CONFIGURATION GUIDE

Extreme Network OS QoS and Traffic Management Configuration Guide, 7.1.0

Supporting Network OS 7.1.0

53-1004594-03
15 June 2018

Contents

Preface

Document conventions

The document conventions describe text formatting conventions, command syntax conventions, and important notice formats used in Extreme technical documentation.

Notes, cautions, and warnings

Notes, cautions, and warning statements may be used in this document. They are listed in the order of increasing severity of potential hazards.

NOTE
A Note provides a tip, guidance, or advice, emphasizes important information, or provides a reference to related information.

ATTENTION
An Attention statement indicates a stronger note, for example, to alert you when traffic might be interrupted or the device might reboot.

CAUTION
A Caution statement alerts you to situations that can be potentially hazardous to you or cause damage to hardware, firmware, software, or data.

DANGER
A Danger statement indicates conditions or situations that can be potentially lethal or extremely hazardous to you. Safety labels are also attached directly to products to warn of these conditions or situations.

Text formatting conventions

Text formatting conventions such as boldface, italic, or Courier font may be used to highlight specific words or phrases.

Format	Description
bold text	Identifies command names. Identifies keywords and operands. Identifies the names of GUI elements. Identifies text to enter in the GUI.
italic text	Identifies emphasis. Identifies variables. Identifies document titles.
`Courier font`	Identifies CLI output.

Format	Description
	Identifies command syntax examples.

Command syntax conventions

Bold and italic text identify command syntax components. Delimiters and operators define groupings of parameters and their logical relationships.

Convention	Description
bold text	Identifies command names, keywords, and command options.
italic text	Identifies a variable.
[ ]	Syntax components displayed within square brackets are optional. Default responses to system prompts are enclosed in square brackets.
{ **x** | **y** | **z** }	A choice of required parameters is enclosed in curly brackets separated by vertical bars. You must select one of the options.
x | **y**	A vertical bar separates mutually exclusive elements.
< >	Nonprinting characters, for example, passwords, are enclosed in angle brackets.
...	Repeat the previous element, for example, *member*[*member*...].
\	Indicates a “soft” line break in command examples. If a backslash separates two lines of a command input, enter the entire command at the prompt without the backslash.

Extreme resources

Visit the Extreme website to locate related documentation for your product and additional Extreme resources.

White papers, data sheets, and the most recent versions of Extreme software and hardware manuals are available at www.extremenetworks.com. Product documentation for all supported releases is available to registered users at www.extremenetworks.com/support/documentation.

Document feedback

Quality is our first concern at Extreme, and we have made every effort to ensure the accuracy and completeness of this document. However, if you find an error or an omission, or you think that a topic needs further development, we want to hear from you.

You can provide feedback in two ways:

- Use our short online feedback form at http://www.extremenetworks.com/documentation-feedback-pdf/
- Email us at internalinfodev@extremenetworks.com

Provide the publication title, part number, and as much detail as possible, including the topic heading and page number if applicable, as well as your suggestions for improvement.

Contacting Extreme Technical Support

As an Extreme customer, you can contact Extreme Technical Support using one of the following methods: 24x7 online or by telephone. OEM customers should contact their OEM/solution provider.

If you require assistance, contact Extreme Networks using one of the following methods:

- GTAC (Global Technical Assistance Center) for immediate support
  - Phone: 1-800-998-2408 (toll-free in U.S. and Canada) or +1 408-579-2826. For the support phone number in your country, visit: www.extremenetworks.com/support/contact.
  - Email: support@extremenetworks.com. To expedite your message, enter the product name or model number in the subject line.
- GTAC Knowledge - Get on-demand and tested resolutions from the GTAC Knowledgebase, or create a help case if you need more guidance.
- The Hub - A forum for Extreme customers to connect with one another, get questions answered, share ideas and feedback, and get problems solved. This community is monitored by Extreme Networks employees, but is not intended to replace specific guidance from GTAC.
- Support Portal - Manage cases, downloads, service contracts, product licensing, and training and certifications.

Before contacting Extreme Networks for technical support, have the following information ready:

- Your Extreme Networks service contract number and/or serial numbers for all involved Extreme Networks products
- A description of the failure
- A description of any action(s) already taken to resolve the problem
- A description of your network environment (such as layout, cable type, other relevant environmental information)
- Network load at the time of trouble (if known)
- The device history (for example, if you have returned the device before, or if this is a recurring problem)
- Any related RMA (Return Material Authorization) numbers

About This Document

Supported hardware and software

In those instances in which procedures or parts of procedures documented here apply to some devices but not to others, this guide identifies exactly which devices are supported and which are not.

Although many different software and hardware configurations are tested and supported by Extreme Networks, Inc. for Network OS, documenting all possible configurations and scenarios is beyond the scope of this document.

The following hardware platforms are supported by this release of Network OS:

- ExtremeSwitching VDX 2746
- ExtremeSwitching VDX 6740
  - ExtremeSwitching VDX 6740-48
  - ExtremeSwitching VDX 6740-64
- ExtremeSwitching VDX 6740T
  - ExtremeSwitching VDX 6740T-48
  - ExtremeSwitching VDX 6740T-64
  - ExtremeSwitching VDX 6740T-1G
- ExtremeSwitching VDX 6940-36Q
- ExtremeSwitching VDX 6940-144S
- ExtremeSwitching VDX 8770
  - ExtremeSwitching VDX 8770-4
  - ExtremeSwitching VDX 8770-8

To obtain information about a Network OS version other than this release, refer to the documentation specific to that version.

Using the Network OS CLI

For complete instructions and support for using the Extreme Network OS command line interface (CLI), refer to the *Extreme Network OS Command Reference*.

What's new in this document

This document describes the concepts and configuration of the traffic policing and QoS features for Network OS.

NOTE
On October 30, 2017, Extreme Networks, Inc. acquired the data center networking business from Brocade Communications Systems, Inc. This document has been updated to remove or replace references to Brocade Communications, Inc. with Extreme Networks., Inc., as appropriate.

QoS

QoS overview

Quality of Service (QoS) provides you with the capability to control how the traffic is moved from switch to switch. In a network that has different types of traffic with different needs (specified by Class of Service, or CoS), the goal of QoS is to provide each traffic type with a virtual pipe. FCoE uses traffic class mapping, scheduling, and flow control to provide quality of service.

Traffic running through the switches can be classified as either multicast traffic or unicast traffic. Multicast traffic has a single source but multiple destinations. Unicast traffic has a single source with a single destination. With all this traffic going through inbound and outbound ports, QoS can be set based on egress port and priority level of the CoS.

QoS can also be set on interfaces where the end-station knows how to mark traffic with QoS and it lies with the same trusted interfaces. An untrusted interface occurs when the end-station is untrusted and is at the administrative boundaries.

QoS features

The principal QoS features are as follows:

- Rewriting. Rewriting or marking a frame allows for overriding header fields such as the priority and VLAN ID. Refer to Rewriting on page 22 for more information.
- Queueing. Queueing provides temporary storage for frames while waiting for transmission. Queues are selected based on ingress ports, egress ports, and configured user priority level. Refer to Queueing on page 12 for more information.
- Congestion control. When queues begin filling up and all buffering is exhausted, frames are dropped. This has a detrimental effect on application throughput. Congestion control techniques are used to reduce the risk of queue overruns without adversely affecting network throughput. Congestion control features include IEEE 802.3x Ethernet Pause, Tail Drop, Ethernet Priority Flow Control (PFC), and Random Early Detect (RED). Refer to Congestion control on page 12 for more information.
- Multicast rate limiting. Many multicast applications cannot be adapted for congestion control techniques and the replication of frames by switching devices can exacerbate this problem. Multicast rate limiting controls frame replication to minimize the impact of multicast traffic. This feature is called BUM Storm Control on Extreme VDX 8770-4, VDX 8770-8, and later platforms.
- BUM storm control. A traffic storm occurs when packets flood the LAN, creating excessive traffic and degrading network performance. BUM storm control on page 55 allows you to limit the amount of broadcast, unknown unicast, and multicast (BUM) traffic admitted to the system to prevent disruptions on Layer 2 physical ports. All traffic received at a physical port in excess of a configured maximum rate for BUM traffic will be discarded. You can also specify whether to shut down an interface if the maximum rate has been exceeded within a 5-second sampling period and receive a LOG indication for the disabled interface. This feature is supported on Extreme VDX 8770, VDX 6740, VDX 6740T VDX 6940 series, and VDX 2746 platforms.
- Data Center Bridging. DCB describes an enhanced Ethernet that will enable convergence of various applications in data centers (LAN, SAN, and IPC) onto a single interconnect technology.

Queueing

Queue selection begins by mapping an incoming frame to a configured user priority, then each user-priority mapping is assigned to one of the switch's eight unicast traffic class queues or one of the eight multicast traffic class queues.

User-priority mapping

There are several ways an incoming frame can be mapped into a user-priority. If the neighboring devices are untrusted or unable to properly set QoS, then the interface is considered untrusted. All traffic must be user-priority mapped using explicit policies for the interface to be trusted; if it is not mapped in this way, the IEEE 802.1Q default-priority mapping is used. If an interface is trusted to have QoS set then the CoS header field can be interpreted.

> NOTE
> The user priority mapping discussed in this chapter applies to both unicast and multicast traffic.

Congestion control

Queues can begin filling up for a number of reasons, such as over-subscription of a link or backpressure from a downstream device. Sustained, large queue buildups generally indicate congestion in the network and can affect application performance through increased queuing delays and frame loss.

Congestion control covers features that define how the system responds when congestion occurs or active measures taken to prevent the network from entering a congested state.

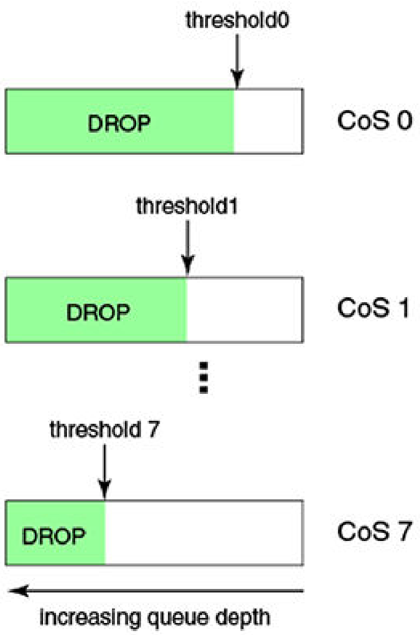
> NOTE
> You cannot configure CoS thresholds and multicast tail drop onExtreme VDX 8770-4 and VDX 8770-8 platforms. Weighted Random Early Detection (WRED) is supported only on Extreme VDX 6740, VDX 6940, VDX 8770-4, and VDX 8770-8 platforms.

Tail drop

Tail drop queuing is the most basic form of congestion control. Frames are queued in FIFO order and queue buildup can continue until all buffer memory is exhausted. This is the default behavior when no additional QoS has been configured.

The basic tail drop algorithm does not have any knowledge of multiple priorities and per traffic class drop thresholds can be associated with a queue to address this. When the queue depth breaches a threshold, then any frame arriving with the associated priority value will be dropped. The figure below illustrates how you can utilize this feature to ensure that lower-priority traffic cannot totally consume the full buffer memory.

FIGURE 1 Queue depth



Thresholds can also be used to bound the maximum queuing delay for each traffic class. Additionally, if the sum of the thresholds for a port is set below 100 percent of the buffer memory, then you can also ensure that a single port does not monopolize the entire shared memory pool allocated to the port. The tail drop algorithm can be extended to support per-priority drop thresholds. When the ingress port CoS queue depth breaches a threshold, then any frame arriving with the associated priority value will be dropped.

CoS thresholds

Every port has associated with it a total of 9 CoS thresholds, one for the port tail-drop threshold and the other eight are thresholds for per priority. To give a fair allocation of buffers for the traffic from all priorities, the port buffers are allocated among different priorities. That is achieved through per-priority tail-drop thresholds. The port tail-drop threshold represents the amount of buffers given to the port, and the per-priority tail-drop threshold (called the CoS tail-drop threshold from here on) represents the buffers allocated to each CoS.

> NOTE
> CoS thresholds apply only to multicast traffic.

Whenever the buffers allocated to a priority are fully exhausted, all the traffic coming in on that priority is dropped. In the absence of per-priority tail-drop thresholds (and with only port tail-drop thresholds), the buffers would be consumed on a first-come, first-served basis, resulting in an unfair share of buffers among all the priorities. If you know which priority traffic is most seen, then providing a sufficient number of buffers for those priorities results in fewer packet drops for those priorities.

