the ingress flow component is responsible for the packet-forwarding decisions (Figure 3). The buffering details are discussed later in this document.

**Figure 3.** Cisco Nexus 3500 Switch-on-Chip Packet Flow



The parser engine parses the incoming packets and extracts the fields required for decisions and passes the information to the Layer 2 and 3 lookup processes. PTP time stamping occurs at this stage.

Next, the packet is sent to the forwarding engine for the learning phase. The source MAC address is learned in the hardware for the given VLAN. Then, depending on the destination MAC address lookup result, the packet is forwarded to the Layer 3 processing engine or the CPU or flooded to all members of a particular VLAN. For Layer 3, the packet arrives at the Layer 3 processing engine, and the source IP address is looked up in the Layer 3 table. The destination IP address is looked up and indexed in the next-hop table, which lists the outgoing interface and the destination MAC address. The outgoing interface provides an index in the Layer 3 interface table that supplies the source MAC address and the VLAN. The other lookup capabilities such as ACL and NAT are also performed at this level. When the packet leaves the queuing engine, this information is used to rewrite the packet after it has been parsed; then the packet is forwarded out of the egress interface. The learning process takes full advantage of the SoC architecture. All the engine operations are performed in the hardware without CPU load.

The packet then proceeds to the admission control check process. The queuing, replication and DCTCP process occur in the admission check component. Depending on the amount of shared buffer space available, the packet will be stored in the reserved per-port location, the reserved per-queue location, or the dynamic shared space. All these spaces are part of the shared buffer. Then the packets are sent to the queue for scheduling. For packet replication, the decision is made in the output buffer. The replication occurs in the queuing engine, as the packets are being placed in queues for scheduling. The replication process does not replicate the data packet. Instead, the data packet is kept with a pointer in memory, and during the packet rewrite process, a different header is created on each outgoing interface. The result is similar latency times across all the outgoing interfaces sending the same data.

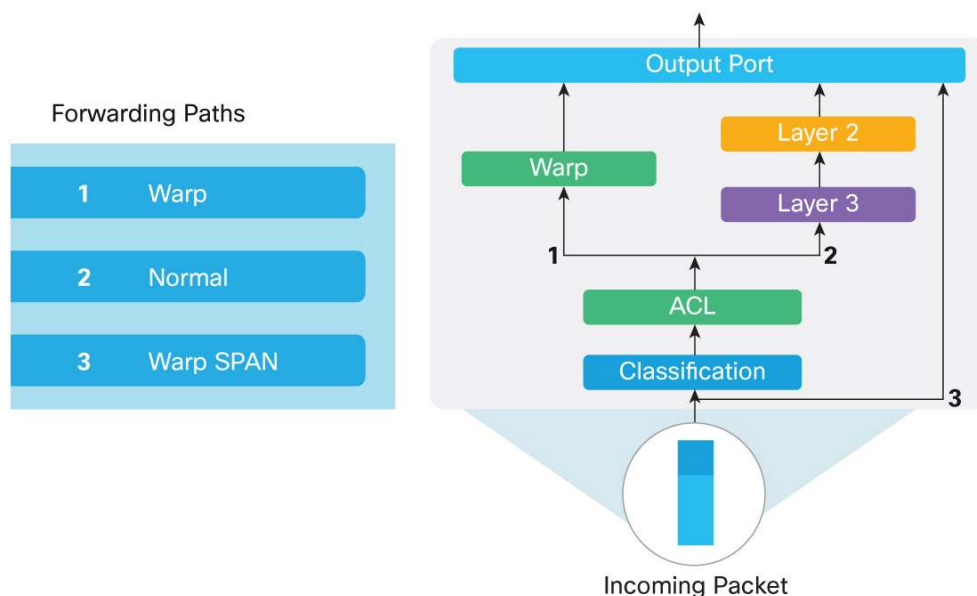
## Cisco Nexus 3548 Warp Mode

The Cisco Nexus 3548 introduces an innovative forwarding mechanism, called warp mode. In warp mode, the access path is shortened through the use of a single memory allocation block, resulting in faster processing of frames and packets. The benefit of warp mode is better performance; latency is reduced by up to 20 percent with latency as low as 190ns. Figure 4 shows the data path difference between normal mode and warp mode in the SoC architecture.

