



CHAPTER 27

Configuring QoS

This chapter describes how to configure quality of service (QoS) by using automatic QoS (auto-QoS) commands or by using standard QoS commands on the Catalyst 2960 switch. With QoS, you can provide preferential treatment to certain types of traffic at the expense of others. Without QoS, the switch offers best-effort service to each packet, regardless of the packet contents or size. It sends the packets without any assurance of reliability, delay bounds, or throughput.

You can configure QoS only on physical ports. Configure the QoS settings, such as classification, queuing, and scheduling, and apply a trusted class of service (CoS) or a CoS override setting.



Note

For complete syntax and usage information for the commands used in this chapter, see the command reference for this release.

This chapter consists of these sections:

- [Understanding QoS, page 27-1](#)
- [Configuring Standard QoS, page 27-8](#)
- [Displaying Standard QoS Information, page 27-18](#)

The switch supports some of the modular QoS CLI (MQC) commands. For more information about the MQC commands, see the “Modular Quality of Service Command-Line Interface Overview” at this site:

http://www.cisco.com/en/US/products/sw/iosswrel/ps1835/products_configuration_guide_chapter09186a00800bd908.html

Understanding QoS

Typically, networks operate on a best-effort delivery basis, which means that all traffic has equal priority and an equal chance of being delivered in a timely manner. When congestion occurs, all traffic has an equal chance of being dropped.

When you configure the QoS feature, you can select specific network traffic, prioritize it according to its relative importance, and use congestion-management and congestion-avoidance techniques to provide preferential treatment. Implementing QoS in your network makes network performance more predictable and bandwidth utilization more effective.

The QoS implementation is based on the Differentiated Services (Diff-Serv) architecture, an emerging standard from the Internet Engineering Task Force (IETF). This architecture specifies that each packet is classified upon entry into the network.

The classification is carried in the IP packet header, using 6 bits from the deprecated IP type of service (ToS) field to carry the classification (*class*) information. Classification can also be carried in the Layer 2 frame. These special bits in the Layer 2 frame are described here and shown in [Figure 27-1](#).

Prioritization bits in Layer 2 frames:

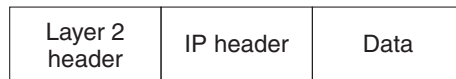
- Layer 2 Inter-Switch Link (ISL) frame headers have a 1-byte User field that carries an IEEE 802.1p CoS value in the three least-significant bits. On ports configured as Layer 2 ISL trunks, all traffic is in ISL frames.
- Layer 2 IEEE 802.1Q frame headers have a 2-byte Tag Control Information field that carries the CoS value in the three most-significant bits, which are called the User Priority bits. On ports configured as Layer 2 IEEE 802.1Q trunks, all traffic is in IEEE 802.1Q frames except for traffic in the native VLAN.
- Other frame types cannot carry Layer 2 CoS values.
- Layer 2 CoS values range from 0 for low priority to 7 for high priority.



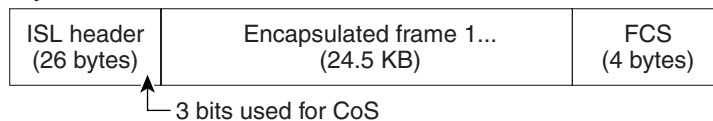
Note IPv6 QoS is not supported in this release.

Figure 27-1 QoS Classification Layers in Frames and Packets

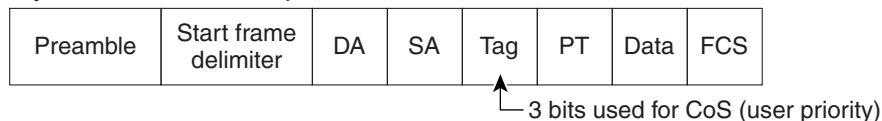
Encapsulated Packet



Layer 2 ISL Frame



Layer 2 802.1Q and 802.1p Frame



All switches and routers that access the Internet rely on the class information to provide the same forwarding treatment to packets with the same class information and different treatment to packets with different class information. The class information in the packet can be assigned by end hosts or by switches or routers along the way, based on a configured policy, detailed examination of the packet, or both. Detailed examination of the packet is expected to happen closer to the edge of the network so that the core switches and routers are not overloaded with this task.

Switches and routers along the path can use the class information to limit the amount of resources allocated per traffic class. The behavior of an individual device when handling traffic in the DiffServ architecture is called per-hop behavior. If all devices along a path provide a consistent per-hop behavior, you can construct an end-to-end QoS solution.

Implementing QoS in your network can be a simple or complex task and depends on the QoS features offered by your internetworking devices, the traffic types and patterns in your network, and the granularity of control that you need over incoming and outgoing traffic.

Basic QoS Model

To implement QoS, the switch must distinguish packets or flow from one another (classify), assign a label to indicate the given quality of service as the packets move through the switch, make the packets comply with the configured resource usage limits (police and mark), and provide different treatment (queue and schedule) in all situations where resource contention exists. The switch also needs to ensure that traffic sent from it meets a specific traffic profile (shape).

