



CHAPTER 30

Configuring QoS

This chapter describes how to configure the Quality of Service (QoS) features on the Cisco Nexus 5000 Series switch.

This chapter includes the following sections:

- [Information About QoS, page 30-1](#)
- [Configuration Guidelines and Limitations, page 30-5](#)
- [Configuring PFC and LLC, page 30-5](#)
- [Configuring System Classes, page 30-7](#)
- [Configuring Egress Policy, page 30-11](#)

Information About QoS

The Cisco Nexus 5000 Series switch provides Quality of Service (QoS) capabilities such as traffic prioritization and egress bandwidth allocation.

The default QoS configuration on the switch provides lossless service for Fibre Channel and FCoE (Fibre Channel Over Ethernet) traffic and best-effort service for Ethernet traffic. QoS can be configured to provide additional classes of service for Ethernet traffic. Cisco Nexus 5000 Series QoS features are configured using Cisco Modular QoS CLI (MQC).

This section includes the following topics:

- [MQC, page 30-2](#)
- [System Classes, page 30-2](#)
- [Default System Classes, page 30-2](#)
- [Link-Level Flow Control, page 30-3](#)
- [Priority Flow Control, page 30-3](#)
- [MTU, page 30-4](#)
- [Trust Boundaries, page 30-4](#)
- [Egress Policies, page 30-4](#)
- [Policy for Fibre Channel Interface, page 30-5](#)
- [Configuration Guidelines and Limitations, page 30-5](#)

[Send feedback to nx5000-docfeedback@cisco.com](mailto:nx5000-docfeedback@cisco.com)

MQC

The Cisco Modular QoS CLI (MQC) provides a standard set of commands for configuring QoS.

In the Cisco Nexus 5000 Series switch, the default QoS configuration automatically provides lossless service for Fibre Channel and FCoE traffic. The default for standard Ethernet traffic is best-effort service.

You can use MQC to define additional traffic classes and to configure QoS policies on individual interfaces. Configuring a QoS policy with MQC consists of the following steps:

1. Define traffic classes using the **class-map** command.

The class map classifies incoming or outgoing packets based on matching criteria, such as the IEEE 802.1p CoS value. Unicast and multicast packets are classified.

2. Associate policies or actions with each class of traffic using the **policy-map** command.

The policy map defines a set of actions to take on the associated traffic class, such as limiting the bandwidth or dropping packets.

3. Attach policies to MQC targets using the **service-policy** command.

An MQC target is an entity (such as an Ethernet interface or virtual interface) that represents a flow of packets. The service policy associates a policy map with the MQC target, and specifies whether to apply the policy on incoming or outgoing packets. This enables the configuration of interface-specific QoS policies such as policing and bandwidth allocation.

System Classes

The system class is a new type of MQC target. A service policy can associate a policy map with a system class, which enables application of a QoS policy across the whole switch.

Parameters in system classes need to be configured consistently across the switch and the whole network to ensure that packets in a specific traffic class receive the same treatment when they are transported across the network.

If QoS parameters are configured at the interface level, the interface-level policy always takes precedence over system class configuration or defaults.

The following features can be specified for a system class:

- Drop

No drop specifies lossless service for the system class. Drop specifies that tail drop is used when a queue for this system class is full.

- MTU

The system class MTU defines the maximum packet size for any packet classified into the system class. Each system class has a default MTU and the system class MTU is configurable.

- Match CoS value

The match cos value specifies the IEEE 802.1p CoS value to associate with this system class.

Default System Classes

The Cisco Nexus 5000 Series switch provides the following default system classes:

- FCoE system class

Send feedback to nx5000-docfeedback@cisco.com

All Fibre Channel and FCoE control and data traffic is automatically classified into the FCoE system class, which provides no-drop service.

This class is created automatically when the system starts up (the class is named **class-foe** in the CLI). You cannot delete this class, and you can only modify the IEEE 802.1p CoS value to associate with this class.

The switch classifies packets into the FCoE system class as follows:

- FCoE traffic is classified based on EtherType.
- Native Fibre Channel traffic is classified based on the physical interface type.
- Drop system class

By default, all unicast and multicast Ethernet traffic is classified into the default drop system class.

This class is created automatically when the system starts up (the class is named **class-default** in the CLI). You cannot delete this class and you cannot change the CoS value associated with the default class.

There are two reserved system classes for internal system use.

Link-Level Flow Control

The IEEE 802.3x link-level flow control capability allows a congested receiver to communicate the far end to pause its data transmission for a short period of time. The link-level flow control feature applies to all the traffic on the link.

