



Configuring Queuing and Scheduling

- [About Queuing and Scheduling](#), on page 1
- [Modifying Class Maps](#), on page 1
- [Congestion Avoidance](#), on page 2
- [Congestion Management](#), on page 2
- [Explicit Congestion Notification](#), on page 2
- [Traffic Shaping](#), on page 5
- [Prerequisites for Queuing and Scheduling](#), on page 6
- [Guidelines and Limitations for Queuing and Scheduling](#), on page 6
- [Configuring Queuing and Scheduling](#), on page 11
- [Configuring Congestion Management](#), on page 20
- [Applying a Queuing Policy on a System](#), on page 31
- [Verifying the Queuing and Scheduling Configuration](#), on page 31
- [Controlling the QoS Shared Buffer](#), on page 32
- [Managing Dynamic Buffer Sharing](#), on page 32
- [Monitoring the QoS Packet Buffer](#), on page 33
- [Configuration Examples for Queuing and Scheduling](#), on page 34

About Queuing and Scheduling

Traffic queuing is the ordering of packets and applies to both input and output of data. Device modules can support multiple queues, which you can use to control the sequencing of packets in different traffic classes. You can also set weighted random early detection (WRED) and taildrop thresholds. The device drops packets only when the configured thresholds are exceeded.

Traffic scheduling is the methodical output of packets at a desired frequency to accomplish a consistent flow of traffic. You can apply traffic scheduling to different traffic classes to weight the traffic by priority.

The queuing and scheduling processes allow you to control the bandwidth that is allocated to the traffic classes so that you achieve the desired trade-off between throughput and latency for your network.

Modifying Class Maps

System-defined queuing class maps are provided.



Note The provided system-defined queuing class maps cannot be modified.

Congestion Avoidance

You can use the following methods to proactively avoid traffic congestion on the device:

- Apply WRED to TCP or non-TCP traffic.
- Apply tail drop to TCP or non-TCP traffic.

Congestion Management

For egress packets, you can choose one of the following congestion management methods:

- Specify a bandwidth that allocates a minimum data rate to a queue.
- Impose a minimum and maximum data rate on a class of traffic so that excess packets are retained in a queue to shape the output rate.
- Allocate all data for a class of traffic to a priority queue. The device distributes the remaining bandwidth among the other queues.

For information about configuring congestion management, see the [Configuring WRED on Egress Queues](#) section.

Explicit Congestion Notification

ECN is an extension to WRED that marks packets instead of dropping them when the average queue length exceeds a specific threshold value. When configured with the WRED ECN feature, routers and end hosts use this marking as a signal that the network is congested to slow down sending packets.



Note The ECN feature is not supported on the Cisco Nexus 9508 switch (NX-OS 7.0(3)F3(3)).



Note Enabling WRED and ECN on a class on a network-qos policy implies that WRED and ECN is enabled for all ports in the system.



Note On extended output queues (EOQ), the approximate fair-drop (AFD) feature for bandwidth management is always enabled. The WRED configuration is ignored on EOQs. The configuration for EOQs is based on the system queuing policy and not on the per port policy.

Approximate Fair Drop



Note Approximate Fair Drop is not supported on the Cisco Nexus 9508 switch (Cisco NX-OS Release NX-OS 7.0(3)F3(3)).

Approximate Fair Drop (AFD) is an Active Queue Management (AQM) algorithm that acts on long lived large flows (elephant flows) in the case of congestion, and does not impact short flows (mice flows).

When congestion occurs, the AFD algorithm maintains the queue occupancy at the configured queue desired value by probabilistically dropping packets from the large elephant flows and not impacting small mice flows.

The probability of dropping packets depends upon the arrival rate calculation of a flow at ingress. This is calculated by Elephant Trap (ETrap).

Explicit Congestion Notification (ECN) can be enabled with AFD on a particular class of traffic to mark the congestion state instead of dropping the packets.

Elephant Trap (ETrap)

The Elephant Trap (ETrap) identifies and hashes flows and forwards the arrival rate per flow to AFD for drop probability computation. When the number of bytes received in a flow exceeds the number of bytes specified by the Elephant trap byte-count-threshold, the flow is considered an elephant flow.

The AFD algorithm is applicable only on the flows that are qualified as elephant flows. Mice flows are protected and are not subject to AFD dropping.

For a flow to continue to be an elephant flow, the configured `bw_threshold` number of bytes has to be received in the configured timer period. Otherwise, the flow is evicted from the ETrap hash table.

The ingress rate of every elephant flow is calculated and forwarded to egress for the AFD algorithm to consume.

When ECN is enabled with AFD, the packets are marked to signal congestion instead of being dropped.

ETrap has three parameters that can be configured:

- Byte-count

Byte-count Is used to identify elephant flows. When number of bytes received in a flow exceeds the number of bytes specified by the byte-count-threshold, the flow is considered an elephant flow. (Default byte-count is ~ 1 MB.)

- Age-period and Bandwidth-threshold

Age-period and Bandwidth-threshold are used together to track the activeness of an elephant flow.

When the average bandwidth during the age-period time is lower than the configured bandwidth-threshold, an elephant flow is considered inactive and is timed-out and removed from the elephant flow table. (Default age-period is 50 μ sec. Default bandwidth-threshold is 500 bytes.)

Example:

```
switch (config)# hardware qos etrap age-period 50 usec
switch (config)# hardware qos etrap bandwidth-threshold 500 bytes
switch (config)# hardware qos etrap byte-count 1048555
```

AFD User Profiles

Three user profiles are provided with AFD:

- Mesh (Aggressive)

AFD and ETRAP timers are set to be aggressive, so that the queue depth does not grow much and is kept close to the queue-desired value.

- Burst (Default)

AFD and ETRAP timers are neither aggressive nor conservative, so that the queue depth could be observed to be hovering near the queue-desired value.

- Ultra-burst (Conservative)

AFD and ETRAP timers are set to be conservative, so that more bursts are absorbed and fluctuations for queue depth can be observed around the queue-desired value.

These profiles set the ETrap and AFD timers to pre-configured values for different traffic profiles such as, very bursty or not-so bursty traffic. For more configuration flexibility, the ETrap period set by the profile can be overridden by configuring the ETrap age-period with the **hardware qos etrap** command. However, the AFD timer cannot be changed.

The following is an example of configuring the ETrap age-period:

```
switch(config)# hardware qos etrap age-period 50 usec
```

The following are examples of configuring the AFD user profiles:

- Mesh (Aggressive with ETrap age-period: 20 μ sec and AFD period: 10 μ sec)

```
switch(config)# hardware qos afd profile mesh
```

- Burst (Default with ETrap age-period: 50 μ sec and AFD period: 25 μ sec)

```
switch(config)# hardware qos afd profile burst
```

- Ultra-burst (Conservative with ETrap age-period: 100 μ sec and AFD period: 50 μ sec)

```
switch(config)# hardware qos afd profile ultra-burst
```

AFD Guidelines and Limitations

AFD has the following configuration guidelines and limitations:

- Beginning with Cisco NX-OS Release 9.3(3), Cisco Nexus 9300-GX platform switches support the AFD and ETrap features.
- If an AFD policy has already been applied in system QoS and you are configuring two unique AFD queuing policies, you must apply each unique AFD policy on ports on the same slice.

The following is an example of the system error if you do not create and apply an unique AFD policy on the same slice:

```
Eth1/50 1a006200 1 0 40 255 196 -1 1 0 0 <<<slice 1
Eth1/51 1a006400 1 0 32 255 200 -1 0 32 56 <<<slice 0
Eth1/52 1a006600 1 0 64 255 204 -1 1 24 48 <<<slice 1
Eth1/53 1a006800 1 0 20 255 208 -1 0 20 40 <<<slice 0
```

```

switch(config)# interface ethernet 1/50
switch(config-if)# service-policy type queuing output LM-out-40G
switch(config)# interface ethernet 1/51
switch(config-if)#service-policy type queuing output LM-out-100G
switch(config)# interface ethernet 1/52
switch(config-if)# service-policy type queuing output LM-out-100G
Unable to perform the action due to incompatibility: Module 1 returned status "Max
profiles reached for unique values of queue management parameters (alpha, beta,
max-threshold) in AFD config"

```

- If no AFD policy has already been applied in system QoS—then you can configure the same AFD policy on ports on a different slice, or configure different AFD policies on ports in the same slice.



Note You cannot configure an AFD queuing in the System QoS at a later time.