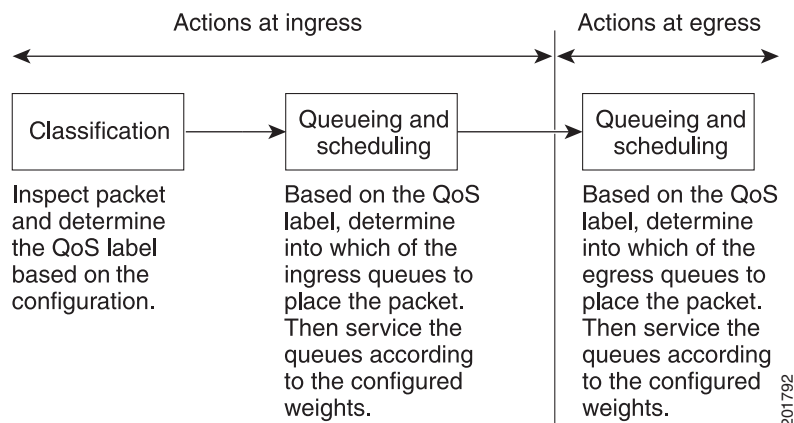
Figure 27-2 shows the basic QoS model. Actions at the ingress port include classifying traffic and queuing:

- Classifying a distinct path for a packet by associating it with a QoS label. The switch maps the CoS in the packet to a QoS label to distinguish one kind of traffic from another. The QoS label that is generated identifies all future QoS actions to be performed on this packet. For more information, see the “Classification” section on page 27-3.
- Queuing evaluates the QoS label and the corresponding CoS value to select into which of the two ingress queues to place a packet. Queuing is enhanced with the weighted tail-drop (WTD) algorithm, a congestion-avoidance mechanism. If the threshold is exceeded, the packet is dropped. For more information, see the “Queuing Overview” section on page 27-4.

The action at the egress port is queuing:

Queuing evaluates the QoS packet label and the corresponding CoS value before selecting which of the four egress queues to use. Because congestion can occur when multiple ingress ports simultaneously send data to an egress port, WTD differentiates traffic classes and subjects the packets to different thresholds based on the QoS label. If the threshold is exceeded, the packet is dropped. For more information, see the “Queuing Overview” section on page 27-4.

Figure 27-2 Basic QoS Model



Classification

Classification is the process of distinguishing one kind of traffic from another by examining the fields in the packet. Classification is enabled only if QoS is globally enabled on the switch. By default, QoS is globally disabled, so no classification occurs.

During classification, the switch performs a lookup and assigns a QoS label to the packet. The QoS label identifies all QoS actions to be performed on the packet and from which queue the packet is sent.

The QoS label is based on the CoS value in the packet and decides the queueing and scheduling actions to perform on the packet. The label is mapped according to the trust setting and the packet type.

You specify which fields in the frame or packet that you want to use to classify incoming traffic. For non-IP traffic, you have this classification option:

Trust the CoS value in the incoming frame (configure the port to trust CoS). Layer 2 ISL frame headers carry the CoS value in the 3 least-significant bits of the 1-byte User field. Layer 2 IEEE 802.1Q frame headers carry the CoS value in the 3 most-significant bits of the Tag Control Information field. CoS values range from 0 for low priority to 7 for high priority.

For configuration information on port trust states, see the [“Configuring Classification Using Port Trust States” section on page 27-11](#).

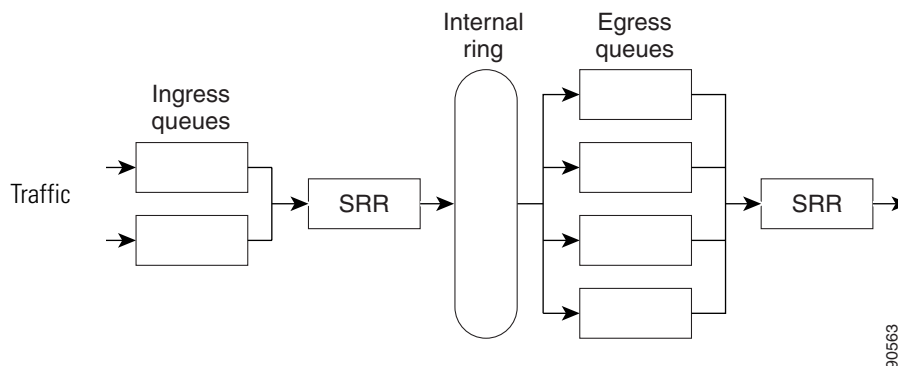
After classification, the packet is sent to the ingress queueing and scheduling stages.

For information about the CoS ingress queue threshold maps, see the [“Queueing on Ingress Queues” section on page 27-5](#). For information about the CoS egress queue threshold maps, see the [“Queueing on Egress Queues” section on page 27-7](#).

Queueing Overview

The switch has queues at specific points to help prevent congestion, as shown in [Figure 27-3](#).

Figure 27-3 Ingress and Egress Queue Location



Because the total inbound bandwidth of all ports can exceed the bandwidth of the internal ring, ingress queues are located after the packet is classified, policed, and marked and before packets are forwarded into the switch fabric. Because multiple ingress ports can simultaneously send packets to an egress port and cause congestion, outbound queues are located after the internal ring.

Weighted Tail Drop

Both the ingress and egress queues use an enhanced version of the tail-drop congestion-avoidance mechanism called weighted tail drop (WTD). WTD is implemented on queues to manage the queue lengths and to provide drop precedences for different traffic classifications.