Instead of using the standard priority values, you can assign anywhere from 0% through 100% priority to any threshold, as long as the sum of all eight priorities does not exceed 100%. For example, using the priorities 5 5 5 5 50 20 2 8 adds up to 100%, as shown in the following example:

```
switch(conf-if-te-0/1)# qos rcv-queue cos-threshold 5 5 5 5 50 20 2 8
switch(conf-if-te-0/1)# do show qos in te 0/1
Interface TenGigabitEthernet 0/1
CoS-to-Traffic Class map 'default'
               In-CoS: 0   1   2   3   4   5   6   7
 -----------------------------------------------------
 Out-CoS/TrafficClass: 0/1 1/0 2/2 3/3 4/4 5/5 6/6 7/7
Per-Traffic Class Tail Drop Threshold (bytes)
        TC: 0     1     2     3     4      5       6    7
 ------------------------------------------------------------------
 Threshold: 10180 10180 10180 10180 101808 40723  4072 16289
```

NOTE
Tail drop thresholds are not allowed to collectively exceed 100%, but the sum can be below 100%. For example, if the tail drop thresholds entered sum to less than 100%, then the buffer allocation is made on the basis of what has been configured.

Weighted Random Early Detection

NOTE
This feature is only supported on the Extreme VDX 8770 series, VDX 6740 series, VDX 6940 series, and VDX 2746 devices.

Traditionally, Weighted Random Early Detection (WRED) is used for TCP traffic streams, which are generally more aggressive, as well as reactive, to network drops. If WRED is not configured, queues build up at the switch and become full, resulting in tail drop. Tail drop situations can cause head-of-line blocking issues at the switch, which is not desirable. By configuring WRED, you set a probability for dropping packets before traffic in the queue reaches a specific threshold. This allows congestion to ease more gradually, avoids retransmit synchronization, resolves "bursty" TCP connections during congestion conditions, and controls packet latency.

Configure WRED using the following parameters:

- WRED profile identification (0-384)
- Minimum threshold of a queue (0-100%)
- Maximum threshold of a queue (0-100%)
- Drop probability (0-100%)

NOTE
Beginning with Network OS 5.0.0, the maximum number of WRED profiles at the system level is 3 on the Extreme VDX 8770 series, 16 on the Extreme VDX 6740 series and the VDX 2746, and 15 on the VDX 6940 series devices.

The ASIC driver maps the configured minimum and maximum percentages to the actual queue size in bytes, depending on the bandwidth of the port (buffers are allocated to a port according to port speed). When buffers in the queue build up to the set minimum threshold, packets being queued are randomly dropped. The drop probability parameter defines the randomness of the drops. When the queues exceed the minimum threshold, packets are dropped according to the configured drop probability value. When the queue buffers exceed the set maximum threshold, packets are dropped with 100% probability. The higher the probability set, the more likely packets will be dropped when the minimum percentage is reached.

You can also map a specific CoS priority value (0 through 7) to a specific WRED profile.

Configuring dynamic buffer sharing

If there is bursty, lossy traffic for certain flows in the system, you can borrow the buffers from less bursty flows, in order to reduce the traffic loss. The **qos** command is used to configure the egress or ingress queue limit, such as the maximum number of kilobytes of data that can be queued in the egress or ingress queue. The **tx-queue** keyword controls the egress, and the **rcv-queue** keyword controls the ingress.

This command only functions on the Extreme VDX 6740 series, VDX 6940 series, and VDX 2746 devices. This configuration is applied on individual RBridges.

To configure dynamic buffer sharing, perform the following steps from privileged EXEC mode.

1. Enter global configuration mode.

   ```
   device# configure terminal
   ```

2. Enter RBridge ID configuration mode.

   ```
   device(config)# rbridge-id 154
   ```

3. Configure the dynamic buffer sharing. The following example sets the egress limit to 256 kilobytes. (The default is 512 kilobytes.)

   ```
   device(config-rbridge-id-154)# qos tx-queue limit 256
   ```

   The following example sets the ingress limit to 1024 kilobytes. (The default is 285 kilobytes.)

   ```
   device(config-rbridge-id-154)# qos rcv-queue limit 1024
   ```

4. Return to privileged EXEC mode.

   ```
   device(config-rbridge-id-154)# end
   ```

Enabling drop logging

Use this procedure to enable RASlog messages for dropped data on an interface.

> NOTE
> The following drop types are logged:
> - Extreme VDX 2746, VDX 6740 series, and VDX 6940 series devices—Random Early Detect (RED) and tail drops
> - Extreme VDX 8770 device (internal port interface ASICs)—Tail drops only

1. Enter the **configure terminal** command to access global configuration mode.

   ```
   device# configure terminal
   ```

2. Enter the **interface** command to access interface subtype configuration mode for the interface on which you are implementing drop logging.

   ```
   device(config)# interface tengigabitethernet 2/2/1
   ```

3. Enter the **qos drop-monitor enable** command.

   ```
   device(conf-if-te-2/2/1)# qos drop-monitor enable
   ```

Ethernet Pause

Ethernet Pause is an IEEE 802.3 standard flow-control mechanism for back pressuring a neighboring device. Pause messages are sent by utilizing the optional MAC control sublayer. A PAUSE frame contains a 2-byte pause number, which states the length of the pause in units of 512 bits. When a device receives a PAUSE frame, it must stop sending any data on the interface for the specified length of time, once it completes the transmission of any frame in progress. You can use this feature to reduce Ethernet frame losses by using a standardized mechanism. However, the pause mechanism does not have the ability to selectively back-pressure data sources multiple hops away, or to exert any control per VLAN or per priority, so it is disruptive to all traffic on the link.

Ethernet Pause features

Ethernet Pause includes the following features:

- All configuration parameters can be specified independently per interface.
- Pause On/Off can be specified independently for TX and RX directions. No support is provided for disabling autonegotiation.
- Pause generation is based on input (receive) queueing. Queue levels are tracked per input port. When the instantaneous queue depth crosses the high-water mark, then a PAUSE frame is generated. If any additional frames are received and the queue length is still above the low-water mark, then additional PAUSE frames are generated. Once the queue length drops below the low-water mark, then the generation of PAUSE frames ceases.
- A PAUSE frame that is received and processed halts transmission of the output queues associated with the port for the duration specified in the PAUSE frame.

1-Gbps pause negotiation

When a 1-Gbps local port is already online, and the **qos flowcontrol** command is issued, the pause settings take effect immediately on that local port. However, when the link is toggled, pause is renegotiated. The local port will advertise the most recent **qos flowcontrol** settings. After autonegotiation completes, the local port pause settings may change, depending on the outcome of the pause negotiation, per 802.3 Clause 28B, as shown in the table below.

TABLE 1 Pause negotiation results

Advertised LOCAL cfg	Advertised REMOTE cfg	Negotiated result
Rx=off Tx=on	Rx=on Tx=on	asymmetrical: LOCAL Tx=on -> pause -> REMOTE Rx=on
Rx=on Tx=on	Rx=off Tx=on	asymmetrical: LOCAL Rx=on <- pause <- REMOTE Tx=on
Rx=on Tx=n/a	Rx=on Tx=n/a	symmetrical: LOCAL Tx/Rx=on <- pause -> REMOTE Tx/Rx=on
Rx=n/a Tx=n/a	Rx=off Tx=off	disable pause both sides

Ethernet Priority Flow Control

Ethernet Priority Flow Control (PFC) is a basic extension of Ethernet Pause. The Pause MAC control message is extended with eight 2-byte pause numbers and a bitmask to indicate which values are valid. Each pause number is interpreted identically to the base Pause protocol; however, each number is applied to the corresponding Ethernet priority/class level. For example, the Pause number 0 applies to priority zero, Pause number 1 applies to priority one, and so on. This addresses one shortcoming of the Ethernet Pause mechanism, which is disruptive to all traffic on the link. However, it still suffers from the other Ethernet Pause limitations.

NOTE
The Extreme VDX 6740 series and VDX 6940 series switches support a maximum of two PFC (lossless) CoS profiles on an interface. If any of these switches is present in a VCS cluster, the user should not create more than two lossless CoS profiles in the CEE configuration.

Ethernet Priority Flow Control includes the following features:

- Everything operates exactly as in Ethernet Pause described above, except there are eight high-water and low-water thresholds for each input port. This means queue levels are tracked per input port plus priority.
- Pause On/Off can be specified independently for TX and RX directions per priority.
- Pause time programmed into Ethernet MAC is a single value covering all priorities.

- Both ends of a link must be configured identically for Ethernet Pause or Ethernet Priority Flow Control because they are incompatible.

Scheduling

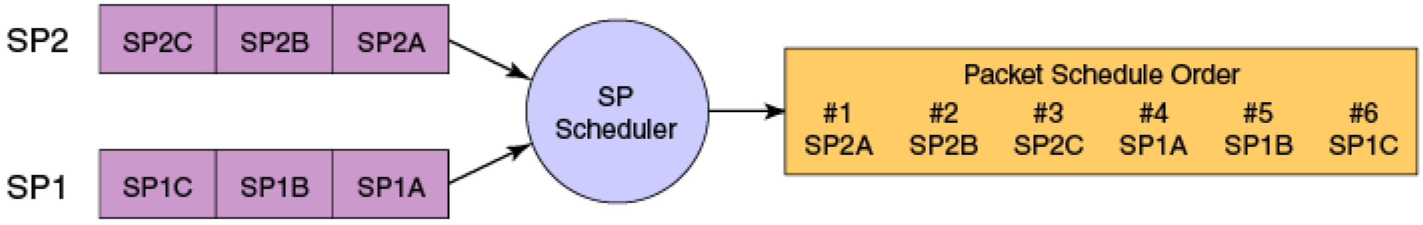
Scheduling arbitrates among multiple queues waiting to transmit a frame. The Extreme device supports both Strict Priority (SP) and Deficit Weighted Round Robin (DWRR) scheduling algorithms. Also supported is the flexible selection of the number of traffic classes using SP-to-DWRR. When there are multiple queues for the same traffic class, then scheduling takes these equal-priority queues into consideration.

Strict priority scheduling

Strict priority scheduling is used to facilitate support for latency-sensitive traffic. A strict priority scheduler drains all frames queued in the highest-priority queue before continuing on to service lower-priority traffic classes. A danger with this type of service is that a queue can potentially starve out lower-priority traffic classes.

The following figure displays the frame scheduling order for an SP scheduler servicing two SP queues. The higher-numbered queue, SP2, has a higher priority.

FIGURE 2 Strict priority schedule — two queues

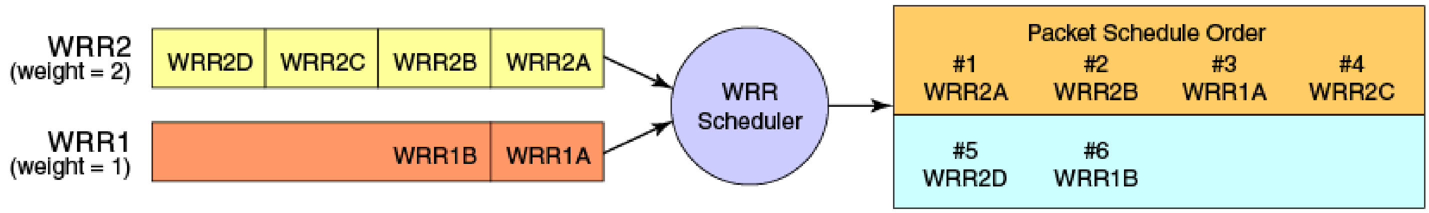


Weighted Round Robin scheduling

Weighted Round Robin (WRR) scheduling is used to facilitate controlled sharing of the network bandwidth. WRR assigns a weight to each queue; that value is then used to determine the amount of bandwidth allocated to the queue. The round robin aspect of the scheduling allows each queue to be serviced in a set order, sending a limited amount of data before moving onto the next queue and cycling back to the highest-priority queue after the lowest-priority queue is serviced.

The following figure displays the frame scheduling order for a WRR scheduler servicing two WRR queues. The higher-numbered queue is considered higher priority (WRR2), and the weights indicate the network bandwidth should be allocated in a 2:1 ratio between the two queues. In this figure WRR2 receives 66 percent of the bandwidth and WRR1 receives 33 percent. The WRR scheduler tracks the extra bandwidth used and subtracts it from the bandwidth allocation for the next cycle through the queues. In this way, the bandwidth utilization statistically matches the queue weights over longer time periods.

FIGURE 3 WRR schedule — two queues



Deficit Weighted Round Robin (DWRR) is an improved version of WRR. DWRR remembers the excess used when a queue goes over its bandwidth allocation and reduces the queue's bandwidth allocation in the subsequent rounds. This way the actual bandwidth usage is closer to the defined level when compared to WRR.

Traffic class scheduling policy

Traffic classes are numbered from 0 to 7, with higher-numbered traffic classes treated as having a higher priority. Extreme devices provide full flexibility in controlling the number of SP-to-WRR queues. The number of SP queues is specified as SP1 through 8, then the highest-priority traffic classes are configured for SP service and the remaining eight are WRR serviced. The supported scheduling configurations listed in the table below describes the set of scheduling configurations supported.