The following is an example of the system error when AFD queuing is already configured in the system:

```

interface Ethernet1/50
  service-policy type queuing output LM-out-40G
interface Ethernet1/51
  service-policy type queuing output LM-out-40G
interface Ethernet1/52
  service-policy type queuing output LM-out-100G
interface Ethernet1/53
  service-policy type queuing output LM-out-100G
interface Ethernet1/54
  service-policy type queuing output LM-out-100G

(config-sys-qos)# service-policy type queuing output LM-out
Unable to perform the action due to incompatibility: Module 1 returned status "Max
profiles reached for unique values of queue management parameters (alpha, beta,
max-threshold) in AFD config"

```

WRED and AFD Differences

Although WRED and AFD are both AQM algorithms, they have different approaches to help manage congestion:

- WRED computes a random drop probability and drops the packets indiscriminately across all the flows in a class of traffic.
- AFD computes drop probability based on the arrival rate of incoming flows, compares it with the computed fair rate, and drops the packets from the elephant flows while not impacting the mice flows.



Note AFD and WRED cannot be applied at the same time. Only one can be used in a system.

Traffic Shaping

Traffic shaping allows you to control the traffic going out of an interface in order to match its flow to the speed of the remote target interface and to ensure that the traffic conforms to policies contracted for it. You

can shape traffic that adheres to a particular profile to meet downstream requirements. Traffic shaping eliminates bottlenecks in topologies with data-rate mismatches.

Traffic shaping regulates and smooths out the packet flow by imposing a maximum traffic rate for each port's egress queue. Packets that exceed the threshold are placed in the queue and are transmitted later. Traffic shaping is similar to traffic policing, but the packets are not dropped. Because packets are buffered, traffic shaping minimizes packet loss (based on the queue length), which provides better traffic behavior for TCP traffic.

Using traffic shaping, you can control access to available bandwidth, ensure that traffic conforms to the policies established for it, and regulate the flow of traffic to avoid congestion that can occur when the egress traffic exceeds the access speed of its remote, target interface. For example, you can control access to the bandwidth when policy dictates that the rate of a given interface should not, on average, exceed a certain rate even though the access rate exceeds the speed.

Queue length thresholds are configured using the WRED configuration.



Note Traffic shaping is not supported on ALE enabled device 40G front panel ports. When traffic shaping is configured for the system level, the setting is ignored and no error message is displayed. When traffic shaping commands are configured for the port level, the setting is rejected and an error message is displayed.

Prerequisites for Queuing and Scheduling

Queuing and scheduling have the following prerequisites:

- You must be familiar with using modular QoS CLI.
- You are logged on to the device.

Guidelines and Limitations for Queuing and Scheduling

Queuing and scheduling have the following configuration guidelines and limitations:



Note For scale information, see the release-specific *Cisco Nexus 9000 Series NX-OS Verified Scalability Guide*.

- The minimum egress shaper granularity is 200 Mbps per queue for Cisco Nexus 9300-GX2/HX platform switches and line cards.
- **show** commands with the **internal** keyword are not supported.
- PVLANS do not provide support for PVLAN QoS.
- The device supports a system-level queuing policy, so all ports in the system are impacted when you configure the queuing policy.
- A type queuing policy can be attached to the system or to individual interfaces for input or output traffic.

- Changes are disruptive. The traffic passing through ports of the specified port type experience a brief period of traffic loss. All ports of the specified type are affected.
- Performance can be impacted. If one or more ports of the specified type do not have a queuing policy applied that defines the behavior for the new queue, the traffic mapping to that queue can experience performance degradation.
- When there is a link flap on a port with active traffic, it results in a packet/traffic loss flowing through other ports on the same or different slices. To avoid the flow discards, make sure you reduce the queue limit from the default value to a lower value and apply it at the system level.
- Traffic shaping can increase the latency of packets due to queuing because it falls back to store-and-forward mode when packets are queued.
- Traffic shaping is not supported on the Cisco Nexus 9300 ALE 40G ports. For more information on ALE 40G uplink ports, see the [Limitations for ALE 40G Uplink Ports on the Cisco Nexus 9000 Series Switches](#).
- When configuring priority for one class map queue (SPQ), configure the priority for QoS group 3. When configuring priority for more than one class map queue, configure the priority on the higher numbered QoS groups. In addition, the QoS groups must be next to each other. For example, if you want to have two SPQs, you have to configure the priority on QoS group 3 and on QoS group 2.
- About queue limits for 100G enabled devices (such as the Cisco Nexus 9300 platform switch with the N9K-M4PC-CFP2 GEM):
 - The maximum dynamic queue-limit alpha value can be greater than 8. However 8 is the maximum alpha value supported. If you configure the alpha value to a value greater than 8, it is overridden and set to the maximum.

No message is issued when the alpha value is overridden.
 - The static queue-limit has a maximum of 20,000 cells. Any value specified greater than the maximum 20,000 cell limit is overridden by the 20,000 cell limit.

No message is issued when the cell limit is overridden.
- 100G enabled devices (such as the Cisco Nexus 9300 Series switch with the N9K-M4PC-CFP2 GEM), the WRED threshold has a maximum of 20,000 cells. Any value specified greater than the maximum 20,000 cell limit is overridden by the 20,000 cell limit.

No message is issued when the cell limit is overridden.
- FEX support for:
 - System input (ingress) level queuing for HIF to NIF traffic.
 - System output (egress) level queuing for NIF to HIF traffic and HIF to HIF traffic.
- Egress queuing feature works only for base ports and not for FEX ports.
- When the switch supported system queuing policy is configured, the FEX uses the default policy.
- The FEX QoS system level queuing policy does not support WRED, queue-limit, shaping, or policing features.
- The FEX QoS system level queuing policy does not support multiple priority levels.

- Assigning a high alpha value on a Cisco Nexus 9200 platform switch uses more than the expected 50% of the available buffer space.

Assigning a lower alpha value (7 or less) assures the usage of the expected 50% of the available buffer space.

- For Cisco Nexus 9200 platform switches, when a static limit is configured on a queue, both the static limit and the dynamic limit are calculated using the dynamic threshold (alpha value).
- Maximum queue occupancy for Leaf Spine Engine (LSE) enabled switches are limited to 64K cells (~13MB).
- For the following Cisco Nexus series switches and line cards, the lowest value that the egress shaper can manage, per queue, is 100 Mbps:
 - Cisco Nexus 9200 platform switches
 - Cisco Nexus 9300-EX/FX/FX2/GX platform switches
 - Cisco Nexus 9700-EX/FX line cards
- Beginning with Cisco NX-OS Release 10.1(2), Scheduling is supported on the N9K-X9624D-R2 and N9K-C9508-FM-R2 platform switches.
- For R2, though different priority levels can be set through CLI, only priority level 1 is supported in queuing policy.
- Beginning with Cisco NX-OS Release 10.3(1)F, Queuing statistics is supported on the Cisco Nexus 9808 platform switches. The queue depth counter per queue is not supported but additional queuing counters on VOQ tail drops are supported.
- Beginning with Cisco NX-OS Release 10.4(1)F, Queuing statistics is supported on the Cisco Nexus 9804 platform switches. The queue depth counter per queue is not supported but additional queuing counters on VOQ tail drops are supported.
- Cisco Nexus 9804 platform switches have the following limitations for Queuing and scheduling support:
 - Eight queues - SPAN and CPU Queues that are overloaded with eight user queues are supported.
 - SP, DWRR, WRED, and ECN are supported. However, the shaper and DWRR accuracy will have 5% variance
 - Maximum shaper and static limit are supported.
 - Micro-Burst Monitoring is not supported.
 - Link Level Flow Control is not supported.
 - Priority Flow Control is not supported.
 - Dynamic queue-limit is not supported.
 - Multicast Queuing statistics is not supported.
- Cisco Nexus 9808 switches support only the eight queue configuration in Queuing and Scheduling policies. Fewer queues can be configured but are not supported.
- Cisco Nexus 9804 switches support only the eight queue configuration in Queuing and Scheduling policies. Fewer queues can be configured but are not supported.

- The **queue-limit** configuration is applicable only in ingress queuing policy on Cisco Nexus 9500 switches with 9600-R/RX line cards.
- The **bandwidth percent** configuration is applicable only in egress queuing policy on Cisco Nexus 9500 switches with 9600-R/RX line cards.
- If granted buffer is not carved out using a custom input queuing policy for a specified group, only global shared buffers are used.
- From Cisco NX-OS Release 10.4(1)F, Queuing and Scheduling policies are supported on Cisco Nexus C9348GCFX3.
- From Cisco NX-OS Release 10.4(1)F, below limitations apply on Cisco Nexus and C9348GC-FX3PH switch:
 - Queuing and Scheduling policies are supported on switch except for the ports 41-48.
 - Configuring WRED on egress queues is not supported on the switch.
- From Cisco NX-OS Release 10.4(2)F, Queuing and Scheduling policies are supported on Cisco Nexus C93108TC-FX3 switch.

Buffer-Boost

The buffer-boost feature enables the line card to use extra buffers. This capability is enabled by default on line cards such as the Cisco Nexus 9564PX.

- The command to enable the buffer-boost feature is **buffer-boost**.
- The command to disable the buffer-boost feature is **no buffer-boost**.