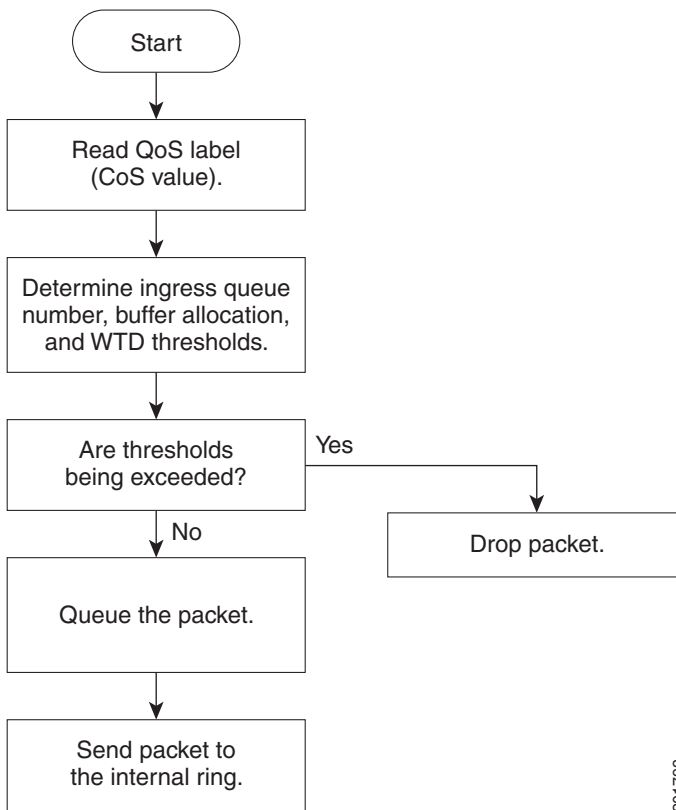
As a frame is enqueued to a particular queue, WTD uses the assigned QoS label for the frame to subject it to different thresholds. If the threshold is exceeded for that QoS label (the space available in the destination queue is less than the size of the frame), the switch drops the frame.

For more information, see the [“Mapping CoS Values to an Ingress Queue”](#) section on page 27-14 and the [“Mapping CoS Values to an Egress Queue and to a Threshold ID”](#) section on page 27-17.

Queueing on Ingress Queues

Figure 27-4 shows the queueing and scheduling flowchart for ingress ports.

Figure 27-4 Queueing Flowchart for Ingress Ports



Note

Shaped round robin (SRR) services the priority queue for its configured share before servicing the other queue.

The switch supports two configurable ingress queues, which are serviced by SRR only in shared mode. [Table 27-1](#) describes the queues.

Table 27-1 *Ingress Queue Types*

Queue Type ¹	Function
Normal	User traffic that is considered to be normal priority. You can configure three different thresholds to differentiate among the flows. You can use the mls qos srr-queue input cos-map global configuration command.
Expedite	High-priority user traffic such as differentiated services expedited forwarding or voice traffic. You can configure the bandwidth required for this traffic as a percentage of the total traffic. You can configure the bandwidth required for this traffic as a percentage of the total traffic by using the mls qos srr-queue input priority-queue global configuration command. The expedite queue has guaranteed bandwidth.

1. The switch uses two nonconfigurable queues for traffic that is essential for proper network operation.

You assign each packet that flows through the switch to a queue and to a threshold. Specifically, you map CoS values to an ingress queue and map CoS values to a threshold ID. You use the **mls qos srr-queue input cos-map queue queue-id {cos1...cos8 | threshold threshold-id cos1...cos8}** global configuration command. You can display the CoS input queue threshold map by using the **show mls qos maps** privileged EXEC command.

WTD Thresholds

The queues use WTD to support distinct drop percentages for different traffic classes. Each queue has three predefined default drop thresholds that are not changeable. For more information about how WTD works, see the [“Weighted Tail Drop” section on page 27-4](#).

Priority Queueing

The priority queue should be used for traffic (such as voice) that requires guaranteed delivery because this queue is guaranteed part of the bandwidth regardless of the load on the internal ring.

SRR services the priority queue for its configured weight as specified by the **bandwidth** keyword in the **mls qos srr-queue input priority-queue queue-id bandwidth weight** global configuration command.

You can combine the commands described in this section to prioritize traffic by placing packets with particular CoSs into certain queues. For configuration information, see the [“Configuring Ingress Queue Characteristics” section on page 27-14](#).