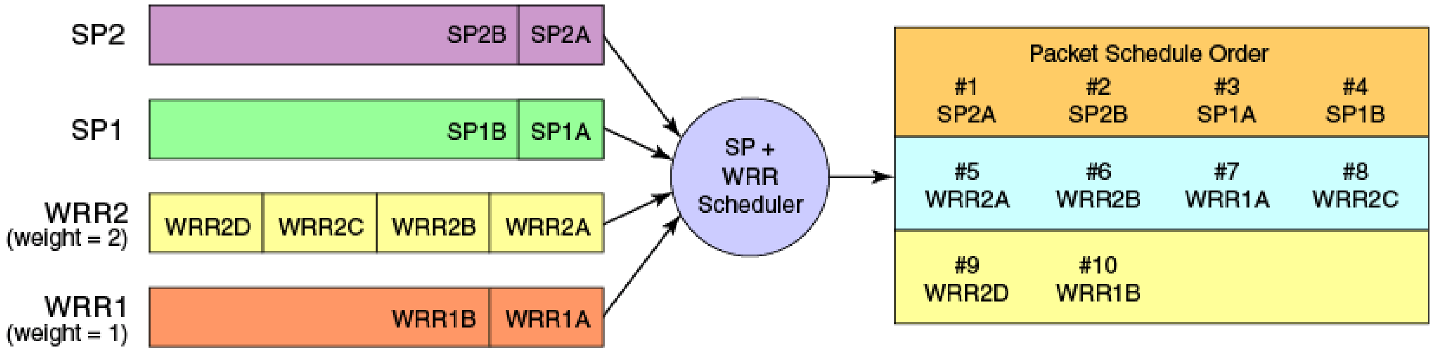
When you configure the QoS queue to use strict priority 4 (SP4), then traffic class 7 will use SP4, traffic class 6 will use SP3, and so on down the list. You use the strict priority mappings to control how the different traffic classes will be routed in the queue.

TABLE 2 Supported scheduling configurations

Traffic Class	SP0	SP1	SP2	SP3	SP4	SP5	SP6	SP8
7	WRR8	SP1	SP2	SP3	SP4	SP5	SP6	SP8
6	WRR7	WRR7	SP1	SP2	SP3	SP4	SP5	SP7
5	WRR6	WRR6	WRR6	SP1	SP2	SP3	SP4	SP6
4	WRR5	WRR5	WRR5	WRR5	SP1	SP2	SP3	SP5
3	WRR4	WRR4	WRR4	WRR4	WRR4	SP1	SP2	SP4
2	WRR3	WRR3	WRR3	WRR3	WRR3	WRR3	SP1	SP3
1	WRR2	WRR2	WRR2	WRR2	WRR2	WRR2	WRR2	SP2
0	WRR1	WRR1	WRR1	WRR1	WRR1	WRR1	WRR1	SP1

The figure below shows that extending the frame scheduler to a hybrid SP+WRR system is fairly straightforward. All SP queues are considered strictly higher priority than WRR so they are serviced first. Once all SP queues are drained, then the normal WRR scheduling behavior is applied to the non-empty WRR queues.

FIGURE 4 Strict priority and Weighted Round Robin scheduler



Multicast queue scheduling

The multicast traffic classes are numbered from 0 to 7; higher numbered traffic classes are considered higher priority. A fixed mapping from multicast traffic class to equivalent unicast traffic class is applied to select the queue scheduling behavior. The Multicast traffic class equivalence mapping table below presents the multicast traffic class with the equivalence mapping applied.

Once the multicast traffic class equivalence mapping has been applied, then scheduling and any scheduler configuration are inherited from the equivalent unicast traffic class. Refer to the table below for details on exact mapping equivalencies.

TABLE 3 Multicast traffic class equivalence mapping

Multicast traffic class	Equivalent unicast traffic class
0	0
1	1
2	2
3	3
4	4
5	5
6	6
7	7

Unicast ingress and egress queueing utilizes a hybrid scheduler that simultaneously supports SP+WRR service and multiple physical queues with the same service level. Multicast adds additional multicast expansion queues. Because multicast traffic classes are equivalent to unicast service levels, they are treated exactly as their equivalent unicast service policies.

Data Center Bridging QoS

Data Center Bridging (DCB) QoS covers frame classification, priority and traffic class (queue) mapping, congestion control, and scheduling. Under the DCB provisioning model, all of these features are configured on the basis of two configuration tables, Priority Group Table and Priority Table.

The DCB Priority Group Table defines each Priority Group ID (PGID) and its scheduling policy (Strict Priority versus DWRR, DWRR weight, relative priority), and partially defines the congestion control (PFC) configuration. There are 16 rows in the DCB Priority Group Table.

The table below presents the default DCB Priority Group Table configuration.

TABLE 4 Default DCB Priority Group Table configuration

PGID	Bandwidth%	PFC
15.0	–	Y
15.1	–	N
15.2	–	N
15.3	–	N
15.4	–	N
15.5	–	N
15.6	–	N
15.7	–	N
0	0	N
1	0	Y
2	0	N
3	0	N
4	0	N
5	0	N
6	0	N
7	0	N

NOTE
Only a single CoS can be mapped to a PFC-enabled priority queue. The switch automatically maps the CoS number to the same TC number when PFC is enabled. The PGID can be anything from 0 through 7. If your configuration violates this restriction an error message displays and the Priority Group Table is set back to the default values. When the DCB map is applied, and the interface is connected to the CNA, only one Strict Priority PGID (PGID 15.0 through PGID 15.7) is allowed.

Strict Priority versus DWRR is derived directly from the PGID value. All PGIDs with prefix 15 receive Strict Priority scheduling policy, and all PGIDs in the range 0 through 7 receive DWRR scheduling policy. Relative priority between Priority Group is exactly the ordering of entries listed in the table, with PGID 15.0 being highest priority and PGID 7 being lowest priority. Congestion control configuration is partially specified by toggling the PFC column On or Off. This provides only partial configuration of congestion control because the set of priorities mapped to the Priority Group is not known, which leads into the DCB Priority Table.

The DCB Priority Table defines each CoS mapping to Priority Group, and completes PFC configuration. The table below shows an example of mapping in the DCB Priority Table.

TABLE 5 Example of mapping DCB priority table values

CoS	PGID
0	15.6
1	15.7
2	15.5
3	15.4
4	15.3
5	15.2
6	15.1
7	15.0

Extreme VCS Fabric QoS

Extreme VCS Fabric QoS requires very little user configuration. The only options to modify are the fabric priority and the lossless priority.

Extreme VCS Fabric reserves a mapping priority and fabric priority of seven (7). Any traffic that enters the Extreme VCS Fabric cluster from upstream that is using the reserved priority value is automatically remapped to a lower priority.

Changing the mapping or fabric priority is not required. By default the values are set to zero (0) for both of the remapped priorities.

In Extreme VCS Fabric mode:

- All incoming priority 7 tagged packets are redefined to the default or user-defined value.
- Untagged control frames are counted in queue 7 (TC7).

All switches in the Extreme VCS Fabric cluster must have matching remapping priority values and the same priority-group-table values.

Restrictions for Layer 3 features in VCS mode

When the switch is in VCS mode, the lossless priority for carrying FCoE traffic and the fabric priority for carrying fabric traffic must be isolated from any Layer 3 QoS markings and classification. Therefore, specific restrictions apply to some Layer 3 DSCP QoS features when the switch is working in VCS mode:

The following are restrictions for using applicable Layer 3 DSCP-Traffic-Class map, DSCP-CoS map, and DSCP Trust features in VCS mode. Note that DSCP mutation maps and the WRED feature are not affected in VCS mode.

- DSCP trust is disabled in VCS mode as it is for CoS trust.
- There are no default DSCP maps in VCS mode.
- A nondefault DSCP-Traffic-Class map has the following restrictions:
  - A DSCP value cannot be classified to Traffic Class 7.
  - A DSCP value cannot be classified to a queue that carries lossless traffic (by default Traffic Class 3).
- A nondefault DSCP-CoS map has the following restrictions:
  - A DSCP value cannot be marked to CoS 7.
  - A DSCP value cannot be marked to lossless priority (by default CoS 3).
- Lossless priorities are identified through the CEE map.
- To enable DSCP based marking or classification, a nondefault DSCP-Traffic-Class map and a DSCP-CoS map have to be applied on the interface.
- To apply a DSCP-Traffic-Class or DSCP-CoS map to an interface, the CoS and Traffic Class values have to be remarked for lossless priorities. For example, when DSCP-Traffic-Class map "abcd" is created, it will have the default contents. When this map is applied to an interface, an error will display that the fabric and lossless priorities are used in the map and it cannot be applied on the interface.
- When a valid DSCP-Traffic-Class map and DSCP-CoS map are applied on the interface, then DSCP trust is enabled with the configured maps.

Port-based Policer

The port-based Policer feature controls the amount of bandwidth consumed by an individual flow or aggregate of flows by limiting the inbound and outbound traffic rate on an individual port according to criteria defined by the user. The Policer provides rate control by prioritizing or dropping ingress and egress packets classified according to a two-rate, three-color marking scheme defined by RFC 4115.

Rewriting

Rewriting a frame header field is typically performed by an edge device. Rewriting occurs on frames as they enter or exit a network because the neighboring device is untrusted, unable to mark the frame, or is using a different QoS mapping.

The frame rewriting rules set the Ethernet CoS and VLAN ID fields. Egress Ethernet CoS rewriting is based on the user-priority mapping derived for each frame as described later in the queueing section.

Port-based Policer features

The Policer supports the following features.

- A color-based priority mapping scheme for limiting traffic rates.
  - One-rate, two-color policing with "conform" color options. "Violate" color traffic will be dropped.
  - Two-rate, three-color policing with "conform" and "exceed" color options. "Violate" color traffic will be dropped.
- A policing option that allows packet headers to be modified for IP precedence.
- Policing options that allow packet headers to be modified for Class of Service (CoS).
- Policing options that allow packet headers to be modified for Differentiated Services Code Point (DSCP).
- Policing options that allow packets to be assigned to a traffic class (0-7).

Color-based priority

The following is the color-based priority mapping scheme for limiting traffic rate:

- Traffic flagged to the green or "conform" color priority conforms to the committed information rate (CIR) as defined by the *cir-rate* variable for the policy-map (refer to Policing parameters on page 22). This rate can be anything from 40000 bps to 100 Gbps.
- Traffic flagged as yellow or "exceed" exceeds the CIR, but conforms to the Excess Information Rate (EIR) defined by the *eir-rate* variable for the policy-map (refer to Policing parameters on page 22). This rate can be set from 0 through 100 Gbps.
- Traffic flagged as red or "violate" are not compared to CIR or EIR and will be dropped.

Using policing parameters, you can define metering rates, such as CIR and EIR, and actions for traffic flagged as conforming or exceeding the rates. As a simple example, traffic within the "conform" rate may be sent at a certain CoS priority, traffic flagged at the "exceed" rate may be sent at a lower priority, and traffic that violates the set rates can be dropped (default and only option).

Policing parameters

Policing parameters provide values for CIR, CBS, EIR, and EBS, for classifying traffic by a specific class for color-based priority mapping. They also specify specific actions to perform on traffic with a color-class priority, such as having packet DSCP priority, traffic class (internal queue assignment), or traffic class (internal queue assignment) set to specific values.

CIR and CBS

The Committed Information Rate (CIR) is the maximum number of bits that a port can receive or send during one-second over an interface. For CIR, there are two parameters that define the available traffic: CIR and the Committed Burst Size (CBS). The CIR represents a portion of the interface's total bandwidth expressed in bits per second (bps). It cannot be larger than the interface's total bandwidth. CBS controls the bursty nature of the traffic. Traffic that does not use the configured CIR accumulates credits until the credits reach the configured CBS. These credits can be used when the rate temporarily exceeds the configured CIR. When credits are not available, the traffic is either dropped or subject to the policy set for the Excess Information Rate (EIR). The traffic limited by the CIR can have its priority, traffic class, and DSCP values changed.

CIR is mandatory policing parameter for configuring a class map.

cir *cir-rate*

The **cir** command defines the value of the CIR as the rate provided in the *cir-rate* variable. Acceptable values are in multiples of 40000 in the range 40000-100000000000 bps.

cbs *cbs-size*

The **cbs** command defines the value of the CBS as the rate provided in the *cbs-size* variable. Acceptable values are 1250-12500000000 bytes in increments of 1 byte.

EIR and EBS

The Excess Information Rate (EIR) provides an option for traffic that has exceeded the CIR. For EIR, there are two parameters that define the available traffic: the EIR and the Excess Burst Size (EBS). The EIR and EBS operate exactly like the CIR and CBS, except that they act only upon traffic that has been passed to the EIR because it could not be accommodated by the CIR. Like the CIR, the EIR provides an initial bandwidth allocation to accommodate inbound and outbound traffic. Like the CBS, the bandwidth available for burst traffic from the EBS is subject to the amount of bandwidth that is accumulated during periods when traffic allocated by the EIR policy is not used. When inbound or outbound traffic exceeds the bandwidth available (accumulated credits or tokens), it is be dropped. The traffic rate limited by the EIR can have its priority, traffic class, and DSCP values changed.

EIR and EBS parameters are optional policing parameters. If not set, they are considered disabled.

eir *eir-rate*

The **eir** parameter defines the value of the EIR as the rate provided in the *eir-rate* variable. Acceptable values are in multiples of 40000 in the range 0-100000000000 bps.

ebs *ebs-size*

The **ebs** parameter defines the value of the EBS as the rate provided in the *ebs-size* variable. Acceptable values are 1250-12500000000 bytes in increments of 1 byte.