Generally, we recommend not to disable the buffer-boost feature. However, disabling the buffer-boost is necessary when there is a need to port channel two different member ports, from Cisco Nexus 9636PQ based line cards and Cisco Nexus 9564PX based line cards. However, we do not recommend to port channel such a configuration between ACI-capable leaf line cards and NX-OS line cards.



Note Line cards like the Cisco Nexus 9636PQ and similar, do not offer the buffer-boost feature.

Order of Resolution

The following describes the order of resolution for the pause buffer configuration and the queue-limit for a priority-group.

- Pause Buffer Configuration

The pause buffer configuration is resolved in the following order:

- Interface ingress queuing policy (if applied, and pause buffer configuration is specified for that class).
- System ingress queuing policy (if applied, and pause buffer configuration is specified for that class).
- System network-QoS policy (if applied, and pause buffer configuration is specified for that class).
- Default values with regard to the speed of the port.

- Queue-limit for Priority-Group

The queue-limit for a priority-group is resolved in the following order:

- Interface ingress queuing policy (if applied, and queue-limit configuration is specified for that class).
- System ingress queuing policy (if applied, and queue-limit configuration is specified for that class).
- The **hardware qos ing-pg-share** configuration provided value.
- System default value.

Ingress Queuing

The following are notes about ingress queuing:

- No default system ingress queuing policy exists.
- The ingress queuing policy is used to override the specified pause buffer configuration.
- When downgrading to an earlier release of Cisco Nexus 9000 NX-OS, all ingress queuing configurations have to be removed.
- The ingress queuing feature is supported only on platforms where priority flow control is supported.
- Ingress queuing is not supported on devices with 100G ports.
- The ingress queuing policy is supported only at the system level (and not at the interface level) for Cisco Nexus 9508 switches with the Cisco Nexus 9732C-EX line card and Cisco Nexus 93108TC-EX and 93180YC-EX switches.
- The Cisco Nexus 9636C-R and 9636Q-R line cards and the Cisco Nexus 9508-FM-R fabric module (in a Cisco Nexus 9508 switch) support ingress queuing.
- Beginning with Cisco NX-OS Release 10.3(1)F, Ingress queuing is supported on Cisco Nexus 9800 switches.
- The Cisco Nexus 9500 switches with 9600-R/RX line cards supports only burst-mode to utilize the big-buffer provided by hardware.



Note Recommendation is to use the same port speeds at ingress and egress side.

- Deep Buffer - Nexus 9332D-H2R platform switches support Deep buffers for Unicast traffic. Deep buffer allows the user to handle huge bursts of traffic in the switch by providing 8GB additional buffer in-addition to the existing buffers (40MB) in the switch. Deep buffer is enabled by default in the system across all the queues such that any queue has flexibility to occupy those buffers during congestion scenarios. Multicast traffic is not supported on deep buffers.
- On Nexus 9332D-H2R platform switches, there are two special 33, 34 ports, mostly for management traffic. These ports have all the same features as any regular ports except for:
 - MACSEC and PTP and Frequency Synchronization are not supported.
 - Deep buffer is not supported on these ports, due to limitation on the total number of queues.
 - These are lower bandwidth ports.

- These ports operate in store-and-forward mode only.
- These ports don't support shaper Min/Max rate guarantees.
- PFC and no-drop classes is not supported on these ports.
- FC mode is not supported on these ports.

Configuring Queuing and Scheduling

Queuing and scheduling are configured by creating policy maps of type queuing that you apply to an egress interface. You cannot modify system-defined class maps, which are used in policy maps to define the classes of traffic to which you want to apply policies.

System-defined class maps match based on QoS groups that can be customized using a type qos policy. By default, there is no type QoS policy and all traffic matches to qos-group 0. One consequence is that all traffic will hit the system-defined default-class of type network-qos and type queuing (assigns 100% bandwidth to qos-group 0). Since system-defined classes of type queuing and type network-qos are predefined to match based on distinct qos-groups and cannot be modified, the way to ensure that traffic hits a given type queuing/network-qos class is to configure a type qos policy that sets the corresponding qos-group for that traffic. For traffic classified into a system-defined class map matching on a qos-group other than 0, create a type QoS policy that sets the QoS groups. Once the traffic has been mapped, it will be subject to the default type network-qos and type queuing policies that operate on the non-default qos-group X ($X \neq 0$). You may need to further customize those type queuing and type network-qos policies in order to ensure the desired actions (e.g. re-allocate some bandwidth). For more information on setting the qos-group, see "Example of set qos-groups" in the Using Modular QoS CLI chapter.

For information about configuring policy maps and class maps, see the Using Modular QoS CLI chapter.

You can configure the congestion-avoidance features, which include tail drop and WRED, in any queue.

You can configure one of the egress congestion management features, such as priority, traffic shaping, and bandwidth in output queues.



Note WRED is not supported on ALE enabled device front panel 40G uplink ports. When WRED is configured for the system level, the setting is ignored and no error message is displayed. When WRED is configured at the port level, the setting is rejected and an error message displays.

The system-defined policy map, default-out-policy, is attached to all ports to which you do not apply a queuing policy map. The default policy maps cannot be configured.

Configuring Type Queuing Policies

Type queuing policies for egress are used for scheduling and buffering the traffic of a specific system class. A type queuing policy is identified by its QoS group and can be attached to the system or to individual interfaces for input or output traffic.



Note Ingress queuing policy is used to configure pause buffer thresholds. For more details, see the [Priority Flow Control](#) section.

SUMMARY STEPS

1. **configure terminal**
2. **policy-map type queuing** *policy-name*
3. **class type queuing** *class-name*
4. **priority**
5. **no priority**
6. **shape** {*kbps* | *mbps* | *gbps*} *burst size* **min** *minimum bandwidth*
7. **bandwidth percent** *percentage*
8. **no bandwidth percent** *percentage*
9. **priority level** *level*
10. **queue-limit** *queue size* [**dynamic** *dynamic threshold*]

DETAILED STEPS

	Command or Action	Purpose
Step 1	configure terminal	Enters global configuration mode.
Step 2	policy-map type queuing <i>policy-name</i>	Creates a named object that represents a set of policies that are to be applied to a set of traffic classes. Policy-map names can contain alphabetic, hyphen, or underscore characters, are case sensitive, and can be up to 40 characters.
Step 3	class type queuing <i>class-name</i>	Associates a class map with the policy map, and enters configuration mode for the specified system class.
Step 4	priority	Specifies that traffic in this class is mapped to a strict priority queue.
Step 5	no priority	(Optional) Removes the strict priority queuing from the traffic in this class.
Step 6	shape { <i>kbps</i> <i>mbps</i> <i>gbps</i> } <i>burst size</i> min <i>minimum bandwidth</i>	Specifies the burst size and minimum guaranteed bandwidth for this queue.
Step 7	bandwidth percent <i>percentage</i>	Assigns a weight to the class. The class will receive the assigned percentage of interface bandwidth if there are no strict-priority queues. If there are strict-priority queues, however, the strict-priority queues receive their share of the bandwidth first. The remaining bandwidth is shared in a weighted manner among the class configured with a bandwidth percent. For example, if strict-priority queues take 90 percent of the bandwidth, and you configure 75

	Command or Action	Purpose
		percent for a class, the class will receive 75 percent of the remaining 10 percent of the bandwidth. Note Before you can successfully allocate bandwidth to the class, you must first reduce the default bandwidth configuration on class-default and class-fcoe.
Step 8	no bandwidth percent <i>percentage</i>	(Optional) Removes the bandwidth specification from this class.
Step 9	priority level <i>level</i>	(Optional) Specifies the strict priority levels for the Cisco Nexus 9000 Series switches. These levels can be from 1 to 7.
Step 10	queue-limit <i>queue size</i> [dynamic <i>dynamic threshold</i>]	(Optional) Specifies either the static or dynamic shared limit available to the queue for Cisco Nexus 9000 Series switches. The static queue limit defines the fixed size to which the queue can grow. Note The minimum <i>queue size</i> must be at least 50 kilobytes. The dynamic queue limit allows the queue's threshold size to be decided depending on the number of free cells available, in terms of the alpha value. Note Cisco Nexus 9200 Series switches only support a class level dynamic threshold configuration with respect to the alpha value. This means that all ports in a class share the same alpha value. Note Starting from Release 10.4(1)F, enhanced queue limit range is 0 to 9437184. The maximum threshold supported in Nexus 9332D-H2R platform switches is 256MB.

Configuring Congestion Avoidance

You can configure congestion avoidance with tail drop or WRED features. Both features can be used in egress policy maps.



Note WRED and tail drop cannot be configured in the same class.

Configuring Tail Drop on Egress Queues

You can configure tail drop on egress queues by setting thresholds. The device drops any packets that exceed the thresholds. You can specify a threshold based on the queue size or buffer memory that is used by the queue.