Queueing on Egress Queues

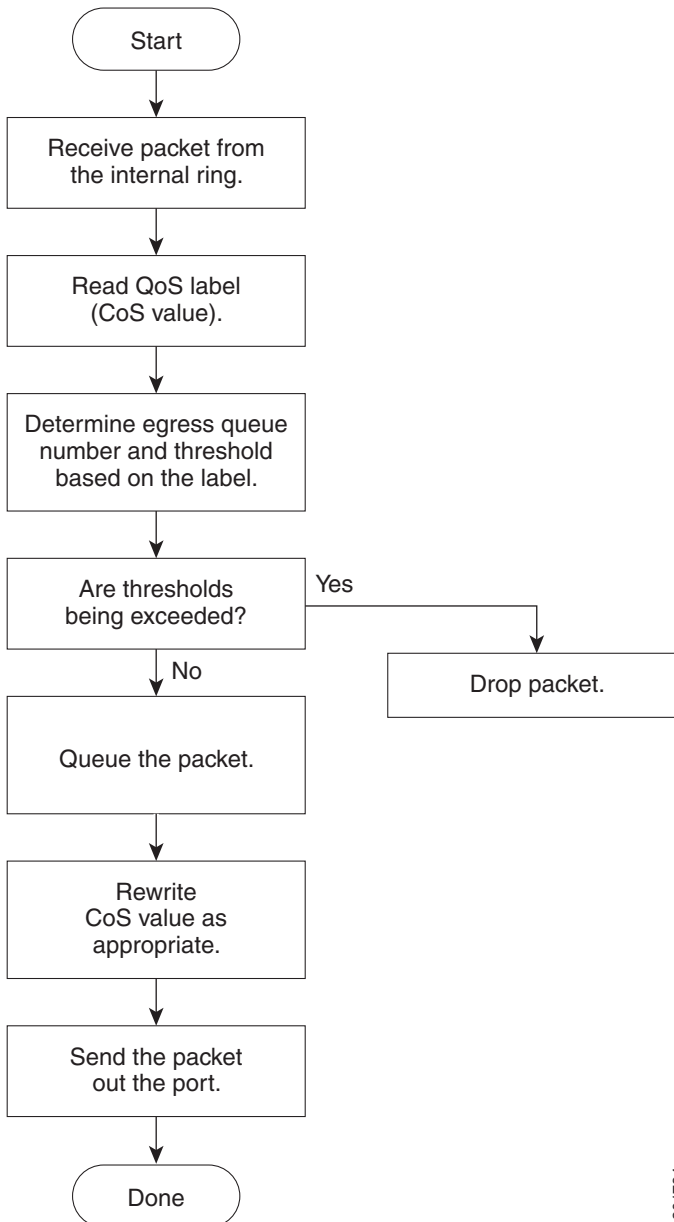
Figure 27-5 shows the queueing and scheduling flowchart for egress ports.



Note

If the expedite queue is enabled, SRR services it until it is empty before servicing the other three queues.

Figure 27-5 Queueing Flowchart for Egress Ports



201794

Each port supports four egress queues, one of which (queue 1) can be the egress expedite queue. These queues are assigned to a queue set. All traffic exiting the switch flows through one of these four queues and is subjected to a threshold based on the QoS label assigned to the packet.

The buffer space is divided between the common pool and the reserved pool. The switch uses a buffer allocation scheme to reserve a minimum amount of buffers for each egress queue, to prevent any queue or port from consuming all the buffers and depriving other queues, and to control whether to grant buffer space to a requesting queue. The switch detects whether the target queue has not consumed more buffers than its reserved amount (under limit), whether it has consumed all of its maximum buffers (over limit), and whether the common pool is empty (no free buffers) or not empty (free buffers). If the queue is not over limit, the switch can allocate buffer space from the reserved pool or from the common pool (if it is not empty). If there are no free buffers in the common pool or if the queue is over limit, the switch drops the frame.

WTD Thresholds

You can assign each packet that flows through the switch to a queue and to a threshold. Specifically, you map CoS values to an egress queue and map CoS values to a threshold ID. You use the **mls qos srr-queue output cos-map queue** *queue-id* {*cos1...cos8* | **threshold** *threshold-id* *cos1...cos8*} global configuration command. You can display the CoS output queue threshold map by using the **show mls qos maps** privileged EXEC command.

The queues use WTD to support distinct drop percentages for different traffic classes. Each queue has three predefined default drop thresholds that are not changeable. For more information about how WTD works, see the “[Weighted Tail Drop](#)” section on page 27-4.

Packet Modification

A packet is classified, policed, and queued to provide QoS. Packet modifications can occur during this process. For IP and non-IP packets, classification involves assigning a QoS label to a packet based on the CoS of the received packet. However, the packet is not modified at this stage; only an indication of the assigned CoS value is carried along.

Configuring Standard QoS

Before configuring standard QoS, you must have a thorough understanding of these items:

- The types of applications used and the traffic patterns on your network.
- Traffic characteristics and needs of your network. Is the traffic bursty? Do you need to reserve bandwidth for voice and video streams?
- Bandwidth requirements and speed of the network.
- Location of congestion points in the network.

These sections contain this configuration information:

- [Default Standard QoS Configuration](#), page 27-9
- [Enabling QoS Globally](#), page 27-11 (required)
- [Configuring Classification Using Port Trust States](#), page 27-11 (required)
- [Configuring Ingress Queue Characteristics](#), page 27-14 (optional)
- [Configuring Egress Queue Characteristics](#), page 27-16 (optional)

Default Standard QoS Configuration

QoS is disabled. There is no concept of trusted or untrusted ports because the packets are not modified (the CoS and IP precedence values in the packet are not changed). Traffic is switched in pass-through mode (packets are switched without any rewrites and classified as best effort without any policing).

When QoS is enabled with the **mls qos** global configuration command and all other QoS settings are at their defaults, traffic is classified as best effort (the CoS value is set to 0) without any policing. The default port trust state on all ports is untrusted. The default ingress and egress queue settings are described in the “[Default Ingress Queue Configuration](#)” section on page 27-9 and the “[Default Egress Queue Configuration](#)” section on page 27-10.