The transmit and receive directions are separately configurable. By default, link-level flow control is disabled for both directions.

On the Cisco Nexus 5000 Series switch, Ethernet interfaces do not auto-detect the link-level flow control capability. You must configure the capability explicitly on the Ethernet interfaces.

On each Ethernet interface, the switch can enable either priority flow control or link-level flow control (but not both).

Priority Flow Control

The priority flow control (PFC) capability allows you to apply pause functionality to specific classes of traffic on a link (instead of all the traffic on the link). PFC applies pause functionality based on the IEEE 802.1p CoS value. When the switch enables PFC, it communicates to the adapter which CoS values to apply the pause.

Ethernet interfaces use PFC to provide lossless service to no-drop system classes. PFC implements Pause frames on a per-class basis and uses the IEEE 802.1p CoS value to identify the classes that require lossless service.

In the switch, each system class has an associated IEEE 802.1p CoS value (assigned by default or configured on the system class). If PFC is enabled, the switch sends the no-drop CoS values to the adapter, which then applies PFC to these CoS values.

The default CoS value for the FCoE system class is 3 and this value is configurable. The CoS value 0 is used for the default drop system class and this value cannot be changed.

If PFC is not enabled on an interface, you can enable IEEE 802.3X link-level pause. By default, link-level pause is disabled.

Send feedback to nx5000-docfeedback@cisco.com

MTU

The Cisco Nexus 5000 Series switch is a Layer 2 switch, and it does not support packet fragmentation. Configuration mismatch of the MTU between ingress and egress interfaces may result in packets being truncated.

When configuring MTU, follow these guidelines:

- MTU is specified per system class. You cannot configure MTU on the interfaces.
- Fibre Channel and FCoE payload MTU is 2112 bytes across the switch. As a result, the rxbufsize for Fibre Channel interfaces is fixed at 2112 bytes. If the Cisco Nexus 5000 Series switch receives an rxbufsize from a peer different than 2112 bytes, it will fail ELP negotiation and not bring the link up.
- The **system jumbomtu** command defines the upper bound of any MTU in the system. System jumbo MTU has a default value of 9216 bytes. The minimum MTU is 2240 bytes and the maximum MTU is 9216 bytes.
- The system class MTU sets the MTU for all packets in the class. The system class MTU cannot be configured larger than the global jumbo MTU.
- The FCoE system class (for Fibre Channel and FCoE traffic) has a default MTU of 2240 bytes. This value cannot be modified.
- The default drop system class has a default MTU of 1500 bytes. You can configure this value.
- The switch does not send an MTU configuration to the network adapter on the server port. You must configure MTU explicitly on the adapters.

Trust Boundaries

The trust boundary is enforced by the incoming interface as follows:

- All Fibre Channel and virtual Fibre Channel interfaces are automatically classified into the FCoE system class.
- By default, all physical and virtual Ethernet interfaces are trusted interfaces. A packet tagged with a 802.1p CoS value is classified into a system class using the value in the packet.
- Any packet not tagged with an 802.1p CoS value is classified into the default drop system class. If the untagged packet is sent over a trunk, it is tagged with a CoS value of zero (which maps to the default Ethernet drop system class).

After the system applies the untagged CoS value, QoS functions the same as if a packet that entered the system tagged with the CoS value.

Egress Policies

You can associate an egress policy map with a physical Ethernet interface, to limit the bandwidth for the specified traffic class or to configure the egress queues.

The bandwidth allocation limit applies to all traffic on the interface (including any FCoE traffic).

Each physical Ethernet interface supports up to eight queues (one for each system class). The queues have the following default configuration:

- Queue zero is configured as a strict priority queue. Control traffic destined for the CPU uses this queue.

Send feedback to nx5000-docfeedback@cisco.com

- FCoE traffic (traffic that maps to the FCoE system class) is assigned a queue. This queue uses WRR scheduling with 50 percent of the bandwidth.
- Standard Ethernet traffic (in the default drop system class) is assigned a queue. This queue uses WRR scheduling with 50 percent of the bandwidth.

If you add a system class, a queue is assigned to the class. You must reconfigure the bandwidth allocation on all affected interfaces. Bandwidth is not dedicated automatically to user-defined system classes.

You can configure an additional strict priority queue. This queue is serviced before all other queues except queue zero (which carries control traffic, not data traffic).

Policy for Fibre Channel Interface

The egress queues are not configurable for each physical Fibre Channel interface. Two queues are available as follows:

- A strict priority queue to serve high-priority control traffic.
- A queue to serve all data traffic and low-priority control traffic.