SUMMARY STEPS

1. **configure terminal**
2. **hardware qos q-noise percent** *value*
3. **policy-map** [**type queuing**] [**match-first**] [*policy-map-name*]
4. **class type queuing** *class-name*
5. **queue-limit** {*queue-size* [**bytes** | **kbytes** | **mbytes**] | **dynamic** *value*}
6. (Optional) Repeat Steps 3 and 4 to assign tail drop thresholds for other queue classes.
7. **show policy-map** [**type queuing**] [*policy-map-name* | **default-out-policy**]
8. **copy running-config startup-config**

DETAILED STEPS

	Command or Action	Purpose
Step 1	configure terminal Example: <pre>switch# configure terminal switch(config)#</pre>	Enters global configuration mode.
Step 2	hardware qos q-noise percent <i>value</i> Example: <pre>switch(config)# hardware qos q-noise percent 30</pre>	Tunes the random noise parameter. The default value is 20 percent. This command is supported for Cisco Nexus 9200 and 9300-EX Series switches beginning with Cisco NX-OS Release 7.0(3)I4(4).
Step 3	policy-map [type queuing] [match-first] [<i>policy-map-name</i>] Example: <pre>switch(config)# policy-map type queuing shape_queues switch(config-pmap-que)#</pre>	Configures the policy map of type queuing and then enters policy-map mode for the policy-map name you specify. Policy-map names can contain alphabetic, hyphen, or underscore characters, are case sensitive, and can be up to 40 characters.
Step 4	class type queuing <i>class-name</i> Example: <pre>switch(config-pmap-que)# class type queuing c-out-q1 switch(config-pmap-c-que)#</pre>	Configures the class map of type queuing and then enters policy-map class queuing mode. Class queuing names are listed in the previous System-Defined Type queuing Class Maps table.
Step 5	queue-limit { <i>queue-size</i> [bytes kbytes mbytes] dynamic <i>value</i> } Example: <pre>switch(config-pmap-c-que)# queue-limit 1000 mbytes</pre>	Assigns a tail drop threshold based on the queue size in bytes, kilobytes, or megabytes or allows the queue's threshold size to be determined dynamically depending on the number of free cells available. The device drops packets that exceed the specified threshold.

Command or Action	Purpose																																																																													
	<p>The valid values for byte-based queue size are from 1 to 83886080. The valid values for dynamic queue size are from 0 to 10 as follows:</p> <table border="1" data-bbox="901 388 1526 1281"> <thead> <tr> <th rowspan="2">Value of alpha</th> <th colspan="2">Network Forwarding Engine (NFE) enabled switches</th> <th colspan="3">Leaf Spine Engine (LSE) enabled switches</th> </tr> <tr> <th>Definition</th> <th>Max % per queue</th> <th>Definition</th> <th>Max % per queue</th> <th>ASIC value</th> </tr> </thead> <tbody> <tr><td>0</td><td>1/128</td><td>~0.8%</td><td>1/8</td><td>~11%</td><td>0</td></tr> <tr><td>1</td><td>1/64</td><td>~1.5%</td><td>1/4</td><td>~20%</td><td>1</td></tr> <tr><td>2</td><td>1/32</td><td>~3%</td><td>1/2</td><td>~33%</td><td>3</td></tr> <tr><td>3</td><td>1/16</td><td>~6%</td><td>3/4</td><td>~42%</td><td>5</td></tr> <tr><td>4</td><td>1/8</td><td>~11%</td><td>1 1/8</td><td>~53%</td><td>8</td></tr> <tr><td>5</td><td>1/4</td><td>20%</td><td>1 3/4</td><td>~64%</td><td>14</td></tr> <tr><td>6</td><td>1/2</td><td>~33%</td><td>3</td><td>~75%</td><td>16</td></tr> <tr><td>7</td><td>1</td><td>50%</td><td>5</td><td>~83%</td><td>18</td></tr> <tr><td>8</td><td>2</td><td>~66%</td><td>8</td><td>~89%</td><td>21</td></tr> <tr><td>9</td><td>4</td><td>~80%</td><td>14</td><td>~92.5%</td><td>27</td></tr> <tr><td>10</td><td>8</td><td>~89%</td><td>18</td><td>~95%</td><td>31</td></tr> </tbody> </table> <p>For example, if you configure a dynamic queue size of 6, then the alpha value is 1/2. If you configure a dynamic queue size of 7, then the alpha value is 1.</p> <p>To calculate the queue-limit consider the following: $\text{queue-limit} = (\text{alpha}/(1 + \text{alpha})) \times \text{total buffers}$ For example, if you configure a queue-limit with a dynamic queue size of 7, then the queue-limit can grow up to $(1/(1+1)) \times \text{total buffers}$. This means that $\text{queue-limit} = 1/2 \times \text{total buffers}$.</p> <p>Note Although the above calculations determine the maximum queue occupancy, the maximum queue occupancy is limited to 64K cells in all cases for Application Spine Engine (ASE2, ASE3) and Leaf Spine Engine (LSE) enabled switches.</p>	Value of alpha	Network Forwarding Engine (NFE) enabled switches		Leaf Spine Engine (LSE) enabled switches			Definition	Max % per queue	Definition	Max % per queue	ASIC value	0	1/128	~0.8%	1/8	~11%	0	1	1/64	~1.5%	1/4	~20%	1	2	1/32	~3%	1/2	~33%	3	3	1/16	~6%	3/4	~42%	5	4	1/8	~11%	1 1/8	~53%	8	5	1/4	20%	1 3/4	~64%	14	6	1/2	~33%	3	~75%	16	7	1	50%	5	~83%	18	8	2	~66%	8	~89%	21	9	4	~80%	14	~92.5%	27	10	8	~89%	18	~95%	31
Value of alpha	Network Forwarding Engine (NFE) enabled switches		Leaf Spine Engine (LSE) enabled switches																																																																											
	Definition	Max % per queue	Definition	Max % per queue	ASIC value																																																																									
0	1/128	~0.8%	1/8	~11%	0																																																																									
1	1/64	~1.5%	1/4	~20%	1																																																																									
2	1/32	~3%	1/2	~33%	3																																																																									
3	1/16	~6%	3/4	~42%	5																																																																									
4	1/8	~11%	1 1/8	~53%	8																																																																									
5	1/4	20%	1 3/4	~64%	14																																																																									
6	1/2	~33%	3	~75%	16																																																																									
7	1	50%	5	~83%	18																																																																									
8	2	~66%	8	~89%	21																																																																									
9	4	~80%	14	~92.5%	27																																																																									
10	8	~89%	18	~95%	31																																																																									

	Command or Action	Purpose
		Note Setting the threshold on ALE enabled devices is only supported for the system level. It is not supported for the port level.
Step 6	(Optional) Repeat Steps 3 and 4 to assign tail drop thresholds for other queue classes.	
Step 7	show policy-map [type queuing [policy-map-name default-out-policy]] Example: switch(config-pmap-c-que)# show policy-map type queuing shape_queues	(Optional) Displays information about all configured policy maps, all policy maps of type queuing, a selected policy map of type queuing, or the default output queuing policy.
Step 8	copy running-config startup-config Example: switch(config)# copy running-config startup-config	(Optional) Saves the running configuration to the startup configuration.

Configuring WRED on Egress Queues

You can configure WRED on egress queues to set minimum and maximum packet drop thresholds. The frequency of dropped packets increases as the queue size exceeds the minimum threshold. When the maximum threshold is exceeded, all packets for the queue are dropped.



Note WRED and tail drop cannot be configured in the same class.



Note AFD and WRED cannot be applied at the same time. Only one can be used in a system.