Parameters that apply actions to conform and exceed traffic

Following are policing parameters that apply actions to conform or exceed color traffic:

- **conform-set-dscp** *dscp-num*.

  The **conform-set-dscp** parameter specifies that traffic with bandwidth requirements within the rate configured for CIR will have its packet DSCP priority set to the value specified in the *dscp-num* variable. Acceptable values for *dscp-num* are 0–63.
- **conform-set-prec** *prec-num*.

  The **conform-set-prec** parameter specifies that traffic with bandwidth requirements within the rate configured for CIR will have its packet IP precedence value (first 3 bits of DSCP) set to the value in the *prec-num* variable. Acceptable values for *prec-num* are 0–7.
- **conform-set-tc** *trafficlass*.

  The **conform-set-tc** parameter specifies that traffic with bandwidth requirements within the rate configured for CIR will have its traffic class (internal queue assignment) set to the value in the *trafficlass* variable. Acceptable values for *trafficclass* are 0–7.
- **exceed-set-dscp** *dscp-num*.

  The **exceed-set-dscp** parameter specifies that traffic with bandwidth requirements that exceeds the rate configured for CIR and sent to the EIR bucket will have its packet DSCP priority set to the value in the *dscp-num* variable. Acceptable values for *dscp-num* are 0–63.

- **exceed-set-prec** *prec-num*.

  The **exceed-set-prec** parameter specifies that traffic with bandwidth requirements that exceed the rate configured for CIR and sent to the EIR bucket will have its packet IP precedence set to the value in the *prec-num* variable. Acceptable values for *prec-num* are 0–7.
- **exceed-set-tc** *trafficclass*.

  The **exceed-set-tc** parameter specifies that traffic with bandwidth requirements that exceed the rate configured for CIR and is in the limit of what is configured for EIR will have its traffic class (internal queue assignment) set to the value in the *trafficlass* variable. Acceptable values for *trafficlass* are 0–7.
- **set-priority** *priority-mapname*.

  The **set-priority parameter** specifies the mapping used for setting QoS priority (802.1p priority) in the packet. The *priority-mapname* name variable should be same as configured for the priority-map (police-priority-map), which will have a set priority and color type (conform or exceed).

Policer considerations and limitations

Consider the topics discussed below when configuring the port-based Policer feature.

Best practices for Policer

Follow these best practices when configuring the port-based Policer feature:

- Avoid mapping lossy priority to lossless priority in conform and exceed CoS maps.
- Configure rate (CIR or EIR) and burst size (CBS or EBS) based on interface speed.
- Set conform and exceed token count (Tc) to the same values to avoid any reordering issues.

Configuration rules and considerations for Policer

The following are rules for configuring maps and using policing parameters for the Policer feature:

- A policy-map, class map, priority-map name must be unique among all maps of that type.
- A policy-map is not supported on an ISL port.
- A Policer name must begin with a-z, or A-Z. You can use underscore, hyphen, and numeric values 0-9 except as the first character.
- You cannot delete a policy-map, class map, or priority-map if is active on the interface.
- You cannot delete a class map from a policy-map when the policy-map is active on the interface.
- Configure **CIR** and **EIR** in multiples of 40000 bps.
- Percentage as a rate limit is not supported,
- Policer actions are applicable only to data traffic. Control traffic, FCoE, and internal VLAN traffic is not subjected to policing.
- The egress Policer can overwrite ingress Policer results such as CoS mapping and DSCP mapping.
- If a policy-map is applied to an interface and no Policer attributes are present in that policy-map, then ingress and egress packets on that interface is marked as green (conforming).
- If the configured CBS value is less than 2*MTU value, then 2*MTU is programmed as the CBS in the hardware. For example, if you configure CBS at 4000 bytes and the MTU on an interface is 3000 bytes, when a policy-map is applied on this interface, the CBS programmed in the hardware is 2*MTU (6000 bytes).

- If CBS and EBS values are not configured, then these values are derived from CIR and EIR values, respectively. Burst size calculation is as follows: `Burst size (cbs or ebs) = 1.2*information rate (CIR/EIR)/8`
- If you do not configure EIR and EBS, then the single-rate, two-color scheme is applied (packets are marked as either green or red).
- You must configure rate limit threshold values on an interface based on interface speed. No validation is performed for user-configured values against interface speed.

Limitations for Policer

- The incremental step size for CIR or EIR is set to 40000 bps.
- The Policer operates in color-blind mode. In other words, color is evaluated at ingress and egress Policers independently. This may result in packets that are marked as yellow in the inbound Policer to be evaluated as green at the outbound Policer, depending on Policer settings.
- Because inbound queue scheduling is performed before outbound policing, setting traffic class (**set-conform-tc** or **set-exceed-tc**) based on policing results does not affect packet forwarding at the outbound side.
- Packets drops caused by any action other than ACLs are included in Policer counters.
- Layer 3 control packets are policed at the outbound side.
- Policing is enabled on lossless priorities at the outbound side.
- Policing is not enabled for control traffic that is trapped or dropped.

Considerations for vLAGs with Policer

Because a virtual link aggregation group (vLAG) spans multiple switches, it is not possible to associate flows on each LAG member port to a common Policer. Instead, apply the same policy-map on individual member ports so that traffic flow on member ports is controlled by a Policer configured on that member port. The total rate-limit threshold value on a vLAG consists of the cumulative values of rate-limit thresholds on all member ports.

Lossless traffic with Policer

The following are considerations for lossless traffic:

- Policing is applicable only for lossy traffic. Lossless traffic should not get policed. For port-based policing, apply a policy-map to an interface even if PFC is configured on that interface. The CoS value (priority) on which PFC is applied is excluded from being policed.
- Remapped priority values should not include lossless priorities. Do not remap lossy traffic priorities to lossless traffic priorities and vice-versa.
- Policer attributes **conform-set-tc** and **exceed-set-tc** should not be set to a lossless traffic class.

Configuring QoS

The following sections discuss configuring QoS, including fundamentals, traffic class mapping, congestion control, rate limiting, BUM storm control, scheduling, DCB QoS, Extreme VCS Fabric QoS, policer functions, and Auto QoS.

Configuring QoS fundamentals

NOTE

Refer to User-priority mapping on page 12.

Understanding default user-priority mappings for untrusted interfaces

When Layer 2 QoS trust is set to **untrusted** , then the default is to map all Layer 2 switched traffic to the port default user priority value of 0 (best effort), unless the user priority is configured to a different value.

The following table presents the Layer 2 QoS **untrusted** user-priority generation table.

TABLE 6 Default priority value of untrusted interfaces

Incoming CoS	User Priority
0	port <user priority> (default 0)
1	port <user priority> (default 0)
2	port <user priority> (default 0)
3	port <user priority> (default 0)
4	port <user priority> (default 0)
5	port <user priority> (default 0)
6	port <user priority> (default 0)
7	port <user priority> (default 0)

NOTE

Nontagged Ethernet frames are interpreted as having an incoming CoS value of 0 (zero).

You can override the default user-priority mapping by applying explicit user-priority mappings.

When neighboring devices are trusted and able to properly set QoS, then Layer 2 QoS trust can be set to **CoS** and the IEEE 802.1Q default-priority mapping is applied.

The following table presents the Layer 2 CoS user-priority generation table conforming to 802.1Q default mapping. You can override this default user-priority table per port if you want to change (mutate) the CoS value.

TABLE 7 IEEE 802.1Q default priority mapping

Incoming CoS	User Priority
0	0
1	1
2	2
3	3
4	4
5	5
6	6
7	7

Configuring QoS mappings

Consider the topics discussed below when configuring the QoS mappings.

Configuring user-priority mappings

To configure user-priority mappings, perform the following steps from privileged EXEC mode.

1. Enter global configuration mode.

   ```
   device# configure terminal
   ```

2. Specify the Ethernet interface.

   ```
   device(config)# interface tengigabitethernet 1/2/2
   ```

3. Configure the interface to priority 3.

   ```
   device(conf-if-te-1/2/2)# qos cos 3
   ```

4. Return to privileged EXEC mode.

   ```
   device(conf-if-te-1/2/2)# end
   ```

Creating a CoS-to-CoS mutation QoS map

To create a CoS-to-CoS mutation, perform the following steps from privileged EXEC mode.

1. Enter global configuration mode.

   ```
   device# configure terminal
   ```

2. Create the CoS-to-CoS mutation QoS map. In this example "test" is the map name.

   ```
   device(config)# qos map cos-mutation test 0 1 2 3 4 5 6 7
   ```

Applying a CoS-to-CoS mutation QoS map to an interface

To apply a CoS-to-CoS mutation QoS map, perform the following steps from privileged EXEC mode.

1. Enter global configuration mode.

   ```
   device# configure terminal
   ```

2. Specify the Ethernet interface.

   ```
   device(config)# interface tengigabitethernet 2/1/2
   ```

3. Activate or apply changes made to the CoS-to-CoS mutation QoS map. In this example, "test" is the map name.

   ```
   device(conf-if-te-2/1/2)# qos cos-mutation test
   ```

   NOTE
   To deactivate the mutation map from an interface, enter the **no qos cos-mutation** command.

4. Return to privileged EXEC mode.

   ```
   device(conf-if-te-2/1/2)# end
   ```

Verifying CoS-to-CoS mutation QoS mapping

To verify applied QoS maps, you can use one or both of the following options from global configuration mode.

- Verify the CoS mutation mapping for a specific map by using the **do show qos maps qos-mutation** command and the map name.

  ```
  switch(config)# do show qos maps cos-mutation test
  ```

- Verify all QoS mapping by using the **do show qos maps** command with just the **cos-mutation** parameter only.

  ```
  switch(config)# do show qos maps cos-mutation
  ```

- Verify the interface QoS mapping by using the **do show qos interface** command.

  ```
  switch(config)# do show qos interface te 2/1/2
  ```

Configuring DSCP mappings

Consider the topics discussed below when configuring the DSCP mappings.

Configuring the DSCP trust mode

Like QoS trust mode, the Differentiated Services Code Point (DSCP) trust mode controls the user-priority mapping of incoming traffic.

The user priority is based on the incoming DSCP value. When this feature is not enabled, DSCP values in the packet are ignored.

When DSCP trust is enabled, the following table shows default mapping of DSCP values to user priority.

TABLE 8 Default DSCP priority mapping

DSCP Values	User Priority
0-7	0
8-15	1
16-23	2
24-31	3
32-39	4
40-47	5
48-55	6
56-63	7

> NOTE
>
> Note the restrictions for using this feature in VCS mode under Restrictions for Layer 3 features in VCS mode on page 21.

To configure DSCP trust mode, perform the following steps from privileged EXEC mode.

1. Enter global configuration mode.

   ```
   device# configure terminal
   ```

2. Specify the Ethernet interface.

   ```
   device(config)# interface tengigabitethernet 10/0/2
   ```

3. Set the interface mode to QoS DSCP trust.

   ```
   device(conf-if-te-10/0/2)# qos trust dscp
   ```

   NOTE
   To deactivate the DSCP trust mode from an interface, enter **no qos trust dscp**.

4. Return to privileged EXEC mode.

   ```
   device(conf-if-te-10/0/2)# end
   ```

Verifying DSCP trust

To verify applied DSCP trust, you can enter the following command from global configuration mode, where **tengigabitethernet 10/0/2** is the interface name.

```
switch(config)# do show qos interface tengigabitethernet 10/0/2
```

Creating a DSCP mutation map

To create a DSCP mutation map and re-map the incoming DSCP value of the ingress packet to egress DSCP values, perform the following steps from privileged EXEC mode.

NOTE
This feature is only supported on Extreme VDX 8770-4, VDX 8770-8, VDX 6740, VDX 6740T, and VDX 6740T-1G devices.

1. Enter global configuration mode.

   ```
   device# configure terminal
   ```

2. Create the DSCP mutation map by specifying a map name. The following command uses "test" as the map name and places the system in DSCP mutation mode so that you can map to traffic classes.

   ```
   device(config)# qos map dscp-mutation test
   ```

3. Once the system is in DSCP mutation mode for the configured map (in this case dscp-mutation-test), you can map ingress DSCP values to egress DSCP values by using the **mark** command as in the following examples:

   ```
   device(dscp-mutation-test)# mark 1,3,5,7 to 9
   device(dscp-mutation-test)# mark 11,13,15,17 to 19
   device(dscp-mutation-test)# mark 12,14,16,18 to 20
   device(dscp-mutation-test)# mark 2,4,6,8 to 10
   ```

   This sets the following:

   - DSCP values 1, 3, 5, and 7 are set to output as DSCP number 9.
   - DSCP values 11, 13, 15, and 17 are set to output as DSCP number 19.
   - DSCP values 12, 14, 16, and 18 are set to output as DSCP number 20
   - DSCP values 2, 4, 6, and 8 are set to output as DSCP number 10.

Applying a DSCP mutation map to an interface

To apply a configured DSCP mutation map to an interface, perform the following steps from privileged EXEC mode.