## Cisco Nexus 3548 Warp SPAN

The Cisco Nexus 3548 introduces an innovative SPAN solution, called Warp SPAN. In Warp SPAN, the latency is reduced to the strictest minimum. Traffic arriving at one dedicated ingress port is replicated to a user configurable group of egress ports. The packet replication happens without any filters or lookup mechanisms (Figure 4). Unlike normal SPAN, the incoming traffic is replicated before any Classification or ACL occurs. The latency for the replicated packets is as low as 50ns. Warp SPAN function is to provide a bypass splitting operation independent and simultaneous of normal traffic forwarding. For example, the incoming source traffic can be switched, routed, multicast replicated etc. while at the same time this incoming traffic is warp spanned to multiple destination ports.

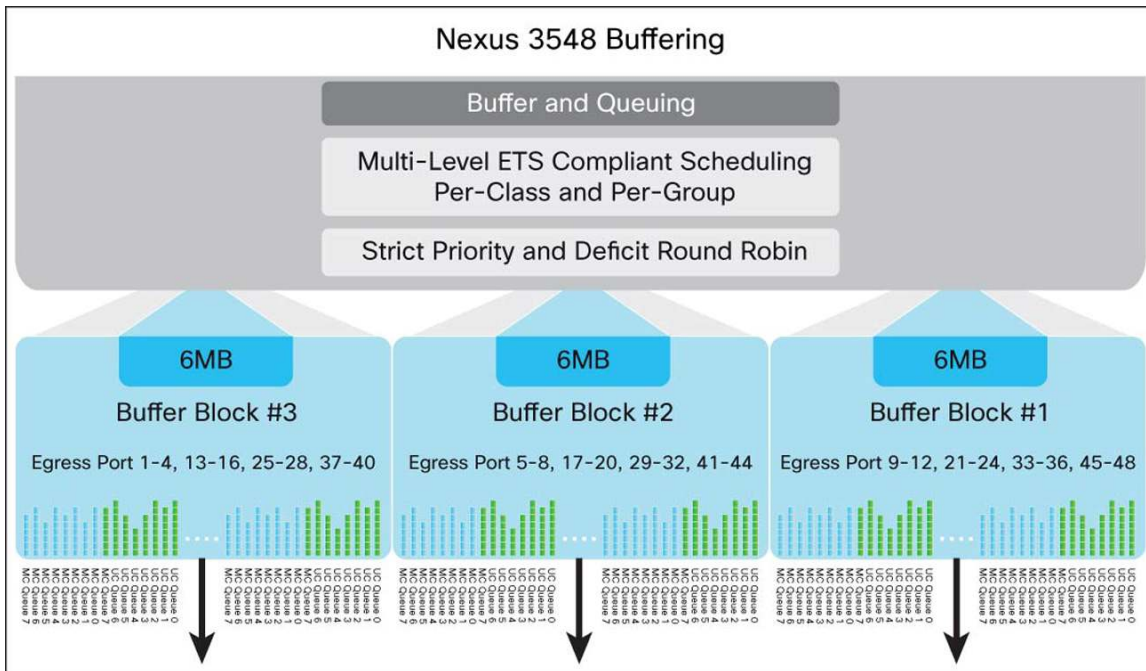
**Figure 4.** Warp Mode Hardware Data Path



## Cisco Nexus 3548 Buffering

Each Cisco Nexus 3548 interface is supplied with a buffer block that resides in the queuing engine (Figure 5). The buffer block has a per-port dedicated buffer and a dynamic shared buffer pool, with a total of 18 MB of buffer space for the platform. There are three buffer pools of 6 MB each, and each group of 16 ports share a buffer pool. For optimal performance, a percentage of the buffer space is statically reserved for each port. The remaining buffer space is shared among all the ports and can be used during congestion events. Eight hardware queues are available for use for unicast or multicast, with multilevel scheduling per port and per group using a Deficit Weighted Round Robin (DWRR) mechanism. The buffer partition is shared, and the buffer limits are configurable on a per-port and per-queue basis.