SUMMARY STEPS

1. **configure terminal**
2. **policy-map type queuing** {[match-first] *policy-map-name*}
3. **class type queuing** *class-name*
4. **random-detect** [minimum-threshold *min-threshold* {packets | bytes | kbytes | mbytes} maximum-threshold *max-threshold* {packets | bytes | kbytes | mbytes} drop-probability *value* weight *value*] [threshold {burst-optimized | mesh-optimized}] [ecn | non-ecn]
5. (Optional) Repeat Steps 3 and 4 to configure WRED for other queuing classes.
6. (Optional) **congestion-control random-detect forward-nonecn**

DETAILED STEPS

	Command or Action	Purpose
Step 1	<p>configure terminal</p> <p>Example:</p> <pre>switch# configure terminal switch(config)#</pre>	Enters global configuration mode.
Step 2	<p>policy-map type queuing {[match-first] <i>policy-map-name</i>}</p> <p>Example:</p> <pre>switch(config)# policy-map type queuing pl switch(config-pmap-que)#</pre>	Configures the policy map of type queuing and then enters policy-map mode for the policy-map name you specify. Policy-map names can contain alphabetic, hyphen, or underscore characters, are case sensitive, and can be up to 40 characters.
Step 3	<p>class type queuing <i>class-name</i></p> <p>Example:</p> <pre>switch(config-pmap-que)# class type queuing c-out-ql switch(config-pmap-c-que)#</pre>	Configures the class map of type queuing and then enters policy-map class queuing mode. Class queuing names are listed in the previous System-Defined Type queuing Class Maps table.
Step 4	<p>random-detect [minimum-threshold <i>min-threshold</i> {packets bytes kbytes mbytes} maximum-threshold <i>max-threshold</i> {packets bytes kbytes mbytes} drop-probability <i>value weight value</i> [threshold {burst-optimized mesh-optimized}] [ecn non-ecn]</p> <p>Example:</p> <pre>switch(config-pmap-c-que)# random-detect minimum-threshold 10 mbytes maximum-threshold 20 mbytes</pre> <p>Example:</p> <pre>switch(config-pmap-c-que)# random-detect non-ecn minimum-threshold 1000 kbytes maximum-threshold 4000 kbytes drop-probability 100</pre> <pre>switch(config-pmap-c-que)# show queuing interface eth 1/1 grep WRED WRED Drop Pkts 0 WRED Non ECN Drop Pkts 0 switch(config-pmap-c-que)#</pre>	<p>Configures WRED on the specified queuing class. You can specify minimum and maximum thresholds used to drop packets from the queue. You can configure these thresholds by the number of packets, bytes, kilobytes, or megabytes. The minimum and maximum thresholds must be of the same type. The thresholds are from 1 to 52428800.</p> <p>Alternatively, you can specify a threshold that is optimized for burst or mesh traffic, or you can configure WRED to drop packets based on explicit congestion notification (ECN). Beginning with Cisco NX-OS Release 7.0(3)I6(1), the Network Forwarding Engine (NFE) platform supports the non-ecn option to configure drop thresholds for non-ECN flows.</p> <p>Note The minimum-threshold and maximum-threshold parameters are not supported on the Cisco Nexus 9300 platform switches and Cisco Nexus 9564TX and 9564PX line cards.</p> <p>Note Starting from Release 10.4(1)F, WRED and ECN queue limit range is 0 to 52428800. The maximum threshold supported in Nexus 9332D-H2R platform switches is 256MB.</p> <p>When random-detect is configured under policy-map the default thresholds and drop probabilities are as following:</p> <p>a. On newer platforms, the threshold is 0 and then the drop probabilities would be enforced irrespective of buffer utilization.</p>

	Command or Action	Purpose
		<p>b. On older platforms, the threshold is min 100KB, max 120KB.</p> <p>The drop probabilities are consistently 10% and 90% for burst-optimized and mesh-optimized respectively on all platforms</p>
Step 5	(Optional) Repeat Steps 3 and 4 to configure WRED for other queuing classes.	
Step 6	<p>(Optional) congestion-control random-detect forward-nonecn</p> <p>Example:</p> <pre>switch(config-pmap-c-que) # congestion-control random-detect forward-nonecn</pre>	<p>This is a global CLI command. Allows non-ECN-capable traffic to bypass WRED thresholds and grow until the egress queue-limit and tail drops. This command is intended to be used with a WRED+ECN configuration and when the intention is to avoid WRED drops of non-ECN-capable traffic. This option is available beginning with Cisco NX-OS Release 7.0(3)I4(2) and supported only for Cisco Nexus 9200 platform switches, Cisco Nexus 93108TC-EX and 93180YC-EX switches, and Cisco Nexus 9508 switches with the Cisco Nexus 9732C-EX line card.</p> <p>Beginning with Cisco NX-OS Release 7.0(3)I4(5), this feature is supported on Cisco Nexus 9508 switches with the Cisco Nexus 9636PQ line cards and Cisco Nexus 3164Q switches.</p>

Configuring AFD on Egress Queues

AFD can be configured for an egress queuing policy.



Note Approximate Fair Drop is not supported on the Cisco Nexus 9508 switch (NX-OS 7.0(3)F3(3)).



Note Approximate Fair Drop is not supported on the Cisco Nexus 9808 switch (NX-OS 10.3(1)F).



Note AFD and WRED cannot be applied at the same time. Only one can be used in a system.



Note The following are recommended values for **queue-desired** for different port speeds:

Port Speed	Value for Queue
10G	150 kbytes
40G	600 kbytes
100G	1500 kbytes

Values for queue are user configurable.



Note After AFD is configured, you can apply the policy to the system or to an interface as follows:

- System

```
switch(config)# system qos
switch(config-sys-qos)# service-policy type queuing output afd_8q-out
```

- Interface

```
switch(config)# int e1/1
switch(config-if)# service-policy type queuing output afd_8q-out
```

SUMMARY STEPS

1. **configure terminal**
2. **policy-map type queuing afd_8q-out**
3. **class type queuing c-out-8q-q3**
4. **afd queue-desired <number> [bytes | kbytes | mbytes] [ecn]**

DETAILED STEPS

	Command or Action	Purpose
Step 1	configure terminal Example:	Enters global configuration mode.
Step 2	policy-map type queuing afd_8q-out	Configures the policy map of type queuing.
Step 3	class type queuing c-out-8q-q3	Configures the class map of type queuing and then enters policy-map class queuing mode.
Step 4	afd queue-desired <number> [bytes kbytes mbytes] [ecn]	Specifies desired queue.

	Command or Action	Purpose
		Note Starting from Release 10.4(1)F, AFD queue limit range is 0 to 52428800. The maximum threshold supported in Nexus 9332D-H2R platform switches is 256MB.

Example

- Configuring AFD without ECN

```
switch(config)# policy-map type queuing afd_8q-out
switch(config-pmap-que)# class type queuing c-out-8q-q3
switch(config-pmap-c-que)# afd queue-desired 600 kbytes
```

- Configuring AFD with ECN

```
switch(config)# policy-map type queuing afd-ecn_8q-out
switch(config-pmap-que)# class type queuing c-out-8q-q3
switch(config-pmap-c-que)# afd queue-desired 150 kbytes ecn
```

Configuring Congestion Management

You can configure only one of the following congestion management methods in a policy map:

- Allocate a minimum data rate to a queue by using the **bandwidth** and **bandwidth remaining** commands.
- Allocate all data for a class of traffic to a priority queue by using the **priority** command. You can use the **bandwidth remaining** command to distribute remaining traffic among the nonpriority queues. By default, the system evenly distributes the remaining bandwidth among the nonpriority queues.
- Allocate a minimum and maximum data rate to a queue by using the **shape** command.

In addition to the congestion management feature that you choose, you can configure one of the following queue features in each class of a policy map:

- Tail drop thresholds based on the queue size and the queue limit usage. For more information, see [Configuring Tail Drop on Egress Queues, on page 14](#).
- WRED for preferential packet drops. For more information, see the [Configuring WRED on Egress Queues](#) section.



Note WRED is not supported on the Cisco Nexus 9508 switch (NX-OS 7.0(3)F3(3)).

Configuring Bandwidth and Bandwidth Remaining

You can configure the bandwidth and bandwidth remaining on the egress queue to allocate a minimum percentage of the interface bandwidth to a queue.



Note When a guaranteed bandwidth is configured, the priority queue must be disabled in the same policy map.

SUMMARY STEPS

1. **configure terminal**
2. **policy-map type queuing** {[**match-first**] *policy-map-name*}
3. **class type queuing***class-name*
4. Assign a minimum rate of the interface bandwidth or assign the percentage of the bandwidth that remains:
 - Bandwidth percent:


```
bandwidth {percent percent}
```
 - Bandwidth remaining percent:


```
bandwidth remaining percent percent
```
5. (Optional) Repeat Steps 3 and 4 to assign tail drop thresholds for other queue classes.
6. **exit**
7. **show policy-map** [**type queuing** [*policy-map-name* | **default-out-policy**]]
8. **copy running-config startup-config**

DETAILED STEPS

	Command or Action	Purpose
Step 1	configure terminal Example: <pre>switch# configure terminal switch(config)#</pre>	Enters global configuration mode.
Step 2	policy-map type queuing {[match-first] <i>policy-map-name</i> }	Configures the policy map of type queuing and then enters policy-map mode for the policy-map name you specify. Policy-map names can contain alphabetic, hyphen, or underscore characters, are case sensitive, and can be up to 40 characters.
	Example: <pre>switch(config)# policy-map type queuing shape_queues switch(config-pmap-que)#</pre>	
Step 3	class type queuing <i>class-name</i> Example: <pre>switch(config-pmap-que)# class type queuing c-out-q1 switch(config-pmap-c-que)#</pre>	Configures the class map of type queuing and then enters policy-map class queuing mode. Class queuing names are listed in the previous System-Defined Type queuing Class Maps table.
Step 4	Assign a minimum rate of the interface bandwidth or assign the percentage of the bandwidth that remains:	<ul style="list-style-type: none"> • Bandwidth percent:

	Command or Action	Purpose
	<ul style="list-style-type: none"> Bandwidth percent: bandwidth {percent <i>percent</i>} Bandwidth remaining percent: bandwidth remaining percent <i>percent</i> <p>Example:</p> <ul style="list-style-type: none"> Bandwidth percent: <pre>switch(config-pmap-c-que)# bandwidth percent 25</pre> Bandwidth remaining percent: <pre>switch(config-pmap-c-que)# bandwidth remaining percent 25</pre> 	<p>Assigns a minimum rate of the interface bandwidth to an output queue as the percentage of the underlying interface link rate. The range is from 0 to 100.</p> <p>The example shows how to set the bandwidth to a minimum of 25 percent of the underlying link rate.</p> <ul style="list-style-type: none"> Bandwidth remaining percent: Assigns the percentage of the bandwidth that remains to this queue. The range is from 0 to 100. <p>The example shows how to set the bandwidth for this queue to 25 percent of the remaining bandwidth.</p>
Step 5	(Optional) Repeat Steps 3 and 4 to assign tail drop thresholds for other queue classes.	
Step 6	<p>exit</p> <p>Example:</p> <pre>switch(config-cmap-que)# exit switch(config)#</pre>	Exits policy-map queue mode and enters global configuration mode.
Step 7	<p>show policy-map [type queuing [<i>policy-map-name</i> default-out-policy]]</p> <p>Example:</p> <pre>switch(config-pmap-c-que)# show policy-map type queuing shape_queues</pre>	(Optional) Displays information about all configured policy maps, all policy maps of type queuing, a selected policy map of type queuing, or the default output queuing policy.
Step 8	<p>copy running-config startup-config</p> <p>Example:</p> <pre>switch(config)# copy running-config startup-config</pre>	(Optional) Saves the running configuration to the startup configuration.

Configuring Bandwidth and Bandwidth Remaining for FEX

You can configure the bandwidth and bandwidth remaining on the ingress and egress queue to allocate a minimum percentage of the interface bandwidth to a queue.



Note When a guaranteed bandwidth is configured, the priority queue must be disabled in the same policy map.

Before you begin

Before configuring the FEX, enable **feature-set fex**.

SUMMARY STEPS

1. **configure terminal**
2. **policy-map type queuing** {[**match-first**] *policy-map-name*}
3. **class type queuing***class-name*
4. Assign a minimum rate of the interface bandwidth or assign the percentage of the bandwidth that remains:
 - Bandwidth percent:
bandwidth {**percent** *percent*}
 - Bandwidth remaining percent:
bandwidth remaining percent *percent*
5. (Optional) Repeat Steps 3 and 4 to assign tail drop thresholds for other queue classes.
6. **exit**
7. **show policy-map** [**type queuing** [*policy-map-name* | **default-out-policy**]]
8. **copy running-config startup-config**

DETAILED STEPS

	Command or Action	Purpose
Step 1	configure terminal Example: <pre>switch# configure terminal switch(config)#</pre>	Enters global configuration mode.
Step 2	policy-map type queuing {[match-first] <i>policy-map-name</i> } Example: <pre>switch(config)# policy-map type queuing shape_queues switch(config-pmap-que)#</pre>	Configures the policy map of type queuing and then enters policy-map mode for the policy-map name you specify. Policy-map names can contain alphabetic, hyphen, or underscore characters, are case sensitive, and can be up to 40 characters.
Step 3	class type queuing <i>class-name</i> Example: <pre>switch(config-pmap-que)# class type queuing c-out-ql switch(config-pmap-c-que)#</pre>	Configures the class map of type queuing and then enters policy-map class queuing mode. Class queuing names are listed in the previous System-Defined Type queuing Class Maps table.
Step 4	Assign a minimum rate of the interface bandwidth or assign the percentage of the bandwidth that remains: <ul style="list-style-type: none"> • Bandwidth percent: bandwidth {percent <i>percent</i>} • Bandwidth remaining percent: bandwidth remaining percent <i>percent</i> Example: <ul style="list-style-type: none"> • Bandwidth percent: 	<ul style="list-style-type: none"> • Bandwidth percent: Assigns a minimum rate of the interface bandwidth to an output queue as the percentage of the underlying interface link rate. The range is from 0 to 100. The example shows how to set the bandwidth to a minimum of 25 percent of the underlying link rate. • Bandwidth remaining percent: Assigns the percentage of the bandwidth that remains to this queue. The range is from 0 to 100.

	Command or Action	Purpose
	<pre>switch(config-pmap-c-que)# bandwidth percent 25</pre> <ul style="list-style-type: none"> Bandwidth remaining percent: <pre>switch(config-pmap-c-que)# bandwidth remaining percent 25</pre>	The example shows how to set the bandwidth for this queue to 25 percent of the remaining bandwidth.
Step 5	(Optional) Repeat Steps 3 and 4 to assign tail drop thresholds for other queue classes.	
Step 6	<pre>exit</pre> <p>Example:</p> <pre>switch(config-cmap-que)# exit switch(config)#</pre>	Exits policy-map queue mode and enters global configuration mode.
Step 7	<p>show policy-map [type queuing [policy-map-name default-out-policy]]</p> <p>Example:</p> <pre>switch(config-pmap-c-que)# show policy-map type queuing shape_queues</pre>	(Optional) Displays information about all configured policy maps, all policy maps of type queuing, a selected policy map of type queuing, or the default output queuing policy.
Step 8	<p>copy running-config startup-config</p> <p>Example:</p> <pre>switch(config)# copy running-config startup-config</pre>	(Optional) Saves the running configuration to the startup configuration.

Example

This example shows how to configure the interface bandwidth:

```
switch(config)# policy-map type queuing inq
switch(config-pmap-que)# class type queuing c-in-q3
switch(config-pmap-c-que)# bandwidth percent 30
switch(config-pmap-que)# class type queuing c-in-q2
switch(config-pmap-c-que)# bandwidth percent 20
switch(config-pmap-que)# class type queuing c-in-q1
switch(config-pmap-c-que)# bandwidth percent 10
switch(config-pmap-que)# class type queuing c-in-q-default
switch(config-pmap-c-que)# bandwidth percent 40
```

Configuring Priority

If you do not specify the priority, the system-defined egress pq queues behave as normal queues. For information on the system-defined type queuing class maps, see the "Using Modular QoS CLI" section.

You can configure only one level of priority on an egress priority queue. You use the system-defined priority queue class for the type of module to which you want to apply the policy map.

For the nonpriority queues, you can configure how much of the remaining bandwidth to assign to each queue. By default, the device evenly distributes the remaining bandwidth among the nonpriority queues.



Note When a priority queue is configured, the other queues can only use the remaining bandwidth in the same policy map.



Note When configuring priority for one class map queue (SPQ), you need to configure the priority for QoS group 3. When configuring priority for more than one class map queue, you need to configure the priority on the higher numbered QoS groups. In addition, the QoS groups need to be adjacent to each other. For example, if you want to have two SPQs, you have to configure the priority on QoS group 3 and on QoS group 2.

SUMMARY STEPS

1. **configure terminal**
2. **policy-map type queuing** {[**match-first**] *policy-map-name*}
3. **class type queuing** *class-name*
4. **priority** [*level value*]
5. **class type queuing***class-name*
6. **bandwidth remaining percent** *percent*
7. (Optional) Repeat Steps 5 to 6 to assign the remaining bandwidth for the other nonpriority queues.
8. **exit**
9. **show policy-map** [**type queuing** [*policy-map-name* | **default-out-policy**]]
10. **copy running-config startup-config**

DETAILED STEPS

	Command or Action	Purpose
Step 1	configure terminal Example: <pre>switch# configure terminal switch(config)#</pre>	Enters global configuration mode.
Step 2	policy-map type queuing {[match-first] <i>policy-map-name</i> }	Configures the policy map of type queuing and then enters policy-map mode for the policy-map name you specify. Policy-map names can contain alphabetic, hyphen, or underscore characters, are case sensitive, and can be up to 40 characters.
	Example: <pre>switch(config)# policy-map type queuing priority_queue1 switch(config-pmap-que)#</pre>	
Step 3	class type queuing <i>class-name</i> Example: <pre>switch(config-pmap-que)# class type queuing c-out-q1 switch(config-pmap-c-que)#</pre>	Configures the class map of type queuing and then enters policy-map class queuing mode. Class queuing names are listed in the previous System-Defined Type queuing Class Maps table.