1. Enter global configuration mode.

   ```
   device# configure terminal
   ```

2. Specify the Ethernet interface.

   ```
   device(config)# interface tengigabitethernet 3/1/2
   ```

3. Activate or apply changes made to the DSCP mutation map to the interface. In this example "test" is the map name.

   ```
   device(conf-if-te-3/1/2)# qos dscp-mutation test
   ```

   NOTE
   To deactivate a map from an interface, enter **no qos dscp-mutation** *name*.

4. Specify the DSCP trust mode for incoming traffic.

   ```
   device(conf-if-te-2/1/2)# qos dscp-cos test
   device(conf-if-te-2/1/2)# qos dscp-traffic-class test
   ```

5. Return to privileged EXEC mode.

   ```
   device(conf-if-te-3/1/2)# end
   ```

Verifying the DSCP mutation mapping

To verify applied DSCP maps, you can use one or both of the following options from global configuration mode.

- Verify DSCP mapping for a specific map using the **do show qos maps dscp-mutation** command and the map name.

  ```
  switch(config)# do show qos maps dscp-mutation test
  ```

- Verify all DSCP mapping by using the **do show qos maps** command with the **dscp-mutation** operand only.

  ```
  switch(config)# do show qos maps dscp-mutation
  ```

- Verify DSCP mutation mapping for an interface by using the **show qos interface** command and specifying the interface:

  ```
  switch(config)# do show qos interface te 3/1/2
  ```

Configuring DSCP-to-CoS mappings

Consider the topics discussed below when configuring the DSCP-to-CoS mappings.

Creating a DSCP-to-CoS mutation map

You can use the incoming DSCP value of ingress packets to remap the outgoing 802.1P CoS priority values by configuring a DSCP-to-COS mutation map on the ingress interface.

> NOTE
> The restrictions for using this feature in VCS mode are listed at Restrictions for Layer 3 features in VCS mode on page 21.

1. Enter global configuration mode.

```
device# configure terminal
```

2. Create the DSCP-to-CoS map by specifying a map name. The following command uses "test" as the map name and places the system in dscp-cos map mode so that you can map DSCP values to CoS values.

```
device(configure)# qos map dscp-cos test
```

3. Once the system is in dscp-cos map mode for the configured map (in this case dscp-cos-test), you can map incoming DSCP values to outgoing CoS priority values by using the **mark** command as in the following examples:

```
device(dscp-cos-test)# mark 1,3,5,7 to 3
device(dscp-cos-test)# mark 11,13,15,17 to 5
device(dscp-cos-test)# mark 12,14,16,18 to 6
device(dscp-cos-test)# mark 2,4,6,8 to 7
```

This sets the following:

- DSCP values 1, 3, 5, and 7 are set to output as CoS priority 3.
- DSCP values 11, 13, 15, and 17 are set to output as CoS priority 5
- DSCP values 12, 14, 16, and 18 are set to output as CoS priority 6
- DSCP values 2, 4, 6, and 8 are set to output as CoS priority 7.

Applying a DSCP-to-CoS map to an interface

To apply a DSCP-to-CoS mutation map to an interface, perform the following steps from privileged EXEC mode.

1. Enter global configuration mode.

```
device# configure terminal
```

2. Specify the Ethernet interface.

```
device(config)# interface tengigabitethernet 1/1/2
```

3. Apply the changes made to the DSCP-to-CoS mutation map to enable DSCP trust on the interface. In this example, "test" is the map name.

```
device(conf-if-te-1/1/2)# qos dscp-cos test
```

> NOTE
> To deactivate a map from an interface, enter **no qos dscp-cos** *name*.

4. Apply the changes made to the DSCP-to-Traffic-Class map to enable DSCP trust on the interface. In this example, "traffic_test" is the map name.

```
device(conf-if-te-1/1/2)# qos dscp-traffic-class traffic_test
```

5. Return to privileged EXEC mode.

```
device(conf-if-te-1/1/2)# end
```

Verifying a DSCP-to-CoS mutation map

To verify applied DSCP-to-CoS maps, you can use one or both of the following options from global configuration mode.

- Verify DSCP mapping for a specific map using the **do show qos maps dscp-cos** command and the map name.

```
switch(config)# do show qos maps dscp-cos test
```

- Verify all DSCP mapping by using the **do show qos maps** command with the **dscp-cos** operand only.

```
switch(config)# do show qos maps dscp-cos
```

- Verify DSCP-to-CoS mutation mapping for an interface by using the **show qos interface** command and specifying the interface:

```
switch(config)# do show qos interface te 1/1/2
```

Configuring traffic class mapping for flexibility

Where additional flexibility in queue selection is required, refer to Configuring traffic class mapping on page 32.

Configuring traffic class mapping

The Extreme device supports eight unicast traffic classes to provide isolation and to control servicing for different priorities of application data. Traffic classes are numbered from 0 through 7, with higher values designating higher priorities.

NOTE
For information on user-priority mapping, refer to User-priority mapping on page 12.

The traffic class mapping stage provides some flexibility in queue selection:

- The mapping may be many-to-one, such as mapping one-byte user priority (256 values) to eight traffic classes.
- There may be a nonlinear ordering between the user priorities and traffic classes.

NOTE
The command **qos trust cos** is not applicable in VCS mode. However, the **show qos interface** command will show trusted ports if the **cos-mutation** command and the **cee** *default* command are applied.

Understanding unicast and multicast traffic defaults

The following table displays the Layer 2 default traffic **unicast** class mapping supported for a CoS-based user priority to conform to .

TABLE 9 Default user priority for unicast traffic class mapping

User priority	Traffic class
0	1
1	0
2	2
3	3

TABLE 9 Default user priority for unicast traffic class mapping (continued)

User priority	Traffic class
4	4
5	5
6	6
7	7

You are allowed to override these default traffic class mappings per port. Once the traffic class mapping has been resolved, it is applied consistently across any queuing incurred on the ingress and the egress ports.

The Extreme device supports eight multicast traffic classes for isolation and to control servicing for different priorities of application data. Traffic classes are numbered from 0 through 7, with higher values designating higher priorities. The traffic class mapping stage provides some flexibility in queue selection.

The following table displays the Layer 2 default traffic **multicast** class mapping supported for a CoS-based user priority to conform to 802.1Q default mapping.

TABLE 10 Default user priority for multicast traffic class mapping

User Priority	Traffic class
0	1
1	0
2	2
3	3
4	4
5	5
6	6
7	7

Once the traffic class mapping has been resolved for inbound traffic, it is applied consistently across all queuing incurred on the ingress and egress ports.

> NOTE
> You can configure an interface with a nondefault DSCP-to-traffic class-map. However, configuring an interface with a nondefault CoS-to-traffic class-map is not supported.

Configuring congestion control

Refer also to Congestion control on page 12.

Changing the multicast tail-drop threshold

To change the tail drop threshold, perform the following steps from privileged EXEC mode.

1. Enter global configuration mode.

```
device# configure terminal
```

2. Change the tail drop threshold for each multicast traffic class. In this example, a threshold of 1000 packets is used.

   ```
   device(config)# qos rcv-queue multicast threshold 1000 1000 1000 1000 1000 1000 1000 1000
   ```

3. Return to privileged EXEC mode.

   ```
   device(config)# end
   ```

Configuring Weighted Random Early Detection

Consider the topics discussed below when configuring Weighted Random Early Detection (WRED) mappings.

Understanding WRED profiles

Consider the following when configuring WRED.

- Trunk ports cannot share WRED profiles with any other ports because the bandwidth for a trunk port changes according to the number of active links in the trunk.
- When queue thresholds in a WRED profile are configured by percentage, the switch maps this to a total number of bytes as buffers allocated to a port depend on the port speed.
- A total of 384 WRED profiles are supported per chassis.

Consider the following when using WRED profiles for link aggregation (LAG) interfaces:

- WRED profiles can be enabled on LAG interfaces. However, the profile is configured on the individual member interfaces of the LAG.
- Because LAG members may belong to different port groups, one of the port groups may not have enough resources available to support a new WRED configuration for the member interface. In this case, and error log will indicate that the WRED application failed on the specific member interface. When a new member is added to the port-channel, the same error may occur if the new member belongs to an port groups with all resources used. To apply the WRED profile on the failed member interface, you must remove the WRED configuration on other all interfaces in the port group so that resources are available and remove or add the member interface to the LAG.

Configuring WRED profiles

To configure an egress WRED profile, perform the following steps from privileged EXEC mode.

1. Enter global configuration mode.

   ```
   device# configure terminal
   ```

2. Configure a WRED profile. For the profile ID, 10 is used in this case. The *min-threshold*, *max-threshold*, and *drop-probability* values are percentages.

   ```
   device(config)# qos red-profile 10 min-threshold 10 max-threshold 80 drop-probability 80
   ```

3. Return to privileged EXEC mode.

   ```
   device(config)# end
   ```

Mapping a traffic-class to an WRED profile on an interface

To map a traffic-class value for a port to a WRED profile, perform the following steps from privileged EXEC mode.

1. Enter global configuration mode.

   ```
   device# configure terminal
   ```

2. Specify the Ethernet interface.

   ```
   device(config)# interface tengigabitethernet 1/2/2
   ```

3. Map the profile to use a traffic-class for a port. In the following example, CoS priority 3 is mapped to WRED profile ID 10.

   ```
   device(conf-if-te-1/2/2)# qos random-detect traffic-class 3 red-profile-id 10
   ```

   NOTE
   To deactivate the map from an interface, enter **no qos random-detect traffic-class** *value*

4. Return to privileged EXEC mode.

   ```
   device(conf-if-te-1/2/2)# end
   ```

Verifying WRED profiles

Verify a configured WRED profiles by using the **show qos red profiles** command.

```
switch# show qos red profiles
Red Profile 1
        Minimum Threshold: 10
        Maximum Threshold: 100
        Drop Probability: 100

    Activated on the following interfaces:
Te 1/1/11  traffic-class: 1


Red Profile 2
        Minimum Threshold: 10
        Maximum Threshold: 100
        Drop Probability: 100

    Activated on the following interfaces:
Te 1/1/11  traffic-class: 2


Red Profile 3
        Minimum Threshold: 10
        Maximum Threshold: 100
        Drop Probability: 100

    Activated on the following interfaces:
Te 1/1/11  traffic-class: 3
```

Examine the applied WRED profiles for an interface by using the **show qos red statistics interface** *interface-name* command. This displays the WRED profile, as well as all QoS configurations applied to the interface, such as DSCP trust, DSCP-to-DSCP map, CoS-Traffic Class map, and others.

```
switch# show qos red statistics interface te 1/2/2
interface TenGigabitEthernet 1/2/2
fabric isl enable
fabric trunk enable
qos random-detect traffic-class 1 red-profile-id 1
```

```
qos random-detect traffic-class 2 red-profile-id 2
qos random-detect traffic-class 3 red-profile-id 3
no shutdown
!
```

Configuring FlowControl

This task configures FlowControl in addition to enabling the Ethernet pause frames. Extreme recommends that you also configure the flow control parameters on the connecting device, and not leave the options set to "auto".

Refer to Ethernet Pause on page 15.

Enabling Ethernet Pause

To enable Ethernet Pause, perform the following steps from privileged EXEC mode.

1. Enter global configuration mode.

   ```
   device# configure terminal
   ```

2. Specify the Ethernet interface.

   ```
   device(config)# interface tengigabitethernet 3/0/2
   ```

3. Enable Ethernet Pause on the interface for both TX and RX traffic.

   ```
   device(conf-if-te-3/0/2)# qos flowcontrol tx on rx on
   ```

   NOTE
   To deactivate Ethernet pause on an interface, enter **no qos flowcontrol**

4. Return to privileged EXEC mode.

   ```
   device(conf-if-te-3/0/2)# end
   ```

5. Verify the Ethernet Pause with the **show qos flowcontrol** command.

   ```
   device# show qos flowcontrol interface all 3/0/2
   ```

Configuring dynamic buffer sharing

If there is bursty, lossy traffic for certain flows in the system, you can borrow the buffers from less bursty flows, in order to reduce the traffic loss. The **qos** command is used to configure the egress or ingress queue limit, such as the maximum number of kilobytes of data that can be queued in the egress or ingress queue. The **tx-queue** keyword controls the egress, and the **rcv-queue** keyword controls the ingress.

This command only functions on the Extreme VDX 6740 series, VDX 6940 series, and VDX 2746 devices. This configuration is applied on individual RBridges.

To configure dynamic buffer sharing, perform the following steps from privileged EXEC mode.

1. Enter global configuration mode.

   ```
   device# configure terminal
   ```

2. Enter RBridge ID configuration mode.

   ```
   device(config)# rbridge-id 154
   ```

3. Configure the dynamic buffer sharing. The following example sets the egress limit to 256 kilobytes. (The default is 512 kilobytes.)

   ```
   device(config-rbridge-id-154)# qos tx-queue limit 256
   ```

   The following example sets the ingress limit to 1024 kilobytes. (The default is 285 kilobytes.)

   ```
   device(config-rbridge-id-154)# qos rcv-queue limit 1024
   ```

4. Return to privileged EXEC mode.

   ```
   device(config-rbridge-id-154)# end
   ```

Configuring scheduling

Refer also to Scheduling on page 17.