Default Ingress Queue Configuration

Table 27-2 shows the default ingress queue configuration when QoS is enabled.

Table 27-2 *Default Ingress Queue Configuration*

Feature	Queue 1	Queue 2
Buffer allocation	90 percent	10 percent
Bandwidth allocation ¹	4	4
Priority queue bandwidth ²	0	10
WTD drop threshold 1	100 percent	100 percent
WTD drop threshold 2	100 percent	100 percent

1. The bandwidth is equally shared between the queues. SRR sends packets in shared mode only.
2. Queue 2 is the priority queue. SRR services the priority queue for its configured share before servicing the other queue.

Table 27-3 shows the default CoS input queue threshold map when QoS is enabled.

Table 27-3 *Default CoS Input Queue Threshold Map*

CoS Value	Queue ID–Threshold ID
0–4	1–1
5	2–1
6, 7	1–1

Default Egress Queue Configuration

Table 27-4 shows the default egress queue configuration for each queue-set when QoS is enabled. All ports are mapped to queue-set 1. The port bandwidth limit is set to 100 percent and rate unlimited.

Table 27-4 Default Egress Queue Configuration

Feature	Queue 1	Queue 2	Queue 3	Queue 4
Buffer allocation	25 percent	25 percent	25 percent	25 percent
WTD drop threshold 1	100 percent	200 percent	100 percent	100 percent
WTD drop threshold 2	100 percent	200 percent	100 percent	100 percent
Reserved threshold	50 percent	50 percent	50 percent	50 percent
Maximum threshold	400 percent	400 percent	400 percent	400 percent
SRR shaped weights (absolute) ¹	25	0	0	0
SRR shared weights ²	25	25	25	25

1. A shaped weight of zero means that this queue is operating in shared mode.
2. One quarter of the bandwidth is allocated to each queue.

Table 27-5 shows the default CoS output queue threshold map when QoS is enabled.

Table 27-5 Default CoS Output Queue Threshold Map

CoS Value	Queue ID–Threshold ID
0, 1	2–1
2, 3	3–1
4	4–1
5	1–1
6, 7	4–1

General QoS Guidelines

These are general QoS guidelines:

- You configure QoS only on physical ports; there is no support for it at the VLAN or switch virtual interface level.
- Control traffic (such as spanning-tree bridge protocol data units [BPDUs] and routing update packets) received by the switch are subject to all ingress QoS processing.
- You are likely to lose data when you change queue settings; therefore, try to make changes when traffic is at a minimum.

Enabling QoS Globally

By default, QoS is disabled on the switch.

Beginning in privileged EXEC mode, follow these steps to enable QoS. This procedure is required.

	Command	Purpose
Step 1	configure terminal	Enter global configuration mode.
Step 2	mls qos	Enable QoS globally. QoS runs with the default settings described in the “ Default Standard QoS Configuration ” section on page 27-9, the “ Queueing on Ingress Queues ” section on page 27-5, and the “ Queueing on Egress Queues ” section on page 27-7.
Step 3	end	Return to privileged EXEC mode.
Step 4	show mls qos	Verify your entries.
Step 5	copy running-config startup-config	(Optional) Save your entries in the configuration file.

To disable QoS, use the **no mls qos** global configuration command.

Configuring Classification Using Port Trust States

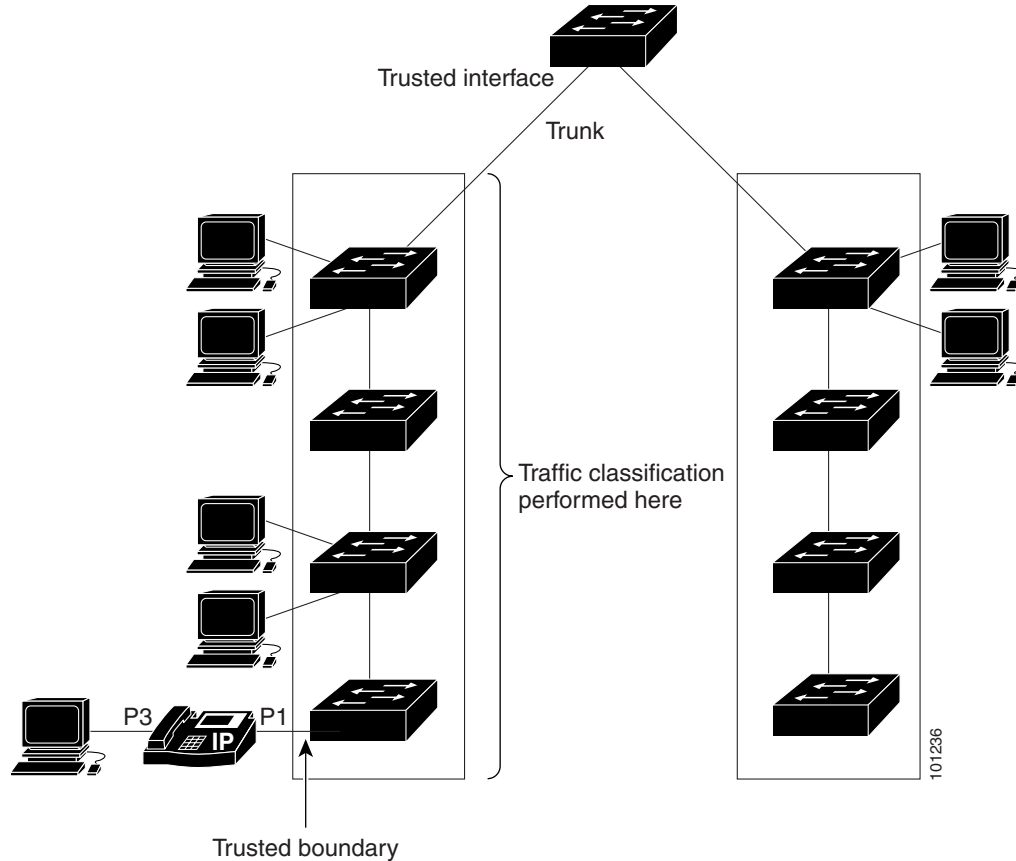
These sections describe how to classify incoming traffic by using port trust states. Depending on your network configuration, you must perform one or more of these tasks.