	Command or Action	Purpose
Step 4	priority [level <i>value</i>] Example: <code>switch(config-pmap-c-que)# priority</code>	Selects this queue as a priority queue. Only one priority level is supported.
Step 5	class type queuing <i>class-name</i> Example: <code>switch(config-pmap-que)# class type queuing c-out-q2 switch(config-pmap-c-que)#</code>	(Optional) Configures the class map of type queuing and then enters policy-map class queuing mode. Class queuing names are listed in the previous System-Defined Type queuing Class Maps table. Choose a nonpriority queue where you want to configure the remaining bandwidth. By default, the system evenly distributes the remaining bandwidth among the nonpriority queues.
Step 6	bandwidth remaining percent <i>percent</i> Example: <code>switch(config-pmap-c-que)# bandwidth remaining percent 25</code>	(Optional) Assigns the percent of the bandwidth that remains to this queue. The range is from 0 to 100.
Step 7	(Optional) Repeat Steps 5 to 6 to assign the remaining bandwidth for the other nonpriority queues.	
Step 8	exit Example: <code>switch(config-cmap-que)# exit switch(config)#</code>	Exits policy-map queue mode and enters global configuration mode.
Step 9	show policy-map [type queuing [<i>policy-map-name</i> default-out-policy]] Example: <code>switch(config)# show policy-map type queuing priority_queue1</code>	(Optional) Displays information about all configured policy maps, all policy maps of type queuing, a selected policy map of type queuing, or the default output queuing policy.
Step 10	copy running-config startup-config Example: <code>switch(config)# copy running-config startup-config</code>	(Optional) Saves the running configuration to the startup configuration.

Configuring Priority for FEX



Note Priority for FEX is not supported on the Cisco Nexus 9508 switch (NX-OS 7.0(3)F3(3)).

If you do not specify the priority, the system-defined egress pq queues behave as normal queues. For information on the system-defined type queuing class maps, see the "Using Modular QoS CLI" section.

You can configure only one level of priority on an egress priority queue. You use the system-defined priority queue class for the type of module to which you want to apply the policy map.

For the nonpriority queues, you can configure how much of the remaining bandwidth to assign to each queue. By default, the device evenly distributes the remaining bandwidth among the non-priority queues.



Note When a priority queue is configured, the other queues can only use the remaining bandwidth in the same policy map.



Note When configuring priority for one class map queue (SPQ), you need to configure the priority for QoS group 3. When configuring priority for more than one class map queue, you need to configure the priority on the higher numbered QoS groups. In addition, the QoS groups need to be adjacent to each other. For example, if you want to have two SPQs, you have to configure the priority on QoS group 3 and on QoS group 2.

Before you begin

Before configuring the FEX, enable **feature-set fex**.

SUMMARY STEPS

1. **configure terminal**
2. **policy-map type queuing** {[match-first] *policy-map-name*}
3. **class type queuing** *class-name*
4. **priority** [level *value*]
5. **class type queuing** *class-name*
6. **bandwidth remaining percent** *percent*
7. (Optional) Repeat Steps 5 to 6 to assign the remaining bandwidth for the other nonpriority queues.
8. **exit**
9. **show policy-map** [type queuing [*policy-map-name* | **default-out-policy**]]
10. **copy running-config startup-config**

DETAILED STEPS

	Command or Action	Purpose
Step 1	configure terminal Example: <pre>switch# configure terminal switch(config)#</pre>	Enters global configuration mode.
Step 2	policy-map type queuing {[match-first] <i>policy-map-name</i> } Example: <pre>switch(config)# policy-map type queuing priority_queue1 switch(config-pmap-que)#</pre>	Configures the policy map of type queuing and then enters policy-map mode for the policy-map name you specify. Policy-map names can contain alphabetic, hyphen, or underscore characters, are case sensitive, and can be up to 40 characters.

	Command or Action	Purpose
Step 3	class type queuing <i>class-name</i> Example: <pre>switch(config-pmap-que)# class type queuing c-out-q3 switch(config-pmap-c-que)#</pre>	Configures the class map of type queuing and then enters policy-map class queuing mode. Class queuing names are listed in the previous System-Defined Type queuing Class Maps table.
Step 4	priority [level <i>value</i>] Example: <pre>switch(config-pmap-c-que)# priority</pre>	Selects this queue as a priority queue. Only one priority level is supported. Note FEX QoS priority is supported only on the c-out-q3 class map.
Step 5	class type queuing <i>class-name</i> Example: <pre>switch(config-pmap-que)# class type queuing c-out-q3 switch(config-pmap-c-que)#</pre>	(Optional) Configures the class map of type queuing and then enters policy-map class queuing mode. Class queuing names are listed in the previous System-Defined Type queuing Class Maps table. Choose a nonpriority queue where you want to configure the remaining bandwidth. By default, the system evenly distributes the remaining bandwidth among the nonpriority queues.
Step 6	bandwidth remaining percent <i>percent</i> Example: <pre>switch(config-pmap-c-que)# bandwidth remaining percent 25</pre>	(Optional) Assigns the percent of the bandwidth that remains to this queue. The range is from 0 to 100.
Step 7	(Optional) Repeat Steps 5 to 6 to assign the remaining bandwidth for the other nonpriority queues.	
Step 8	exit Example: <pre>switch(config-cmap-que)# exit switch(config)#</pre>	Exits policy-map queue mode and enters global configuration mode.
Step 9	show policy-map [type queuing [<i>policy-map-name</i> default-out-policy]] Example: <pre>switch(config)# show policy-map type queuing priority_queue1</pre>	(Optional) Displays information about all configured policy maps, all policy maps of type queuing, a selected policy map of type queuing, or the default output queuing policy.
Step 10	copy running-config startup-config Example: <pre>switch(config)# copy running-config startup-config</pre>	(Optional) Saves the running configuration to the startup configuration.

Example

This example shows how to configure the level of priority:

```

switch(config)# policy-map type queuing inq_pri
switch(config-pmap-que)# class type queuing c-in-q3
switch(config-pmap-c-que)# priority
switch(config-pmap-que)# class type queuing c-in-q2
switch(config-pmap-c-que)# bandwidth remaining percent 20
switch(config-pmap-que)# class type queuing c-in-q1
switch(config-pmap-c-que)# bandwidth remaining percent 40
switch(config-pmap-que)# class type queuing c-in-q-default
switch(config-pmap-c-que)# bandwidth remaining percent 40

```

Configuring Traffic Shaping

You can configure traffic shaping on an egress queue to impose a minimum and maximum rate on it.



Note Configuring traffic shaping for a queue is independent of priority or bandwidth in the same policy map.



Note The system queuing policy is applied to both internal and front panel ports. When traffic shaping is enabled on the system queuing policy, traffic shaping is also applied to the internal ports. As a best practice, do not enable traffic shaping on the system queuing policy.



Note Traffic shaping is not supported on the Cisco Nexus 9300 40G ports.



Note The lowest value that the egress shaper can manage, per queue, is 100 Mbps on Cisco Nexus 9200 series, 9300-EX/FX/FX2/GX, and 9700-EX/FX switches.

Before you begin

Configure random detection minimum and maximum thresholds for packets.

SUMMARY STEPS

1. **configure terminal**
2. **policy-map type queuing** *{[match-first] policy-map-name}*
3. **class type queuing** *class-name*
4. **shape min value** *{bps | gbps | kbps | mbps | pps}* **max value** *{bps | gbps | kbps | mbps | pps}*
5. (Optional) Repeat Steps 3 and 4 to assign tail drop thresholds for other queue classes.
6. **show policy-map** *[type queuing [policy-map-name | default-out-policy]]*
7. **copy running-config startup-config**

DETAILED STEPS

	Command or Action	Purpose
Step 1	configure terminal Example: <pre>switch# configure terminal switch(config)#</pre>	Enters global configuration mode.
Step 2	policy-map type queuing {[match-first] <i>policy-map-name</i> } Example: <pre>switch(config)# policy-map type queuing shape_queues switch(config-pmap-que)#</pre>	Configures the policy map of type queuing and then enters policy-map mode for the policy-map name you specify. Policy-map names can contain alphabetic, hyphen, or underscore characters, are case sensitive, and can be up to 40 characters.
Step 3	class type queuing <i>class-name</i> Example: <pre>switch(config)# class type queuing c-out-q-default switch(config-pmap-c-que)#</pre>	Configures the class map of type queuing and then enters policy-map class queuing mode. Class queuing names are listed in the previous System-Defined Type queuing Class Maps table.
Step 4	shape min <i>value</i> { bps gbps kbps mbps pps } max <i>value</i> { bps gbps kbps mbps pps } Example: <pre>switch(config-pmap-c-que)# shape min 100 mbps max 150 mbps</pre>	<p>Assigns a minimum and maximum bit rate on an output queue. The default bit rate is in bits per second (bps).</p> <p>The example shows how to shape traffic to a minimum rate of 100 megabits per second (mbps) and a maximum rate of 150 mbps.</p> <p>Note Most scenarios where traffic shaping is needed requires the configuration of only the max shaper value. For instance, if you want traffic shaped and limited to a maximum desired rate, configure the min shaper value as 0 and the max shaper value as the maximum desired rate.</p> <p>You should only configure the min shaper value for specific scenarios where a guaranteed rate is desired. For instance, if you want traffic to have a guaranteed rate, configure the min shaper value as the guaranteed rate and the max value as something greater than guaranteed rate (or the maximum of the port speed rate).</p>
Step 5	(Optional) Repeat Steps 3 and 4 to assign tail drop thresholds for other queue classes.	
Step 6	show policy-map [type queuing [<i>policy-map-name</i> default-out-policy]] Example: <pre>switch(config)# show policy-map type queuing shape_queues</pre>	(Optional) Displays information about all configured policy maps, all policy maps of type queuing, a selected policy map of type queuing, or the default output queuing policy.