**Figure 5.** Cisco Nexus 3548 Buffer Block

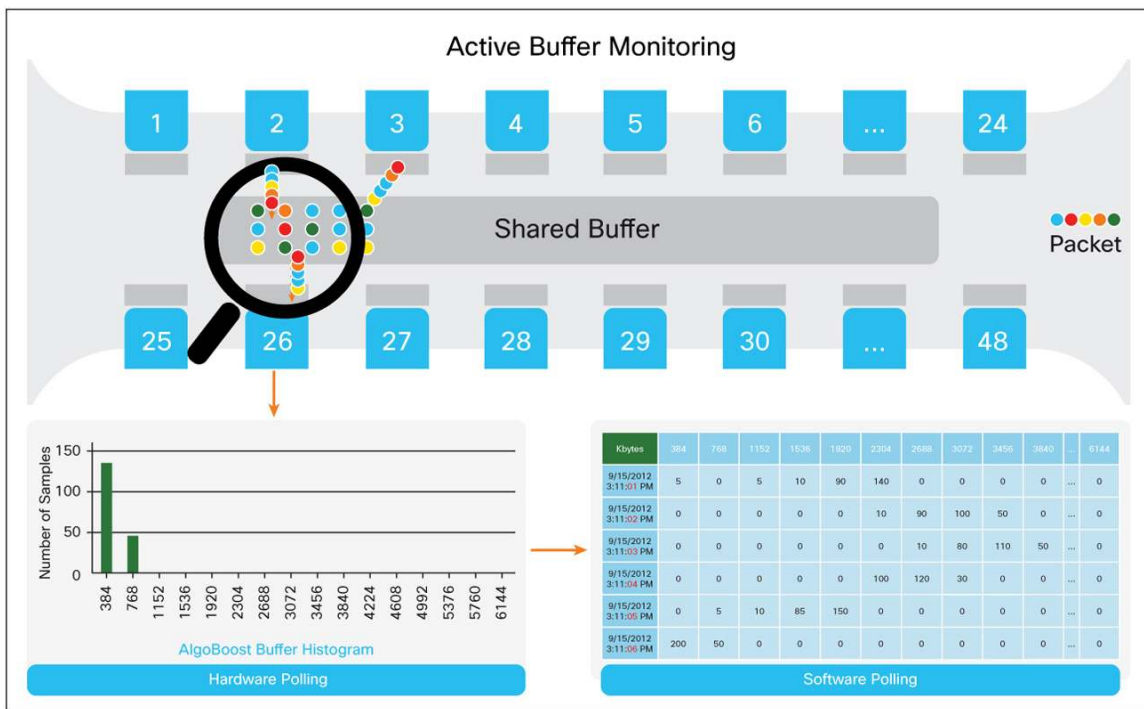


The Cisco Nexus 3548 provides real-time hardware buffer monitoring. It uses a cut-through switching mode that provides line-rate throughput without the need to use packet-buffering space under non-congested conditions. This architecture allows the switch to detect delays from microbursts and other queuing delays that occur in traffic flows when packets are buffered. During moments of congestion, the latency will increase due to queuing. Detecting the presence of any congestion at the hardware level provides very useful visibility into the switch. Awareness of the location, depth, and duration of the congestion allows additional optimizations to occur.

Active buffer monitoring provides a solution the challenges of congestion. In contrast to other buffer monitoring technologies that provide utilization snapshots and watermarks, active buffer monitoring provides detailed buffer histograms for each port, showing the percentage of time that the switch buffers are empty, fully occupied, or anywhere in between, with millisecond specificity. Figure 6 provides a summary of the buffer monitoring capabilities.



**Figure 6.** Active Buffer Monitoring



## Cisco Nexus 3548 Quality of Service

The main Cisco Nexus 3548 QoS functions are:

- Traffic classification
- Marking: class of service (CoS) or differentiated services code point (DSCP)
- Maximum transmission unit (MTU) verification
- Congestion management and DCTCP
- Queuing and scheduling
- Link bandwidth management
- Control-plane policing (CoPP)

In the Cisco Nexus 3000 Series, all data path and QoS system resources can be configured by system class. The QoS configuration uses the Cisco NX-OS command-line interface (CLI) with three QoS types: QoS, network QoS, and queuing.

### Traffic Classification

The first step in QoS processing consists of classifying the incoming packet so that it can be associated with a system class. This classification information will be retained from ingress to egress and will be used for all QoS processing. The classification can be based on the CoS or DSCP bits of the incoming packet or on user-defined QoS ACLs that match the Layer 2, 3, and 4 information. Alternatively the Cisco Nexus 3000 Series allows the user to set a default CoS value for an interface; classification then is performed based on the marked IEEE 802.1p value.

---

## CoS and DSCP Marking

The Cisco Nexus 3548 can mark IEEE 802.1p (CoS) bits or DSCP bits in the IP header. This function can be applied either on ingress or egress and is performed by the forwarding controller block of the unified port controller (UPC).

## MTU

The Cisco Nexus 3000 Series allows multiple types of traffic - unicast and multicast high-performance traffic with ultra-low latency - transmitted through the same interfaces. Each traffic class can have a different MTU requirement in the range 1518 to 9216 bytes. Consequently, the MTU setting needs to be per system class, and not per interface, because multiple traffic classes share the same physical interface. When operating as Layer 2 switches, the Cisco Nexus 3000 Series supports per-system-class MTU values for Layer 2 switch ports and per-interface-MTU values for Layer 3 interfaces.