- [Configuring the Trust State on Ports within the QoS Domain](#), page 27-11
- [Configuring the CoS Value for an Interface](#), page 27-13
- [Enabling DSCP Transparency Mode](#), page 27-13

Configuring the Trust State on Ports within the QoS Domain

Packets entering a QoS domain are classified at the edge of the QoS domain. When the packets are classified at the edge, the switch port within the QoS domain can be configured to one of the trusted states because there is no need to classify the packets at every switch within the QoS domain. [Figure 27-6](#) shows a sample network topology.

Figure 27-6 Port Trusted States within the QoS Domain



Beginning in privileged EXEC mode, follow these steps to configure the port to trust the classification of the traffic that it receives:

	Command	Purpose
Step 1	configure terminal	Enter global configuration mode.
Step 2	interface <i>interface-id</i>	Specify the port to be trusted, and enter interface configuration mode. Valid interfaces include physical ports.
Step 3	mls qos trust [cos]	Configure the port trust state. By default, the port is not trusted. If no keyword is specified, the default is cos . The keyword has this meanings: cos —Classifies an ingress packet by using the packet CoS value. For an untagged packet, the port default CoS value is used. The default port CoS value is 0.
Step 4	end	Return to privileged EXEC mode.
Step 5	show mls qos interface	Verify your entries.
Step 6	copy running-config startup-config	(Optional) Save your entries in the configuration file.

To return a port to its untrusted state, use the **no mls qos trust** interface configuration command.

For information on how to change the default CoS value, see the [“Configuring the CoS Value for an Interface”](#) section on page 27-13.

Configuring the CoS Value for an Interface

QoS assigns the CoS value specified with the **mls qos cos** interface configuration command to untagged frames received on trusted and untrusted ports.

Beginning in privileged EXEC mode, follow these steps to define the default CoS value of a port or to assign the default CoS to all incoming packets on the port:

	Command	Purpose
Step 1	configure terminal	Enter global configuration mode.
Step 2	interface <i>interface-id</i>	Specify the port to be configured, and enter interface configuration mode. Valid interfaces include physical ports.
Step 3	mls qos cos { <i>default-cos</i> override }	Configure the default CoS value for the port. <ul style="list-style-type: none"> For <i>default-cos</i>, specify a default CoS value to be assigned to a port. If the packet is untagged, the default CoS value becomes the packet CoS value. The CoS range is 0 to 7. The default is 0. Use the override keyword to override the previously configured trust state of the incoming packet and to apply the default port CoS value to the port on all incoming packets. By default, CoS override is disabled. <p>Use the override keyword when all incoming packets on specified ports deserve higher or lower priority than packets entering from other ports. Even if a port was previously set to trust CoS or IP precedence, this command overrides the previously configured trust state, and all the incoming CoS values are assigned the default CoS value configured with this command. If an incoming packet is tagged, the CoS value of the packet is modified with the default CoS of the port at the ingress port.</p>
Step 4	end	Return to privileged EXEC mode.
Step 5	show mls qos interface	Verify your entries.
Step 6	copy running-config startup-config	(Optional) Save your entries in the configuration file.

To return to the default setting, use the **no mls qos cos** {*default-cos* | **override**} interface configuration command.

Enabling DSCP Transparency Mode

The switch supports the Differentiated Services Code Point (DSCP) transparency feature. It affects only the DSCP field of a packet at egress. By default, DSCP transparency is disabled. The switch modifies the DSCP field in an incoming packet, and the DSCP field in the outgoing packet is based on the QoS configuration, including the port trust setting, policing and marking, and the DSCP-to-DSCP mutation map.

If DSCP transparency is enabled by using the **no mls qos rewrite ip dscp** command, the switch does not modify the DSCP field in the incoming packet, and the DSCP field in the outgoing packet is the same as that in the incoming packet.

Regardless of the DSCP transparency configuration, the switch modifies the internal DSCP value of the packet, which the switch uses to generate a CoS value that represents the priority of the traffic. The switch also uses the internal DSCP value to select an egress queue and a threshold.