	Command or Action	Purpose
Step 7	copy running-config startup-config Example: <pre>switch(config)# copy running-config startup-config</pre>	(Optional) Saves the running configuration to the startup configuration.

Applying a Queuing Policy on a System

You apply a queuing policy globally on a system.

SUMMARY STEPS

1. **configure terminal**
2. **system qos**
3. **service-policy type queuing output** *{policy-map-name | default-out-policy}*

DETAILED STEPS

	Command or Action	Purpose
Step 1	configure terminal Example: <pre>switch# configure terminal switch(config)#</pre>	Enters global configuration mode.
Step 2	system qos Example: <pre>switch (config)# system qos switch (config-sys-qos)#</pre>	Enters system qos mode.
Step 3	service-policy type queuing output <i>{policy-map-name default-out-policy}</i> Example: <pre>switch (config-sys-qos)# service-policy type queuing map1</pre>	<p>Adds the policy map to the input or output packets of system.</p> <p>Note The output keyword specifies that this policy map should be applied to traffic transmitted from an interface.</p> <p>Note To restore the system to the default queuing service policy, use the no form of this command.</p>

Verifying the Queuing and Scheduling Configuration

Use the following commands to verify the queuing and scheduling configuration:

Command	Purpose
show class-map [type queuing <i>class-name</i>]	Displays information about all configured class maps, all class maps of type queuing, or a selected class map of type queuing.
show policy-map [type queuing <i>policy-map-name</i> default-out-policy]	Displays information about all configured policy maps, all policy maps of type queuing, a selected policy map of type queuing, or the default output queuing policy.
show policy-map system	Displays information about all configured policy maps on the system.

Controlling the QoS Shared Buffer

The QoS buffer provides support per port/queue and shared space. You can control the QoS buffer that is shared by all flows by disabling or restricting reservations.

The **hardware qos min-buffer** command is used to control the QoS shared buffer.

hardware qos min-buffer [all default none]	<ul style="list-style-type: none"> • all Current behavior where all reservations are enabled (ON). • default Enables reservations only for qos-group-0. • none Disables reservations for all qos-groups.
--	--

The **show hardware qos min-buffer** command is used to display the current buffer configuration.

Managing Dynamic Buffer Sharing

Beginning with NX-OS 7.0(3)I7(4), dynamic buffer sharing (egress buffering) across slices is configured with the **hardware qos dynamic-buffer-sharing** command. Following the command, you must reload the switch to enable the dynamic buffering.

Buffer sharing is enabled by dynamic bank allocation (1 bank = 4k cells, 1 cell = 416 bytes) and controlled by a global controller (eCPU) that manages the banks being distributed among slices. Dynamic buffer sharing provides six reserved banks (10MB) for each slice and twelve banks for sharing across slices (20MB).



Note Dynamic Buffer Sharing is supported only on Nexus 9300-FX2 platform switches, see [Nexus Switch Platform Support Matrix](#)

Monitoring the QoS Packet Buffer

The Cisco Nexus 9000 Series device has a 12-MB buffer memory that divides into a dedicated per port and dynamic shared memory. Each front-panel port has four unicast queues and four multicast queues in egress. In the scenario of burst or congestion, each egress port consumes buffers from the dynamic shared memory.

You can display the real-time and peak status of the shared buffer per port. All counters are displayed in terms of the number of cells. Each cell is 208 bytes in size. You can also display the global level buffer consumption in terms of consumption and available number of cells.



Note Monitoring the shared buffer on ALE enabled devices is not supported for the port level.



Note In the examples shown in this section, the port numbers are Broadcom ASIC ports.

This example shows how to clear the system buffer maximum cell usage counter:

```
switch# clear counters buffers
Max Cell Usage has been reset successfully
```

This example shows how to set a buffer utilization threshold for a specific module:

```
switch(config)# hardware profile buffer info port-threshold module 1 threshold 10
Port threshold changed successfully
```



Note The buffer threshold feature is not enabled for ports if they have a no-drop class configured (PFC).



Note The configured threshold buffer count is checked every 5 seconds against all the buffers used by that port across all the queues of that port.



Note You can configure the threshold percentage configuration for all modules or for a specific module, which is applied to all ports. The default threshold value is 90% of the switch cell count of shared pool SP-0. This configuration applies to both Ethernet (front panel) and internal (HG) ports.



Note The buffer threshold feature is not supported for ACI capable device ports.

This example shows how to display the interface hardware mappings:

```
switch# show interface hardware-mappings
```

```
Legends:
```

```

SMod - Source Mod. 0 is N/A
Unit - Unit on which port resides. N/A for port channels
HPort - Hardware Port Number or Hardware Trunk Id:
FPort - Fabric facing port number. 255 means N/A
NPort - Front panel port number
VPort - Virtual Port Number. -1 means N/A

```

```

-----
Name          Ifindex  Smod Unit HPort FPort NPort VPort
-----
Eth2/1        1a080000 4    0   13   255  0    -1
Eth2/2        1a080200 4    0   14   255  1    -1
Eth2/3        1a080400 4    0   15   255  2    -1
Eth2/4        1a080600 4    0   16   255  3    -1
Eth2/5        1a080800 4    0   17   255  4    -1
Eth2/6        1a080a00 4    0   18   255  5    -1
Eth2/7        1a080c00 4    0   19   255  6    -1
Eth2/8        1a080e00 4    0   20   255  7    -1
Eth2/9        1a081000 4    0   21   255  8    -1
Eth2/10       1a081200 4    0   22   255  9    -1
Eth2/11       1a081400 4    0   23   255 10    -1
Eth2/12       1a081600 4    0   24   255 11    -1
Eth2/13       1a081800 4    0   25   255 12    -1
Eth2/14       1a081a00 4    0   26   255 13    -1
Eth2/15       1a081c00 4    0   27   255 14    -1
Eth2/16       1a081e00 4    0   28   255 15    -1
Eth2/17       1a082000 4    0   29   255 16    -1
Eth2/18       1a082200 4    0   30   255 17    -1
Eth2/19       1a082400 4    0   31   255 18    -1
Eth2/20       1a082600 4    0   32   255 19    -1
Eth2/21       1a082800 4    0   33   255 20    -1
Eth2/22       1a082a00 4    0   34   255 21    -1
Eth2/23       1a082c00 4    0   35   255 22    -1
Eth2/24       1a082e00 4    0   36   255 23    -1

```

Configuration Examples for Queuing and Scheduling

In this section, you can find examples of configuring queuing and scheduling.



Note The default system classes type queuing match based on qos-group (by default all traffic matches to qos-group 0, and this default queue gets 100% bandwidth). Create a type QoS policy that first sets the qos-group in order to drive the correct matching for the type queuing classes and policies.

Example: Configuring WRED on Egress Queues

The following example shows how to configure the WRED feature on an egress queue:

```

configure terminal
  class-map type queuing match-any c-out-q1
    match qos-group 1
  class-map type queuing match-any c-out-q2
    match qos-group 1

```

```
policy-map type queuing wred
  class type queuing c-out-q1
    random-detect minimum-threshold 10 bytes maximum-threshold 1000 bytes
  class type queuing c-out-q2
    random-detect threshold burst-optimized ecn
```

Example: Configuring Traffic Shaping

The following example shows how to configure traffic shaping using 500 mbps and 1000 mbps for respective classes::

```
configure terminal
  class-map type queuing match-any c-out-q1
    match qos-group 1
  class-map type queuing match-any c-out-q2
    match qos-group 1
policy-map type queuing pqu
  class type queuing c-out-8q-q3
    bandwidth percent 20
    shape min 100 mbps max 500 mbps
  class type queuing c-out-8q-q2
    bandwidth percent 30
    shape min 200 mbps max 1000 mbps
  class type queuing c-out-8q-q-default
    bandwidth percent 50
  class type queuing c-out-8q-q1
    bandwidth percent 0
  class type queuing c-out-8q-q4
    bandwidth percent 0
  class type queuing c-out-8q-q5
    bandwidth percent 0
  class type queuing c-out-8q-q6
    bandwidth percent 0
  class type queuing c-out-8q-q7
    bandwidth percent 0
system qos
  service-policy type queuing output pqu
```