Scheduling the QoS multicast queue

To schedule the QoS multicast queue, perform the following steps from privileged EXEC mode.

1. Enter global configuration mode.

   ```
   device# configure terminal
   ```

2. Specify the schedule to use and the traffic class to bandwidth mapping.

   ```
   device(config)# qos queue multicast scheduler dwrr 10 20 20 10 10 10 10 10
   ```

3. Return to privileged EXEC mode.

   ```
   device(config)# end
   ```

Configuring DCB QoS

Refer also to Data Center Bridging QoS on page 19.

Creating a DCB map

To create a DCB map, perform the following steps from privileged EXEC mode.

1. Enter global configuration mode.

   ```
   device# configure terminal
   ```

2. Select the DCB map by using the **cee-map** command.

   The only map name allowed is "default."

   ```
   device(config)# cee-map default
   ```

3. Return to privileged EXEC mode.

   ```
   device(config)# exit
   ```

Defining a DCB priority group table

To define a priority group table map, perform the following steps from privileged EXEC mode.

1. Enter global configuration mode.

   ```
   device# configure terminal
   ```

2. Specify the name of the DCB map to define by using the **cee-map** command.

   NOTE
   The only map name allowed is "default."

   ```
   device(config)# cee-map default
   ```

3. Define the DCB map for PGID 0.

   ```
   device(config-cee-map-default)# priority-group-table 0 weight 50 pfc on
   ```

4. Define the DCB map for PGID 1.

   ```
   device(config-cee-map-default)# priority-group-table 1 weight 50 pfc off
   ```

5. Return to privileged EXEC mode.

   ```
   device(config-cee-map-default)# end
   ```

Defining a DCB priority-table map

To define a priority-table map, perform the following steps from privileged EXEC mode.

1. Enter global configuration mode.

   ```
   device# configure terminal
   ```

2. Specify the name of the DCB map to define by using the **cee-map** command.

   ```
   device(config)# cee-map default
   ```

3. Define the map.

   ```
   device(config-cee-map)# priority-table 1 1 1 0 1 1 1 15.0
   ```

4. Return to privileged EXEC mode.

   ```
   device(config-cee-map)# end
   ```

Applying a DCB provisioning map to an interface

To apply a DCB provisioning map, perform the following steps from privileged EXEC mode.

1. Enter global configuration mode.

   ```
   device# configure terminal
   ```

2. Specify the Ethernet interface. In this example, 101/0/2 is used.

   ```
   device(config)# interface tengigabitethernet 101/0/2
   ```

3. Apply the DCB map on the interface.

   ```
   device(conf-if-te-101/0/2)# cee default
   ```

   NOTE
   To deactivate the map on the interface, enter **no cee**.

4. Return to privileged EXEC mode.

   ```
   device(conf-if-te-101/0/2)# end
   ```

Verifying the DCB maps

To verify the CoS DCB map, use the **show cee maps default** command from privileged EXEC mode.

```
switch# show cee maps default
```

Manually enabling lossless RDMA over Ethernet

Manually enabling lossless Remote Direct Memory Access (RDMA) over Ethernet is an alternate to the default automatic method of enabling lossless Ethernet.

In Network OS, lossless 10 Gbps or 40 Gbps Ethernet is enabled automatically when an attached host signals the desire to do so through the Data Center Bridging Capability Exchange protocol (DCBX) extension mechanism

DCBX advertises this lossless request by exchanging protocol information through Type, Length, and Value attributes (DCBX TLV). This is an automatic mechanism that does not require Network OS configuration because both LLDP and DCBX are enabled by default. This is used to enable lossless Ethernet for FCoE and for some iSCSI targets which support the DCBX TLV, such as Dell EqualLogic arrays.

There are cases where lossless 10 Gbps or 40 Gbps Ethernet is desired, but where the host does not support enabling this automatically through DCBX TLV. A common example is to support Remote Direct Memory Access (RDMA) over Ethernet (also known as “RDMA over Converged Enhanced Ethernet” or “RoCE”). One example of this is to support SMB Direct in Microsoft Windows Server 2012.

Networks adapters which support RDMA over Ethernet need to have DCB Priority Flow Control (DCB PFC) enabled, and the network switches in the path also need to have DCB PFC set to the same priority in order to provide lossless Ethernet. The following are the steps needed to manually configure lossless Ethernet in Network OS (Network OS 4.0.1 or greater is required):

1. For VDX switch interfaces that are attached to hosts requiring lossless Ethernet support, disable LLDP at the 10GbE or 40GbE interface level by issuing the **lldp disable** command.
2. Also at the interface level, enable the cee map by issuing the **cee default** command.
3. Configure the host’s network adapter to enable DCB PFC and to set it for PFC class 3 (priority 3). This will match the default PFC class setting in Network OS. These instructions are provided by the network adapter manufacturer.

4. It is important that the DCB PFC settings match on both the VDX switch and the host's network adapter. Ensure that the global NOS setting for the CEE map remains at its default: cee-map default

This simplifies configuration and ensures that lossless Ethernet will be automatically and correctly set up across the VCS fabric. This gives the added benefit of automatic, end-to-end lossless Ethernet throughout the VCS Ethernet fabric. When lossless Ethernet is enabled manually for PFC class 3, this will preclude support of FCoE or other protocols on PFC class 3. For this reason, do not mix FCoE configurations with manual lossless Ethernet configurations.

Configuring Extreme VCS Fabric QoS

Refer also to Extreme VCS Fabric QoS on page 21.

To configure the remapping priorities for the Extreme VCS Fabric, perform the following steps from global configuration mode.

1. Use the **cee-map** command to enter CEE map configuration mode.

   ```
   device(config)# cee-map default
   ```

2. Use the **remap lossless priority** command to set the lossless priority for Extreme VCS Fabric QoS.

   The default lossless remap priority is set to 0 (zero).

   ```
   device(config-cee-map-default)# remap lossless-priority priority 2
   ```

3. Use the **remap fabric priority** command to set the fabric priority for Extreme VCS Fabric QoS.

   The default FCoE remap fabric priority is set to 0 (zero).

   ```
   device(fabric-cee-map-default)# remap fabric-priority priority 2
   ```

4. Use the **exit** command to return to global configuration mode.

   ```
   device(config-cee-map)# exit
   ```

5. Specify the incoming Ethernet interface.

   ```
   device(config)# interface tengigabitethernet 22/0/1
   ```

6. Apply the CEE Provisioning map to the interface.

   ```
   device(conf-if-te-22/0/1)# cee default
   ```

Configuring DCB QoS

NOTE
Flow-based QoS functions only in the ingress direction.

To configure flow-based QoS functions, do the following while the switch is in global configuration mode:

1. Configure a class-map to classify traffic according to the traffic properties required for your flow-based QoS needs. Refer to Configuring a class-map on page 41.
2. Configure a policy-map to associate a policy-map with the class-map and also add the QoS action to be applied on the type of flow determined by the class-map. Refer to Configuring flow-based QoS actions using policy-map on page 41.
3. Bind the policy-map to a specific interface using the **service-policy** command, or bind the policy-map to an Rbridge ID. Refer to Binding the policy-map to an interface on page 45 and Binding flow-based QoS at the system level on page 46.

Configuring flow-based QoS classifications using a class-map

Consider the topics discussed below when configuring the flow-based QoS classifications.

Configuring a class-map

The classification map classifies traffic based on match criteria that you configure. If traffic matches the criteria, it belongs to the class.

1. Enter global configuration mode.

   ```
   device# configure terminal
   ```

2. Create an access-list (either a MAC, IP, or VLAN-based ACL) to define the traffic. Refer to the Network OS Security Guide for details on creating access-lists.

   ```
   device(config)# mac access-list standard ACL1
   device(conf-macl-std)# permit host 0000.00aa.aa00
   device(conf-macl-std)# exit
   ```

3. Create a class-map by providing a class-map name. This enables class-map configuration mode.

   ```
   device(config)# class-map class1
   ```

   The name for the class-map (in this case class1) can be a character string up to 64 characters.

   > NOTE
   > The "default" class-map and "cee" class-map name is reserved and intended to match everything. It is always created and cannot be removed.

4. Provide match criteria for the class. Currently, the only match criterion is "match access-group".

   ```
   device(config-classmap)# match access-group ACL1
   ```

5. Exit the class-map configuration mode.

   ```
   device(config-classmap)# exit
   ```

6. Return to privileged EXEC mode.

   ```
   device(config)# end
   ```

Flow-based QoS actions using policy-map

Consider the topics discussed below when configuring the flow-based QoS actions.

Configuring flow-based QoS actions using policy-map

Configure a rate-limit policy-map to associate a policy-map with the class-map and add a QoS action to be applied to the type of QoS flow defined by the class-map.

1. Enter the global configuration mode.

   ```
   device# configure terminal
   ```

2. Configure a policy-map by providing a policy-map name. This enables policy-map configuration mode.

   ```
   device(config)# policy-map pmap1
   ```

3. Exit the policy-map configuration mode.

   ```
   device(config-policymap)# exit
   ```

4. Return to privileged EXEC mode.

   ```
   device(config)# end
   ```

Configuring QoS policer action

Add color-based priority CoS mapping by configuring a policer priority-map. A policer priority-map remaps frame class of service CoS values (802.1p priority bits in VLAN tag) to conform or exceed color values when rates conform to or exceed limits set in a classification map.

The policer priority-map re-marks CoS values according to color-based green (conform), yellow (exceed), and red (violate) priorities. Creating a policer priority-map is optional. If you do not define priority mapping for a color, the map defaults to priorities of 0, 1, 2, 3, 4, 5, 6, and 7 (in other words, nothing is modified). You can configure a maximum of 32 priority-maps (one is reserved as the default), but only one map can be associated with a policer.

> **NOTE**
> You can set a priority-map when creating a policy-map by using appropriate policer attributes.

1. Enter global configuration mode.

   ```
   device# configure terminal
   ```

2. Create a priority-map by providing a priority-map name. This enables police priority-map configuration mode.

   ```
   device(config)# police-priority-map pmap1
   ```

   The name for the priority-map (in this case pmap1) can be a character string up to 64 characters.

3. Create color-based priority mapping. The following example sets the CoS for traffic that conforms to the CIR set in the policy-map.

   ```
   device(config-policepmap)# conform 0 1 1 2 1 2 2 1 1
   ```

   The following example sets the CoS for traffic that exceeds the CIR setting, but conforms to the EIR set in the policy-map.

   ```
   device(config-policepmap)# exceed 3 3 3 3 4 5 6 7
   ```

4. Return to global configuration mode with the **exit** command.

   ```
   device(config-policepmap)# exit
   device(config)#
   ```

5. Configure a policy-map by providing a policy-map name. This enables policy-map configuration mode.

   ```
   device(config)# policy-map pmap1
   ```

6. Configure a class-map in the policy-map by providing the class-map name. This enables policy class-map configuration mode. Note that the class-map name in the following example matches the name provided when you create the class-map by using the **class-map** command (refer to Configuring a class-map on page 41).

   ```
   device(config-policymap)# class class1
   ```

7. Set QoS and policing parameters for the class-map as shown in the following example. For information on all of the optional parameters for this command, refer to the *Extreme Network OS Command Reference*.

   ```
   device(config-policymap)# police cir 40000 cbs 5000 eir 40000 ebs 3000 set-priority pmap1 conform-
   set-dscp 61 conform-set-tc 7 exceed- set-dscp 63 exceed-set-tc 3
   ```

   The CIR parameter is mandatory for the QoS policer. All other parameters are optional. Note that the parameter for set-priority (pmap1) includes the name for the created priority-map (refer to Configuring QoS policer action). For details on setting QoS and policing parameters, refer to Policing parameters on page 22.

8. Return to privileged EXEC mode with the **end** command.

   ```
   device(config-policymap)# end
   ```

Configuring QoS mutation map actions

You can specify the mutation-map to be used on a port. This can lead to possible contradictions if there are other user-defined classes used in the same policy-map that have a set **cos action** configured. In this case-defined cos takes priority over the **mutation map**.

Perform the following task in global configuration mode.

1. Select the policy-map.

   ```
   switch(config)# policy-map p1
   ```

2. Select the class.

   ```
   switch(config-policymap)# class class-default
   ```

3. Specify the mutation map to be used on the port. Different kinds of mutations can be used depending on the command. For complete information, refer to *Extreme Network OS Command Reference*. The available commands are:

   - dscp-mutation
   - dscp-cos
   - dscp-traffic-class

   ```
   switch(config-policyclass)# map cos-mutation plsmap
   ```

Configuring QoS shaping action

You can specify the shaping rate per port attached to the policy-map. You can use this command to smooth out the traffic that egresses an interface. This command is allowed only for the egress direction.

NOTE
The minimum shaping speed on a VDX 6740is 200,000 Kbps.

Perform the following task in global configuration mode.

1. Select the policy-map.

   ```
   switch(config)# policy-map p1
   ```

2. Select the class.

   ```
   switch(config-policymap)# class class-default
   ```

3. Specify the shaping rate for the port.

```
switch(config-policyclass)# shape 30000
```

Configuring the QoS scheduling action

You can specify the scheduling attributes along with per TC shape rate. There are total of eight queues on an interface. The number of DWRR queues present depends on the SP_COUNT value. For example, if the SP_COUNT is 2, then there are two strict priority queues and six DWRR queues. This command is allowed only for the egress direction.