## Queuing and Scheduling

The Cisco Nexus 3548 uses an output queuing system architecture to meet high-performance and high-density requirements. The architecture implements egress queues. Buffering occurs at the end of the output admission process. By the time the packet reaches the end of the ingress admission process, the egress port information is already known and resolved. The buffering occurs based on the egress port, the internal priority (QoS group), and the traffic type (unicast or multicast). On egress, there are eight egress queues for unicast and eight egress queues for multicast. In the queuing process, the weight of unicast compared to multicast traffic can be configured, to provide more specific control of traffic-type priority when contention occurs between unicast and multicast traffic. Bandwidth percentage allocation is also set in the queuing configuration.

In addition, one queue can be defined as a strict-priority queue; when a priority queue is configured and congestion occurs, any traffic marked in this class will be processed before any other queued traffic is scheduled.

## Congestion Management and DCTCP

Each interface of the Cisco Nexus 3000 Series has eight QoS groups, with 16 queues - 8 egress unicast queues and 8 egress multicast queues - and bandwidth allocation is supported to provide detailed control per queue. Also, the weight of unicast and multicast traffic at the scheduler level can be configured to set a higher priority for one type of traffic compared to the other in the event of contention.

Additionally, DCTCP, which is an extension of ECN, enables ECN to gauge the extent of congestion. As discussed earlier, ECN is an extension to TCP/IP. TCP detects network congestion by observing dropped packets. When congestion is detected, the TCP sender takes action by controlling the flow of traffic. However, dropped packets can lead to long TCP timeouts and consequent loss of throughput. Instead of dropping a packet, the Cisco Nexus 3548 can set a mark in the IP header to signal impending congestion, and it also can apply a configurable threshold that considers the extent of the congestion. The receiver of the packet echoes the congestion indicator to the sender, which responds as though congestion had been indicated by packet drops.

## Link Bandwidth Management

The Cisco Nexus 3548 implements eight egress queues for each interface, with two egress queues corresponding to a system class for control-plane traffic. The six QoS groups share the same link bandwidth, and the user can set the desired bandwidth for each egress queue through the Cisco NX-OS CLI. The QoS group 5 is reserved for SPAN. DWRR is scheduled between the egress queues. In DWRR, each queue is assigned a scheduling weight, and the bandwidth of each port is shared according to the weight.

## Control-Plane Policing

CoPP is a mechanism to protect the control plane (CPU) of the switch. Control traffic using the two hardware queues dedicated for it (inbound high and inbound low) is classified and policed when it tries to access the CPU. Any disruption of traffic to the CPU can result in a significant network outage. For example, in a disruption caused by a denial-of-service (DoS) attack, the CPU may not be able to process genuine control packets (PIM, BGP, OSPF, Spanning Tree Protocol, etc.). With customizable CoPP, the end user can tune the number of packets per second the CPU accepts.

## Cisco Nexus 3548 Hardware Architecture

The Cisco Nexus 3548 control plane runs Cisco NX-OS Software on a dual-core 1.5-GHz Intel Gladden processor with 4 GB of DDR3 memory. The supervisor complex is connected to the data plane in-band through two internal ports running 1-Gbps Ethernet (1 Gbps for each direction), and the system is managed in-band or through the out-of-band 10/100/1000-Mbps management ports. Table 1 summarizes the control-plane specifications.