Configuration Guidelines and Limitations

Switch resources (such as buffers, virtual output queues, and egress queues) are partitioned based on the default and configured system classes. The switch software automatically adjusts the resource allocation to accommodate the configured system classes.

To maintain optimal switch performance, follow these guidelines when configuring system classes and policies:

- If less than four Ethernet classes are defined, up to two of these classes can be configured as no-drop classes. If more than four Ethernet classes are defined, only one of these classes can be configured as a no-drop class. The default drop class is counted as an Ethernet class.
- If priority flow control is enabled on an Ethernet interface, pause will never be applied to traffic with a drop system class. PFC does not apply pause to drop classes and the link-level pause feature is never enabled on an interface with PFC.
- All FCoE traffic on an Ethernet interface is mapped to one no-drop system class. By default, this class is associated with CoS value 3, although you can configure a different value. If you configure standard Ethernet traffic to use the same CoS value as FCoE, the switch does not apply priority flow control to the standard Ethernet traffic. This traffic is mapped to the default drop system class.
- The CoS value 0 is reserved for the default drop system class. This value cannot be mapped to any other class.

Configuring PFC and LLC

Cisco Nexus 5000 Series switches support PFC and LLC on Ethernet interfaces. The Ethernet interface can operate in two different modes: FCoE mode or standard Ethernet mode.

If the interface is operating in FCoE mode, the Ethernet link is connected at the server port using a converged network adapter. Refer to [Chapter 28, “Configuring FCoE”](#) for information about configuring PFC and LLC when the interface is operating in FCoE mode.

Send feedback to nx5000-docfeedback@cisco.com

If the interface is operating in standard Ethernet mode, the Ethernet link is connected at the server port with a standard 10-Gigabit Ethernet NIC.



Note

You must configure a no-drop Ethernet system class for PFC to operate for Ethernet traffic (PFC will be applied to traffic that matches the CoS value configured for this class).

Configuring PFC and LLC for standard Ethernet is covered in the following sections:

- [Configuring Priority Flow Control, page 30-6](#)
- [Configuring IEEE 802.3x Link-Level Flow Control, page 30-6](#)

Configuring Priority Flow Control

By default, Ethernet interfaces negotiate PFC capability with the adapter via DCBX. When PFC is enabled, PFC is applied to traffic that matches the CoS value configured for the no-drop Ethernet class.

You can override the negotiation result by force-enabling the PFC capability. To force-enable the PFC capability, perform this task:

| | Command | Purpose |
|--------|--|---|
| Step 1 | switch# configure terminal | Enters configuration mode. |
| Step 2 | switch(config)# interface <i>type slot/port</i> | Specifies the interface to be changed. |
| Step 3 | switch(config-if)# priority-flow-control mode {auto on} | Sets PFC mode for the selected interface. Specify on to force-enable PFC. Specify auto to negotiate PFC capability. |

The following example shows how to force-enable PFC on an interface:

```
switch# configure terminal
switch(config)# interface ethernet 1/2
switch(config-if)# priority-flow-control mode on
```

To disable PFC capability for this interface, perform this task:

| Command | Purpose |
|---|--|
| switch(config-if)# no priority-flow-control mode | Disables the PFC setting for the selected interface. |

Configuring IEEE 802.3x Link-Level Flow Control

By default, link-level flow control capability on Ethernet interfaces is disabled. You can enable link-level flow-control capability for the transmit and receive directions. To enable link-level flow control capability, perform this task:

Send feedback to nx5000-docfeedback@cisco.com

| | Command | Purpose |
|---------------|--|--|
| Step 1 | switch# configure terminal | Enters configuration mode. |
| Step 2 | switch(config)# interface <i>type slot/port</i> | Selects the interface to change. |
| Step 3 | switch(config-if)# flowcontrol [receive { on off }] [transmit { on off }] | Enables IEEE 802.3x link-level flow control for the selected interface. Set receive and/or transmit on or off . |

The following example shows enabling link-level flow control frames on an interface:

```
switch# configure terminal
switch(config)# interface ethernet 1/2
switch(config-if)# flowcontrol receive on transmit on
```

To disable link-level flow control, perform this task:

| Command | Purpose |
|---|---|
| switch(config-if)# no flowcontrol [receive { on off }] [transmit { on off }] | Disables 802.3x link-level flow control for the selected interface. |

Configuring System Classes

This section describes how to configure system class QoS on the switch. The steps to configure a system class are described in the following sections:

- [Configuring Class Maps, page 30-7](#)
- [Configuring Policy Maps, page 30-8](#)
- [Creating the System Service Policy, page 30-9](#)
- [System Class Example, page 30-9](#)
- [Enabling Jumbo MTU, page 30-10](#)
- [Verifying Jumbo MTU, page 30-10](#)

Configuring Class Maps

The **class-map** command creates a named object that represents a class of traffic. In the class map, you specify a set of match criteria for classifying the packets. For system classes, the only match criteria supported is **match cos**.