Perform the following task in global configuration mode.

1. Select the policy map.

```
switch(config)# policy-map p1
```

2. Select the class.

```
switch(config-policymap)# class class-default
```

3. Specify the scheduling attributes. For complete information, refer to the *Extreme Network OS Command Reference*.

```
switch(config-policyclass)# scheduler strict-priority 3 dwrr 10 10 10 10 60 TC5 35000 TC6 36000 TC7
37000
```

Configuring the sFlow profile action

You can specify the sFlow profile attached to the policy map.

This feature only functions in the the ingress direction. It can be configured both in user-defined class-maps and in the universal class-map "default". If you use the class-map "default", port-based sFlow is enabled.

Perform the following task in global configuration mode.

1. Select the policy-map.

```
device(config)# policy-map p1
```

2. Select the class.

```
device(config-policymap)# class class1
```

3. Specify the sFlow map for the port.

```
device(config-policyclass)# map sflow mysflowmap
```

Configuring the CEE map action

The priority-mapping-table can support features provided by the Cisco Modular Quality of Service (MQC) provisioning mode to bring partial Converged Enhanced Ethernet (CEE) map content into an MQC class. MQC does not allow ingress and egress feature to be present in a same policy-map. By definition, they are two different entities and should be provisioned through two separate policy-maps. However, a CEE map provisions ingress and egress features in a single provisioning command. Because of this conflict, only the following features are inherited from a CEE map.

- Priority-Group Table.
- Priority-Mapping Table.

- Priority Flow Control Configuration.
- Lossless Priority Remapping.
- Fabric Priority Remapping.

NOTE
In Extreme switches, the CEE map scheduler configuration is global. Unless an egress scheduling policy is applied on an interface, the default scheduler is present.

Perform the following task in global configuration mode.

1. Select the policy-map.

```
switch(config)# policy-map p1
```

2. Select the class.

```
switch(config-policymap)# class cee
```

3. Attach the policy-map to the CEE map.

```
switch(config-policyclass)# priority-mapping-table import cee default
```

Configuring flow-based QoS targets

Consider the topics discussed below when configuring the flow-based QoS targets. You may apply one policy-map per interface for inbound traffic direction by using the service-policy command. Flow-based QoS functions only in the ingress direction.

Binding the policy-map to an interface

Use the **service-policy** command to associate a policy-map to an interface to apply policing parameters.

1. Enable the global configuration mode.

```
device# configure terminal
```

2. Specify the Ethernet interface, as in the following 10-gigabit Ethernet example

```
device(config)# interface te 1/1/2
```

3. Bind a policy-map to inress traffic on the interface. The following associates binds policymap1 to outbound traffic on the interface.

```
device(config-if-te-1/1/2)# service-policy in policymap1
```

You can unbind the policy-map by using the **no** keyword.

```
device(config-if-te-1/1/2)# no service-policy in
```

4. Bind a policy-map to inbound traffic on the interface. The following associates binds policymap1 to inbound traffic on the interface.

```
device(config-if-te-1/1/2)# service-policy in policymap1
```

You can unbind the policy-map by using the **no** keyword.

```
device(config-if-te-1/1/2)# no service-policy in
```

5. Return to privileged EXEC mode.

   ```
   device(conf-if-te-1/1/2)# end
   ```

Policer binding rules

Consider the following rules when binding a policy-map to an interface:

- You can bind the same policy-map to multiple interfaces but only one policy per interface per direction is allowed.
- You cannot bind policy-maps to an interface if the policy-map has no class-map associations.

Binding flow-based QoS at the system level

Use the **qos service-policy** command to bind an existing policy-map to a single Rbridge ID or all Rbridge IDs to apply policing parameters to the interfaces in a VCS fabric.

> NOTE
> The class-maps titled "default" and "cee" cannot be bound at the system level.

1. Enable the global configuration mode.

   ```
   device# configure terminal
   ```

2. Bind the policy-map to inbound traffic.

   ```
   device(config)# qos service-policy in pmap1
   ```

3. Bind a policy-map to ingress traffic on the Rbridge ID. The Rbridge ID can be a single ID, a range of IDs, or you may use all to bind the policy-map to all Rbrdge-IDs. The following associates binds policymap1 to inbound traffic on RBridge ID 14 through 18. Only one policy per interface per direction is allowed.

   ```
   device(config-service-policy)# attach rbridge-id add 14-18
   ```

4. Return to privileged EXEC mode.

   ```
   device(config-service-policy)# end
   ```

Displaying flow-based QoS configuration and operational data

Consider the topics discussed below when displaying policing settings and policy-maps.

Displaying policy-maps

In the following example, the **show policymap** command is used to display Policer policies and parameters set for the 10-gigabit Ethernet interface 4/1 inbound traffic.

```
switch(conf-if-te-5/1/33)# do show policy-map interface tengigabitethernet 5/1/33
Ingress Direction :
 Policy-Map pmap1
    Class default
      Police cir 43454
        Stats:
          Operational cir:39944 cbs:6518 eir:0 ebs:0
          Conform Byte:0 Exceed Byte:0 Violate Byte:0
Egress Direction :
 Policy-Map pmap1
    Class default
```

```
        Police cir 43454
          Stats:
            Operational cir:39944 cbs:6518 eir:0 ebs:0
            Conform Byte:0 Exceed Byte:0 Violate Byte:0
```

Entering **show policymap system rbridge-id** displays the policy-map information for the specified Rbridge ID.

```
switch(conf-if-te-2/0/11)# do show policy-map system rbridge-id 14

Ingress Direction :
 Policy-Map pmap1
    Class cmap1
      matches 0 packets
      Police cir 43454
        Stats:
          Operational cir:39944 cbs:6518 eir:0 ebs:0
          Conform Byte:0 Exceed Byte:0 Violate Byte:0
```

Entering **show policymap** without identifying an interface and specify inbound traffic displays policy-maps bound on all switch interfaces.

```
switch(conf-if-te-5/1/33)# do show policy-map
Number of policy maps : 1

Policy-Map pmap1
  Bound To: Te 5/1/33(in), Te 5/1/33(out)
            Rbridges:14,15,16,17,18

switch(conf-if-te-5/1/33)# do show policy-map detail pmap1
Policy-Map pmap1
    Class cmap1
      Police cir 43454

  Bound To: Te 5/1/33(in), Te 5/1/33(in)
            Rbridges:14,15,16,17,18
```

The following example displays the running configured policy-map by means of the **show running-config policy-map** command.

```
switch(conf-if-te-5/1/33)# do show running-config interface tengigabitethernet 5/1/33
interface TenGigabitEthernet 5/1/33
 service-policy in pmap1
 service-policy in pmap1
 fabric isl enable
 fabric trunk enable
 no shutdown
```

The following example displays the running configured service-policy by means of the **show running-config qos** command.

```
switch(conf-if-te-2/0/11)# do show running-config qos service-policy
qos service-policy in pmap1
 attach rbridge-id add 14-18
```

Displaying class-maps

The following example displays the running configured class-map name and configured match attribute by means of the **show running-config class-map** command.

```
switch(config-classmap)# do show running-config class-map
class-map cee
!
class-map class_map1
 match access-group stdacl1
!
class-map default
```

Configuring Auto QoS

Auto QoS automatically classifies traffic based on either a source or a destination IPv4 or IPv6 address. Once the traffic is identified, it is assigned to a separate priority queue. This allows a minimum bandwidth guarantee to be provided to the queue so that the identified traffic is less affected by network traffic congestion than other traffic.

> NOTE
> As this command was created primarily to benefit Network Attached Storage devices, the commands used in the following sections use the term "NAS". However, there is no strict requirement that these nodes be actual NAS devices, as Auto QoS will prioritize the traffic for any set of specified IP addresses.

Auto QoS for NAS

There are four steps to enabling and configuring Auto QoS for NAS:

1. Enable Auto QoS.
2. Set the Auto QoS CoS value.
3. Set the Auto QoS DSCP value.
4. Specify the NAS server IP addresses.

For detailed instructions, refer to Enabling Auto QoS for NAS on page 50.

Auto QoS configuration guidelines

When configuring Auto QoS, the following configuration guidelines should be followed.

- Auto QoS is enabled and disabled globally.
- Auto QoS supports virtual fabrics.
- When Auto QoS is enabled, you can modify the CEE map subject to the following restrictions:
  - You can set the Auto QoS class of service (CoS) to any value other than "fabric" or "fcoe priority".
  - Only one CoS can map to PGID 3.
  - PGID 2 and PGID 3 in the CEE map cannot be deleted.
  - The priority table cannot be deleted. (This sets CoS to be mapped to strict priority PGID.)
  - Strict-priorities cannot be assigned to CoS values, with the exception of strict-priority 15.0 being assigned to cos 7.
  - The Auto QoS CoS value (the PGID mapping) cannot be modified in the priority table.
- Avoid mixing L2 level multitenancy and L3 level multitenacy (VLAN and VRF), as this will cause Auto QoS statistics to not be displayed properly.
- The source and destination server IP addresses identifying NAS traffic are contained in the NAS server IP list. Refer to Specifying NAS server IP addresses for Auto QoS on page 52 and Removing NAS server IP addresses for Auto QoS on page 52 for instructions on adding and removing NAS server IP addresses.
- Enter more specific entries (those using fewer wild card values) first to have statistics display properly. For example , if you enter **nas server-ip** *10.10.0.0/16* followed by **nas server-ip** *10.10.10.0/24*, the number of matched packets will always be zero for the second entry.

Auto QoS restrictions

- Auto QoS can only be activated when the CEE map is set to the default values.
- If long distance ISL is configured, you cannot enable Auto QoS.

- Different NAS traffic types are not distinguished, and all NAS traffic types are treated equally. This means all NAS traffic, whatever the type, share a common bandwidth guarantee. Individual traffic types do not get separate bandwidth guarantees.
- In order to provide a minimum bandwidth guarantee for NAS traffic across switches in the fabric, when Auto QoS is enabled for NAS, the CEE map reserves 20% of the available bandwidth for NAS traffic. After enabling Auto-nas, NAS bandwidth can be modified by modifying CEE-MAP.
- Port-level configuration has higher precedence than the NAS configuration. This means that any interface-specific configuration will override the global configuration.
- The *conform* and *exceed* traffic class values are not set for the NAS traffic class when Auto QoS for NAS is enabled.
- If all QoS Ternary Content-Addressable Memory (TCAM) space is used up already, enabling Auto QoS will not have any impact on NAS traffic.
- The ACL byte counter is not supported on VDX 6740 and VDX 6940 series platforms.
- PFC is always turned OFF for the NAS classified traffic and cannot be made lossless.
- Auto QoS traffic classification is always done in the ingress RBridge. Once the traffic is classified, all other nodes in the cluster automatically provision the default bandwidth guarantee of 20% through their CEE map.
- Automatic Migration of Port Profiles (AMPP) and Auto-NAS cannot be active on the same switch. This means that if the CEE map is part of AMPP port profile, Auto QoS cannot be enabled, and if Auto QoS is enabled and the CEE map is then associated to AMPP port profile, Auto QoS cannot be disabled. In order to disable Auto QoS, you must:
  1. Disassociate the CEE map from the port profile.
  2. Disable Auto QoS.
  3. Associate the CEE map with the AMPP port profile.

Auto QoS and CEE maps

In order to provide minimum bandwidth guarantee for NAS traffic across switches in the fabric, the default CEE map has to be changed to accommodate NAS traffic. By default, the bandwidth division in the CEE map is 40% for FCoE traffic and 60% for LAN traffic. As NAS traffic has to co-exist with FCoE traffic, some of the LAN traffic bandwidth allocation is given over to NAS traffic. When Auto QoS is enabled, the default bandwidth allocations are 40% for FCoE traffic, 20% for NAS traffic and 40% for LAN traffic. The purpose of using a CEE map is to pass along the scheduler weights to ISLs in the fabric so that they will treat the NAS traffic accordingly.

When Auto QoS is enabled, the modified CEE map will be similar to the following:

```
switch# show cee maps

CEE Map 'default'
   Precedence: 1
   Remap Fabric-Priority to Priority 0
   Remap Lossless-Priority to Priority 0
   Priority Group Table
     1:  Weight 40, PFC Enabled, BW% 40
     2:  Weight 40, PFC Disabled, BW% 40
     3:  Weight 20, PFC Disabled, BW% 20
     15.0: PFC Disabled
     15.1: PFC Disabled
     15.2: PFC Disabled
     15.3: PFC Disabled
     15.4: PFC Disabled
     15.5: PFC Disabled
     15.6: PFC Disabled
     15.7: PFC Disabled

     Priority Table
     CoS:     0    1    2    3    4    5    6   7
     ---------------------------------------------
     PGID:    2    2    3    1    2    2    2   15.0
```

In the example above a PGID with an ID of "3" is defined with a bandwidth allocation of 20 which is then mapped to a user-configured CoS value. If the CoS value is not configured, the PGID value is mapped to the default QoS CoS value (2) in the priority table. This shows that when this CEE map is applied to an interface, the CoS will be allotted 20% of the overall bandwidth.