**Table 1.** Control-Plane Specifications

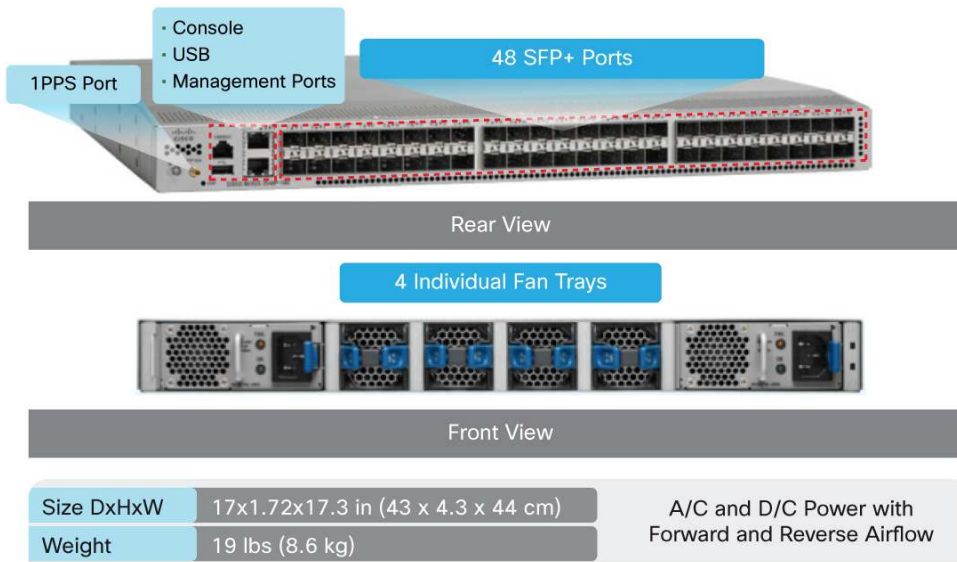
Component	Specification
<b>CPU</b>	1.5-GHz Intel Gladden processor (dual core)
<b>DRAM</b>	4 GB of ECC DDR3 Memory AT 1333 MHz in 4 DIMM slots
<b>Persistent Disk</b>	2 GB of embedded USB (eUSB) flash memory for base system storage
<b>NVRAM</b>	16 MB to store syslog, licensing information, and reset reason
<b>On-Board Fault Log</b>	512 MB of flash memory to store hardware-related fault and reset reasons
<b>Boot and BIOS Flash Memory</b>	64 MB to store upgradable and golden images
<b>Management Interface</b>	RS-232 console port and two 10/100/1000BASE-T management ports: mgmt0 and mgmt1; 1 external USB flash port can be used for updating the system configuration

## Cisco Nexus 3548 Physical Specifications

The Cisco Nexus 3548 is a high-density and high-availability 1RU 1 and 10 Gigabit Ethernet Layer 3 switch that is built to provide throughput with ultra-low latency. The 48 fixed 1 and 10 Gigabit Ethernet ports accept modules and cables meeting the Enhanced Small Form-Factor Pluggable (SFP+) form factor. The switch has a single serial console port and dual out-of-band 10/100/1000-Mbps Ethernet management ports. The Cisco Nexus 3548 is designed with 1+1 redundant and hot-swappable power supplies and four 3+1 redundant fan modules that can be accessed from the front panel, where status lights offer an at-a-glance view of the switch operation. The switch can function with one failed power supply and with one failed fan under any conditions. Cooling is front to back by default, supporting hot- and cold-aisle configurations that help increase cooling efficiency. Reverse cooling is also available. All ports are at the rear of the switch, simplifying cabling and reducing cable length. The front panel (Figure 7) includes status indicators for power supplies, their power entry connections, and cooling modules. All serviceable components are accessible from the front panel, allowing the switch to be serviced while in operation and without disturbing network cabling. The fans and power supplies are color coded to indicate airflow: the front-panel airflow is indicated with blue for fans and power supplies, and the back-panel airflow is indicated with red.

Please refer to the Cisco Nexus 3548 data sheet for information about supported optics.

**Figure 7.** Cisco Nexus 3548 Rear and Front Panels



## Conclusion

Cisco designed the Cisco Nexus 3548 Switch to extend the industry-leading versatility of the Cisco Nexus Family, providing a ultra-low-latency, 10/40 Gigabit Ethernet data center-class switch with a complete and robust Layer 3 built in. The ASIC provides an ultra-low-latency Warp mode while retaining a comprehensive Cisco NX-OS feature set. With the Algo Boost technology, industry innovations are embedded in the hardware. For example, Warp mode allows even more latency reduction in the packet processing. Real-time hardware buffer monitoring provides precise visibility into microbursts. Enhanced traffic mirroring provides a comprehensive monitoring capability, and NAT provides network flexibility. The hardware features are implemented with no additional latency.

## For More Information

- Cisco Nexus 3548 Switches: <http://www.cisco.com/go/nexus3548>
- Cisco NX-OS Software: <http://www.cisco.com/go/nxos>



Americas Headquarters  
Cisco Systems, Inc.  
San Jose, CA

Asia Pacific Headquarters  
Cisco Systems (USA) Pte. Ltd.  
Singapore

Europe Headquarters  
Cisco Systems International BV Amsterdam,  
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at [www.cisco.com/go/offices](http://www.cisco.com/go/offices).

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: [www.cisco.com/go/trademarks](http://www.cisco.com/go/trademarks). Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)