If a system class is configured with no-drop function, the **match cos** command serves an additional purpose. The switch sends the CoS value to the adapter, so that the adapter will apply PFC Pause for this CoS value.

The FCoE system class has a default CoS value of 3. You can add a **match cos** configuration to the FCoE system class to set a different CoS value. PFC Pause will be applied to traffic with the new value.

Send feedback to nx5000-docfeedback@cisco.com

To configure class maps, perform this task:

| | Command | Purpose |
|---------------|--|--|
| Step 1 | switch# configure terminal | Enters configuration mode. |
| Step 2 | switch(config)# class-map <i>name</i> | Creates a named object that represents a class of traffic. Class map names can contain alphabetic, hyphen, or underscore characters, are case sensitive, and can be up to 40 characters. |
| Step 3 | switch(config-cmap)# match cos <i>cos-value</i> | Specifies the CoS value to match for classifying packets into this class. You can configure a CoS value in the range of 1 to 7. Note CoS value 0 is reserved for the default drop system class (class-default). |

Configuring Policy Maps

The **policy-map** command is used to create a named object representing a set of policies that are to be applied to a set of traffic classes.

The switch provides two default system classes: a no-drop class for lossless service and a drop class for best-effort service. You can define up to four additional system classes for Ethernet traffic.

You need to create a policy map to contain the user-defined system classes. In the policy-map, you configure the drop behavior and MTU for user-defined classes. Use the same policy map to modify the configuration of the default classes.



Note Before creating the policy map, define a class map for each new system class.

To configure a policy map, perform this task:

| | Command | Purpose |
|---------------|---|--|
| Step 1 | switch# configure terminal | Enters configuration mode. |
| Step 2 | switch(config)# policy-map <i>name</i> | Creates a named object representing 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 | switch(config-pmap)# class <i>class-name</i> | Associates a class map with the policy map. |
| Step 4 | switch(config-pmap-c)# pause no-drop | (Optional) Configures a no-drop class. If you do not specify this subcommand, the default policy is drop. |
| Step 5 | switch(config-pmap-c)# mtu <i>value</i> | (Optional) Specifies the MTU value in bytes. |

The operation for drop policy is simple tail drop, where arriving packets will be dropped if the queue increases to its allocated size.

[Send feedback to nx5000-docfeedback@cisco.com](mailto:nx5000-docfeedback@cisco.com)



Note

The policy map can also be configured for egress policies. However, different parameters are supported for egress policies.

The following example shows how to create a system class with a no-drop policy map:

```
switch(config)# class-map ethCoS4
switch(config-cmap)# match cos 4
switch(config-cmap)# exit
switch(config)# policy-map ethNoDrop
switch(config-pmap)# class ethCoS4
switch(config-pmap-c)# pause no-drop
switch(config)# system qos
switch(config-system)# service-policy ethNoDrop
```

Creating the System Service Policy

The **service-policy** command is used to associate the system class policy-map as the service policy for the system.

| | Command | Purpose |
|--------|--|---|
| Step 1 | switch# configure terminal | Enters configuration mode. |
| Step 1 | switch(config)# system qos | Enters system class configuration mode. |
| Step 2 | switch(config-sys-qos)# service-policy <i>name</i> | Specifies the policy-map to use as the service policy for the system. |

System Class Example

In the following example, a new Ethernet no-drop system class is created, and the CoS values of the default system classes are changed from their default values:

```
switch(config)# class-map trading-data-no-drop
switch(config-cmap)# match cos 5
switch(config)# class-map class-fcoe
switch(config-cmap)# match cos 2
switch(config)# policy-map system-policy
switch(config-pmap)# class trading-data-no-drop
switch(config-pmap-c)# pause no-drop
switch(config-pmap-c)# mtu 2000
switch(config)# system qos
switch(config-system)# service-policy system-policy
```

In this example, the first **class-map** command defines a new Ethernet system class. Packets from all over the system with 802.1p CoS value of 5 will be classified into this new system class.

The second **class-map** command changes the match value of the default no-drop system class.

The **policy-map** command defines the QoS policy for the system. The new Ethernet class is configured as a no-drop class, with an MTU of 2000 bytes.