Enabling Auto QoS for NAS

Auto QoS (Quality of Service) for NAS creates a minimum bandwidth guarantee for Network Attached Storage traffic. Auto QoS for NAS is disabled by default; you must enable Auto QoS to allow tagged NAS packets to have the correct service levels.

All steps must be performed in route-map configuration mode.

The **cee-map** priority group and priority-map settings must be their default values.

Enabling Auto QoS for NAS:

- Changes the CoS value of tagged NAS packets to 2
- Reduces the weight of PGID 2 from 60 to 40
- Creates a new PGID 3 with a weight of 20
- Modifies the priority table to include PGID 3 for the user-configured NAS CoS, or the default NAS CoS if the CoS has not been otherwise modified

Use the following procedure to enable Auto QoS for NAS traffic.

1. Enable Auto QoS for all NAS traffic by entering **nas auto-qos**.

   ```
   switch(config)# nas auto-qos
   ```

2. Set the Class of Service for all NAS traffic by entering: **set cos** *cos_value*.

   The CoS value affects how Auto QoS operated by specifying the User-Priority field value and traffic-class value in the VLAN packet header. If you do not specify a CoS value, the NAS CoS value is set to the default of 2.

   This example sets the CoS value to 3:

   ```
   switch(config)# set cos 3
   ```

3. Set the DSCP value for all NAS traffic by entering **set dscp** *dscp_value*.

   The Differentiated Services Code Point (DSCP) value affects how Auto QoS operates by specifying the priority value for Network Attached Storage traffic on IP networks. If you do not specify a DSCP value, the DSCP value is set to the default of 0. Higher numbers provide a higher level of priority.

   The following example sets the DSCP value to 56:

   ```
   switch(config)# set dscp 56
   ```

4. Identify the IPv4 network addresses (either origination or destination) used by the NAS devices by entering **nas server-ip** followed by the IP address including the mask and then one of the following:

   - **vlan** *vlan_ID*
   - **vrf** *vrf_Name*

     NOTE
     Associating both a VRF and a VLAN value to the same server IP address is strongly discouraged, as this will cause errors in reporting NAS statistics.

5. Press **Enter** after you add each address entry.

   The following example adds two addresses, one using a VLAN mask, and the other a VRF mask.

   ```
   switch(config)# nas server-ip 10.192.100.100/32 vlan 100
   switch(config)# nas server-ip 10.192.100.101/32 vrf bruce
   ```

Disabling Auto QoS for NAS

Disabling Auto QoS (Quality of Service) for NAS removes the minimum bandwidth guarantee for Network Attached Storage traffic.

Disabling Auto QoS for NAS:

- Disables Auto QoS functionality for NAS
- Restores the default bandwidth settings if you have not made any changes to the CEE map after enabling Auto QoS. If you have made changes, the portion of the bandwidth you assigned to NAS traffic will revert to the LAN bandwidth.
- Replaces PGID 3 with PGID 2 in the priority table
- Deletes PGID 3
- Increases the weight of PGID 2 by the weight of PGID 3, so that the weight of PGID 2 equals the weight of PGID 2 plus the weight of PGID 3

Use the following procedure to disable Auto QoS for NAS traffic and restore the default CoS and DSCP values.

1. Enter **no nas auto-qos** in configuration mode.

   ```
   switch(config)# no nas auto-qos
   ```

2. Enter **no set cos** in route-map configuration mode to disable CoS for NAS and restore the default value.

   ```
   switch(config)# no set cos
   ```

3. Enter **no set dscp** in route map configuration mode to disable DSCP for NAS and restore the default value.

   ```
   switch(config)# no set dscp
   ```

Displaying Auto-NAS configurations

Network OS allows you to display Network Attached Storage server configurations for all NAS servers.

Use the following command to display the Auto-NAS configuration for a switch.

Enter **show system internal nas**.

The following example shows a typical output of this command, showing that Auto-NAS is enabled on two IP address (one using VLAN, and one using VRF), that it has the values of CoS 2 and DSCP 0, and a Traffic Class of 5.

```
switch# show system internal nas
Auto-NAS Enabled
Cos 2
Dscp 0
Traffic Class 5
NAS server-ip 10.192.100.100/32 vlan 100
NAS server-ip 10.192.100.101/32 vrf broceliande
```

Specifying NAS server IP addresses for Auto QoS

To enable Auto QoS for NAS, you must identify the IPv4 network addresses used by the NAS devices and tell the platform to tag all packets with this network address as NAS traffic so that they have the correct priority.

The **nas server-ip** command can accept a single IPv4 address (10.192.100.100/32), or an entire sub-net (10.192.100.0/24 vlan 100) as input. You can specify either a Virtual Routing and Forwarding (VRF) ID, or a VLAN ID. If no ID identifier is specified, the default VRF is assumed. When a subnet is specified, all the servers in the sub-net will have Auto QoS enabled for NAS. Subnet masks are not supported (for example: 10.192.100.0 255.255.255.0 is not a supported address).

> NOTE
> IPv6 addressing is also supported for Auto QoS for NAS.

To add a NAS address to the NAS server IP list:

1. In configuration mode, enter **nas server-ip** followed by the IP address with mask and then one of the following:
   - **vlan** *vlan_ID*
   - **vrf** *vrf_Name*
2. Press **Enter** after you add each address entry.

The following example adds two addresses, one using a VLAN mask, and the other a VRF mask.

```
switch(config)# nas server-ip 10.192.100.100/32 vlan 100
switch(config)# nas server-ip 10.192.100.101/32 vrf broceliande
```

Removing NAS server IP addresses for Auto QoS

Removing NAS server IP addresses sets the Auto QoS values for those addresses back to their defaults.

The **no nas server-ip** command can accept a single IPv4 address (10.192.100.100/32), or an entire sub-net (10.192.100.0/24 vlan 100) as input. When a subnet is specified, all the servers in the sub-net will have Auto QoS for NAS removed. Subnet masks are not supported (for example: 10.192.100.0 255.255.255.0 is not a supported address).

To remove an address from the NAS server IP list:

1. In config mode, enter **no nas server-ip** followed by the IPv4 address including the mask and then one of the following:
   - **vlan** *vlan_ID*
   - **vrf** *vrf_Name*
2. Press **Enter** after you add each individual address entry.

The following example removes two addresses, one using a VLAN mask, and the other a VRF mask.

```
switch(config)# no nas server-ip 10.192.100.100/32 vlan 100
switch(config)# no nas server-ip 10.192.100.101/32 vrf broceliande
```

Displaying NAS server IP addresses

How to display the IP addresses for network-attached storage (NAS) servers.

You must be in config mode to run this command.

Use the following command to display all the configured NAS server IPv4 addresses. IPv6 addresses are not supported:

show running-config nas server-ip.

The following example shows the IP addresses of the active NAS servers.

```
switch(config)# show running-config nas server-ip
nas server-ip 10.192.100.100/32 vlan 100
nas server-ip 10.192.100.101/32 vrf broceliande
```

Displaying NAS server statistics

Network OS allows you to display Network Attached Storage (NAS) server statistics for a single server or for all servers.

The **show nas statistics all** command displays the number of incoming IP packets which are NAS-classified for every RBridge in the cluster. It does not show any egressing NAS-classified packets.

Traffic matching the NAS Auto QoS server IP rule works in the same way as matching sequence entries in user-configured ACLs. If the first entry is hit, it will not proceed with any more entry checks.

The order of Auto QoS server IP addresses in the system is not necessarily (or always) same as the order of server IPs programmed in the hardware.

1. To display all NAS server statistics, enter **show nas statistics all**.

   ```
   switch# show nas statistics all
   RBridge 1
   -----------
   nas server-ip 10.1.1.1/32  vrf default-vrf
   matches 0 packets 0 bytes

   RBridge 2
   -----------
   nas server-ip 10.1.1.1/32 vrf default-vrf
   matches 0 packets 0 bytes
   ```

2. To display NAS server statistics for a single server, enter: **show nas statistics server-ip** followed by the IPv4 address and the appropriate VLAN or VRF mask and RBridge ID.

   This command can accept a single IPv4 address (10.192.100.100/32), or an entire sub-net (10.192.100.0/24 vlan 100) as input. When a subnet is specified, all the servers in the sub-net with Auto-NAS QoS will be shown. Subnet masks are not supported (for example: 10.192.100.0 255.255.255.0 is not a supported address). The first following example shows the statistics for all ports using RBridge ID 1; the second example shows the statistics for 10.1.1.0/24 vrf brad.

   ```
   switch# show nas statistics all rbridge-id 1
   RBridge 1
   -----------
   nas server-ip 10.1.1.1/32  vrf default-vrf
   matches 0 packets 0 bytes

   switch# show nas statistics server-ip 10.1.1.0/24 vrf brad
   nas server-ip 10.1.1.0/24 vrf brad
   matches 2000000 packets 40000000 bytes
   ```

Clearing NAS server statistics

You can use this command to clear the statistics for a single IPv4 address (10.192.100.100/32), or an entire sub-net (10.192.100.0/24 vlan 100). When a subnet is specified, the NAS statistics for all the servers in the sub-net will be cleared. Subnet masks are not supported (for example: 10.192.100.0 255.255.255.0 is not a supported address).

1. To clear NAS server statistics for a single server, enter: **show nas statistics server-ip** followed by the IPv4 address and the appropriate VLAN or VRF mask and RBridge ID.

   The following example shows clearing the NAS statistics for "10.1.1.0/24 vrf brad" and then the effect of this clearing.

   ```
   switch# clear nas statistics server-ip 10.1.1.0/24 vrf brad
   switch# show nas statistics server-ip 10.1.1.0/24 vrf brad
   nas server-ip 10.1.1.0/24 vrf brad
   matches 0000000 packets 00000000 bytes
   ```

2. To clear all NAS server statistics, enter **clear nas statistics** *all*.

Rate Limiting and Shaping

BUM storm control

A traffic storm occurs when packets flood the LAN, creating excessive traffic and degrading network performance. Broadcast, unicast, and unknown multicast (BUM) storm control can prevent disruptions on Layer 2 physical ports. This feature is supported only at the interface level.

BUM storm control allows you to limit the amount of broadcast, unknown unicast, and multicast ingress traffic on a specified interface or on the entire system. All traffic received in excess of the configured rate gets discarded. You also have the option to specify whether to shut down an interface if the maximum defined rate is exceeded within a five-second sampling period. When a port is shut down, you receive a log message. You must then manually re-enable the interface by using the **no shut** command.

BUM storm control considerations and limitations

- BUM storm control must be configured on one of the following physical interfaces:
  - 1-gigabit Ethernet
  - 10-gigabit Ethernet
  - 40-gigabit Ethernet
  - 100-gigabit Ethernet
- BUM storm control and input service-policy are mutually exclusive features. Only one can be enabled at a time on a given interface.
- BUM storm control replaces the multicast rate-limit feature for Extreme VDX 6740 series, VDX 6940 series, VDX 8770-4 and VDX 8770-8, and later platforms.

Configuring rate limiting

To create a receive queue multicast rate-limit, perform the following steps from privileged EXEC mode.

1. Enter global configuration mode.

   ```
   device# configure terminal
   ```

2. Create a lower maximum multicast frame expansion rate. In this example, the rate is set to 10000 pps.

   ```
   device(config)# qos rcv-queue multicast rate-limit 10000
   ```

3. Return to privileged EXEC mode.

   ```
   device(config)# end
   ```

Configuring BUM storm control

Refer also to BUM storm control on page 55.

To configure storm control on the 10-gigabit Ethernet interface 101/0/2, with the **broadcast** traffic type and **limit-rate** of 1000000 bps, perform the following steps:

1. Enter global configuration mode.

```
switch# configure terminal
```

2. Specify the Ethernet interface for the traffic you want to control. In the following example, interface 101/0/2 is in *rbridge-id/slot/port* format:

```
switch(config)# interface tengigabitethernet 101/0/2
```

3. Issue the **storm-control ingress** command to set a traffic limit for broadcast traffic on the interface:

```
switch(conf-if-te-101/0/2)# storm-control ingress broadcast 1000000
```

4. Verify the storm control verification with the **show storm-control** command.

```
switch(conf-if-te-101/0/2)# do show storm-control
Interface  Type            rate (Mbps) conformed   violated    total
Te102/4/1  broadcast       100,000     12500000000 12500000000 25000000000
Te102/4/1  unknown-unicast 100,000     12500000000 12500000000 25000000000
Te102/4/1  multicast       100,000     12500000000 12500000000 25000000000
Te102/4/2  broadcast       100,000     12500000000 12500000000 25000000000
Te102/4/3  broadcast       100,000     12500000000 12500000000 25000000000
Te102/4/4  unknown-unicast 100,000     12500000000 12500000000 25000000000
```

NOTE

To deactivate storm control from an interface, enter **no storm-control ingress** followed by the mode (**broadcast**, **unknown-unicast**, or **multicast**) the limit (**limit-bps** or **limit-percent**), **rate**, and optionally either **monitor** or **shutdown**.