Beginning in privileged EXEC mode, follow these steps to enable DSCP transparency on a switch:

	Command	Purpose
Step 1	configure terminal	Enter global configuration mode.
Step 2	mls qos	Enable QoS globally.
Step 3	no mls qos rewrite ip dscp	Enable DSCP transparency. The switch is configured to not modify the DSCP field of the IP packet.
Step 4	end	Return to privileged EXEC mode.
Step 5	show mls qos interface <i>[interface-id]</i>	Verify your entries.
Step 6	copy running-config startup-config	(Optional) Save your entries in the configuration file.

To configure the switch to modify the DSCP value based on the trust setting by disabling DSCP transparency, use the **mls qos rewrite ip dscp** global configuration command.

If you disable QoS by using the **no mls qos** global configuration command, the CoS and DSCP values are not changed (the default QoS setting).

If you enter the **no mls qos rewrite ip dscp** global configuration command to enable DSCP transparency and then enter the **mls qos trust cos** interface configuration command, DSCP transparency is still enabled.

Configuring Ingress Queue Characteristics

Depending on the complexity of your network and your QoS solution, you might need to perform all of the tasks in the next sections. You will need to make decisions about these characteristics:

- Which packets are assigned by CoS value to each queue?
- Is there traffic (such as voice) that should be given high priority?

These sections contain this configuration information:

- [Mapping CoS Values to an Ingress Queue, page 27-14](#) (optional)
- [Configuring the Ingress Priority Queue, page 27-15](#) (optional)

Mapping CoS Values to an Ingress Queue

You can prioritize traffic by placing packets with particular CoSs into certain queues and adjusting the queue thresholds so that packets with lower priorities are dropped.

Beginning in privileged EXEC mode, follow these steps to map CoS values to an ingress queue and to set WTD thresholds. This procedure is optional.

	Command	Purpose
Step 1	configure terminal	Enter global configuration mode.
Step 2	mls qos srr-queue input cos-map queue <i>queue-id</i> threshold <i>threshold-id</i> cos1...cos8	Map CoS values to an ingress queue and to a threshold ID. By default, CoS values 0 to 4, 6, and 7 are mapped to queue 1 and threshold 1. CoS value 5 is mapped to queue 2 and threshold 1. <ul style="list-style-type: none"> For <i>queue-id</i>, the range is 1 to 2. For <i>threshold-id</i>, the range is 1 to 3. The drop-threshold percentage for threshold 3 is predefined. It is set to the queue-full state. For <i>cos1...cos8</i>, enter up to eight values, and separate each value with a space. The range is 0 to 7.
Step 3	end	Return to privileged EXEC mode.
Step 4	show mls qos maps	Verify your entries. The CoS input queue threshold map shows the CoS value in the top row and the corresponding queue ID and threshold ID in the second row; for example, queue 2 and threshold 2 (2-2).
Step 5	copy running-config startup-config	(Optional) Save your entries in the configuration file.

To return to the default CoS input queue threshold map, use the **no mls qos srr-queue input cos-map** global configuration command. To return to the default WTD threshold percentages, use the **no mls qos srr-queue input threshold *queue-id*** global configuration command.

Configuring the Ingress Priority Queue

You should use the priority queue only for traffic that needs to be expedited (for example, voice traffic, which needs minimum delay and jitter).

The priority queue is guaranteed part of the bandwidth to reduce the delay and jitter under heavy network traffic on an oversubscribed ring (when there is more traffic than the backplane can carry, and the queues are full and dropping frames).

SRR services the priority queue for its configured weight as specified by the **bandwidth** keyword in the **mls qos srr-queue input priority-queue *queue-id* bandwidth *weight*** global configuration command. Then, SRR shares the remaining bandwidth with both ingress queues and services them as specified by the weights configured with the **mls qos srr-queue input bandwidth *weight1 weight2*** global configuration command.

Beginning in privileged EXEC mode, follow these steps to configure the priority queue. This procedure is optional.

	Command	Purpose
Step 1	configure terminal	Enter global configuration mode.
Step 2	mls qos srr-queue input priority-queue <i>queue-id</i> bandwidth <i>weight</i>	Assign a queue as the priority queue and guarantee bandwidth on the internal ring if the ring is congested. By default, the priority queue is queue 2, and 10 percent of the bandwidth is allocated to it. <ul style="list-style-type: none"> For <i>queue-id</i>, the range is 1 to 2. For bandwidth <i>weight</i>, assign the bandwidth percentage of the internal ring. The range is 0 to 40. The amount of bandwidth that can be guaranteed is restricted because a large value affects the entire ring and can degrade performance.
Step 3	end	Return to privileged EXEC mode.
Step 4	show mls qos interface queueing or show mls qos input-queue	Verify your entries.
Step 5	copy running-config startup-config	(Optional) Save your entries in the configuration file.

To return to the default setting, use the **no mls qos srr-queue input priority-queue *queue-id*** global configuration command. To disable priority queueing, set the bandwidth weight to 0, for example, **mls qos srr-queue input priority-queue *queue-id* bandwidth 0**.