The **pause no-drop** command causes PFC to apply pause functionality for packets with IEEE 802.1p priority value 5.

[Send feedback to nx5000-docfeedback@cisco.com](mailto:nx5000-docfeedback@cisco.com)

Enabling Jumbo MTU

To enable jumbo MTU for the whole switch, set the MTU to its maximum size (9216 bytes) in the policy map for the default Ethernet system class (class-default).

In the following example, the default Ethernet system class is configured to support the jumbo MTU:

```
switch(config)# policy-map jumbo
switch(config-pmap)# class class-default
switch(config-pmap-c)# mtu 9216
switch(config)# system qos
switch(config-system)# service-policy jumbo
```



Note

The **system jumbomtu** command defines the maximum MTU size for the switch. However, jumbo MTU is only supported for system classes that have **mtu** configured.

Verifying Jumbo MTU

To verify that jumbo MTU is enabled, enter the **show interface ethernet slot/port** command for an Ethernet interface that carries traffic with jumbo MTU.

The following example shows how to display summary jumbo MTU information for Ethernet 2/1 (the relevant part of the output is shown in bold font):

```
switch# show interface ethernet 2/1
Ethernet2/1 is up
...
Rx
1547805598 Input Packets 1547805596 Unicast Packets 0 Multicast Packets
0 Broadcast Packets 1301767362 Jumbo Packets 33690 Storm Suppression Packets
7181776513802 Bytes
Tx
1186564478 Output Packets 7060 Multicast Packets
0 Broadcast Packets 997813205 Jumbo Packets
4813632103603 Bytes
...
```

The following example shows how to display detailed jumbo MTU information for Ethernet 2/1 (the relevant part of the output is shown in bold font):

```
switch# show interface ethernet 2/1 counters detailed
Rx Packets: 1547805598
Rx Unicast Packets: 1547805596
Rx Jumbo Packets: 1301767362
Rx Bytes: 7181776513802
Rx Storm Suppression: 33690
Rx Packets from 0 to 64 bytes: 169219
Rx Packets from 65 to 127 bytes: 10657133
Rx Packets from 128 to 255 bytes: 21644488
Rx Packets from 256 to 511 bytes: 43290596
Rx Packets from 512 to 1023 bytes: 86583071
Rx Packets from 1024 to 1518 bytes: 83693729
Rx Trunk Packets: 1547805596
Tx Packets: 1186564481
Tx Unicast Packets: 1005445334
Tx Multicast Packets: 7063
Tx Jumbo Packets: 997813205
Tx Bytes: 4813632103819
Tx Packets from 0 to 64 bytes: 137912
```

[Send feedback to nx5000-docfeedback@cisco.com](mailto:nx5000-docfeedback@cisco.com)

```
Tx Packets from 65 to 127 bytes: 8288443
Tx Packets from 128 to 255 bytes: 16596457
Tx Packets from 256 to 511 bytes: 33177999
Tx Packets from 512 to 1023 bytes: 66363944
Tx Packets from 1024 to 1518 bytes: 64186521
Tx Trunk Packets: 1005451729
```

Configuring Egress Policy

Bandwidth allocation control is supported for Ethernet interfaces.

| | Command | Purpose |
|--------|--|--|
| Step 1 | switch# configure terminal | Enters configuration mode. |
| Step 2 | switch(config)# class-map class-name | Defines a class name for the egress policy. |
| Step 3 | switch(config)# policy-map policy1-name | Creates a named object representing 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 4 | switch(config-pmap)# class class-name | Creates a named object that represents a class of traffic. |
| Step 5 | switch(config-pmap-c) bandwidth percent percentage | Specifies the guaranteed percentage of bandwidth allocated to this class. |
| Step 6 | switch(config-pmap-c) priority | Specifies that egress traffic in this class is mapped to a strict priority queue. |
| Step 7 | switch(config)# interface ethernet slot/port | Enters configuration mode for the specified interface. |
| Step 8 | switch(config-if)# server-policy output policy-name | Applies the policy map to the interface. |

The following example shows that the system class best-effort-drop-class is guaranteed 20 percent of the bandwidth on interface eth1/1:

```
switch(config)# class-map best-effort-drop-class
switch(config-cmap)# match cos 5
switch(config)# policy-map policy1-egress
switch(config-pmap)# class best-effort-drop-class
switch(config-pmap-c)# bandwidth percent 20
switch(config)# int eth1/1
switch(config-if)# server-policy output policy1-egress
```

Send feedback to nx5000-docfeedback@cisco.com