This example shows how to assign the ingress bandwidths to the queues. Queue 1 is the priority queue with 10 percent of the bandwidth allocated to it. The bandwidth ratios allocated to queues 1 and 2 is 4/(4+4). SRR services queue 1 (the priority queue) first for its configured 10 percent of the bandwidth. Then SRR equally shares the remaining 90 percent of the bandwidth between queues 1 and 2 by allocating 45 percent to each queue:

```
Switch(config)# mls qos srr-queue input priority-queue 1 bandwidth 10
Switch(config)# mls qos srr-queue input bandwidth 4 4
```

Configuring Egress Queue Characteristics

Depending on the complexity of your network and your QoS solution, you might need to perform all of the tasks in the next sections. You will need to make decisions about these characteristics:

- Which packets are mapped by CoS value to each queue and threshold ID?
- What drop percentage thresholds apply to the queue-set (four egress queues per port), and how much reserved and maximum memory is needed for the traffic type?
- How much of the fixed buffer space is allocated to the queue-set?
- Does the bandwidth of the port need to be rate limited?
- How often should the egress queues be serviced and which technique (shaped, shared, or both) should be used?

These sections contain this configuration information:

- [Configuration Guidelines, page 27-17](#)
- [Mapping CoS Values to an Egress Queue and to a Threshold ID, page 27-17](#) (optional)
- [Configuring the Egress Expedite Queue, page 27-18](#) (optional)
- [Displaying Standard QoS Information, page 27-18](#) (optional)

Configuration Guidelines

Follow these guidelines when the expedite queue is enabled or the egress queues are serviced based on their SRR weights:

- If the egress expedite queue is enabled, it overrides the SRR shaped and shared weights for queue 1.
- If the egress expedite queue is disabled and the SRR shaped and shared weights are configured, the shaped mode overrides the shared mode for queue 1, and SRR services this queue in shaped mode.
- If the egress expedite queue is disabled and the SRR shaped weights are not configured, SRR services this queue in shared mode.

Mapping CoS Values to an Egress Queue and to a Threshold ID

You can prioritize traffic by placing packets with particular costs of service into certain queues and adjusting the queue thresholds so that packets with lower priorities are dropped.



Note

The egress queue default settings are suitable for most situations. You should change them only when you have a thorough understanding of the egress queues and if these settings do not meet your QoS solution.

Beginning in privileged EXEC mode, follow these steps to map CoS values to an egress queue and to a threshold ID. This procedure is optional.

	Command	Purpose
Step 1	<code>configure terminal</code>	Enter global configuration mode.
Step 2	<code>mls qos srr-queue output cos-map queue <i>queue-id</i> threshold <i>threshold-id</i> <i>cos1...cos8</i></code>	Map CoS values to an egress queue and to a threshold ID. By default, CoS values 0 and 1 are mapped to queue 2 and threshold 1. CoS values 2 and 3 are mapped to queue 3 and threshold 1. CoS values 4, 6, and 7 are mapped to queue 4 and threshold 1. CoS value 5 is mapped to queue 1 and threshold 1. <ul style="list-style-type: none"> • For <i>queue-id</i>, the range is 1 to 4. • For <i>threshold-id</i>, the range is 1 to 3. The drop-threshold percentage for threshold 3 is predefined. It is set to the queue-full state. • For <i>cos1...cos8</i>, enter up to eight values, and separate each value with a space. The range is 0 to 7.
Step 3	<code>end</code>	Return to privileged EXEC mode.

	Command	Purpose
Step 4	<code>show mls qos maps</code>	Verify your entries. The CoS output queue threshold map shows the CoS value in the top row and the corresponding queue ID and threshold ID in the second row; for example, queue 2 and threshold 2 (2-2).
Step 5	<code>copy running-config startup-config</code>	(Optional) Save your entries in the configuration file.

To return to the default CoS output queue threshold map, use the `no mls qos srr-queue output cos-map` global configuration command.

Configuring the Egress Expedite Queue

You can ensure that certain packets have priority over all others by queuing them in the egress expedite queue. SRR services this queue until it is empty before servicing the other queues.

Beginning in privileged EXEC mode, follow these steps to enable the egress expedite queue. This procedure is optional.

	Command	Purpose
Step 1	<code>configure terminal</code>	Enter global configuration mode.
Step 2	<code>mls qos</code>	Enable QoS on a switch.
Step 3	<code>interface interface-id</code>	Specify the egress port, and enter interface configuration mode.
Step 4	<code>priority-queue out</code>	Enable the egress expedite queue, which is disabled by default. When you configure this command, the SRR weight and queue size ratios are affected because there is one less queue participating in SRR.
Step 5	<code>end</code>	Return to privileged EXEC mode.
Step 6	<code>show running-config</code>	Verify your entries.
Step 7	<code>copy running-config startup-config</code>	(Optional) Save your entries in the configuration file.

To disable the egress expedite queue, use the `no priority-queue out` interface configuration command.

This example shows how to enable the egress expedite queue.

```
Switch(config)# interface gigabitethernet0/1
Switch(config-if)# priority-queue out
Switch(config-if)# end
```

Displaying Standard QoS Information

To display standard QoS information, use one or more of the privileged EXEC commands in [Table 27-6](#):

Table 27-6 Commands for Displaying Standard QoS Information

Command	Purpose
<code>show mls qos</code>	Display global QoS configuration information.

Table 27-6 *Commands for Displaying Standard QoS Information (continued)*

Command	Purpose
<code>show mls qos maps [cos-input-q cos-output-q]</code>	Display QoS mapping information.
<code>show running-config include rewrite</code>	Display the CoS transparency setting.

