Configuring Priority Flow Control

This chapter describes how to configure priority flow control (PFC) on the Cisco MDS 9000 switches. This chapter includes the following sections:

- Information About Priority Flow Control, page 3-1
- Licensing Requirements for Priority Flow Control, page 3-2
- Guidelines and Limitations for Priority Flow Control, page 3-2
- Default Settings for Priority Flow Control, page 3-2
- Configuring Priority Flow Control, page 3-2
- Information About Long Distance, page 3-4
- Guidelines and Limitations for Long Distance, page 3-4
- Configuring Long Distance, page 3-4
- Configuration Examples for Priority Flow Control, page 3-5

Information About Priority Flow Control

Priority flow control (PFC; IEEE 802.1bb), which is also referred to as Class-based Flow Control (CBFC) or Per Priority Pause (PPP), is a mechanism that prevents frame loss that is due to congestion. PFC is similar to 802.3x Flow Control (pause frames) or link-level flow control (LFC). However, PFC functions on a per class-of-service (CoS) basis.

When a buffer threshold is exceeded due to congestion, LFC sends a pause frame to its peer to pause all data transmission on the link for a specified period of time. When the congestion is mitigated (traffic comes under the configured threshold), a resume frame is generated to restart data transmission on the link.

In contrast, during congestion, PFC sends a pause frame that indicates which CoS value needs to be paused. A PFC pause frame contains a 2-octet timer value for each CoS that indicates the length of time that the traffic needs to be paused. The unit of time for the timer is specified in pause quanta. A quanta is the time that is required for transmitting 512 bits at the speed of the port. The range is from 0 to 65535. A pause frame with a pause quanta of 0 indicates a resume frame to restart the paused traffic.

Note Only certain classes of service of traffic can be flow controlled while other classes are allowed to operate normally.

PFC asks the peer to stop sending frames of a particular CoS value by sending a pause frame to a well-known multicast address. This pause frame is a one-hop frame that is not forwarded when received by the peer. When the congestion is mitigated, PFC can request the peer to restart transmitting frames.

Licensing Requirements for Priority Flow Control

The following table shows the licensing requirements for this feature:

Product	License Requirement
Cisco NX-OS	The QoS feature does not require a license. Any feature not included in a license package is bundled with the Cisco NX-OS system images and is provided at no extra charge to you. For a complete explanation of the Cisco NX-OS licensing scheme, see the *Cisco MDS NX-OS Licensing Guide*.

Guidelines and Limitations for Priority Flow Control

PFC has the following configuration guidelines and limitations:

- If PFC is enabled on a port or a port channel, it does not cause a port flap.
- A flap occurs when both the PFC and LFC are enabled and PFC is disabled before LFC is configured.
- PFC configuration enables PFC in both the send (Tx) and receive (Rx) direction.
- PFC **on** mode is used to support the hosts that support PFC but do not support the Data Center Bridging Capability Exchange Protocol (DCBXP).
- Only an exact match of the no-drop CoS is considered as a successful negotiation of PFC by the DCBXP.

Default Settings for Priority Flow Control

Table Table 3-1 lists the default setting for PFC.

Table 3-1 Default PFC Setting

Parameter	Default
PFC	Auto

Configuring Priority Flow Control

You can configure PFC one of these three modes with the long-distance feature:

- **auto**—Enables the no-drop CoS values to be advertised by the DCBXP and negotiated with the peer. A successful negotiation enables PFC on the no-drop CoS. Any failures because of a mismatch in the capability of peers causes the PFC not to be enabled.
- **on**—Enables PFC on the local port regardless of the capability of the peers.

- **off**—Disables PFC on the local port.

Note You can also enable Link-level Flow Control (LFC) on the same port in which PFC is enabled. However, PFC, if enabled, always gets the priority.

SUMMARY STEPS

1. **configure terminal**
2. **interface ethernet** [*slot/port-number*] or **interface ethernet port-channel** [*port-number*]
3. **[no] priority-flow-control mode** {**auto** | **off** | **on**}
4. **show interface priority-flow-control**

DETAILED STEPS

	Command	Purpose
Step 1	**configure terminal** **Example:** switch# configure terminal switch(config)#	Enters global configuration mode.
Step 2	**interface ethernet** [*slot/port-number*] **Example:** switch(config)# interface ethernet 2/5 switch(config-if)#	Enters interface mode on the interface specified.
Step 3	**priority-flow-control mode** {**auto** | **off** | **on**} **Example:** switch(config-if)# priority-flow-control mode on switch(config-if)#	Sets the PFC to the **auto**, **off**, or **on** mode. By default, PFC mode is set to **auto** on all ports.
Step 4	**show interface priority-flow-control** **Example:** switch# show interface priority-flow-control	Displays the status of PFC on all interfaces.

Verifying the Priority Flow Control Configuration

To display the PFC configuration, perform the following task:

Command	Purpose
show interface priority-flow-control	Displays the status of PFC on all interfaces.

For detailed information about the fields in the output from these commands, see the *Cisco MDS 9000 Family Command Reference*.

Information About Long Distance

In case of congestion, the long distance ingress ports sends a pause frame to its peer to pause all data transmission on the link for a specified period of time in order to absorb the in-flight packets.

When a pause frame is generated from a switch, it takes time to reach the peer switch. Until the packet reaches to peers all the in-flight packets need to be buffered and should not be dropped.

If the distance between the switches is longer, time taken by a pause frame to reach the peer switch is more and the number of the in-flight packets becomes higher. All these in-flights packets need to be buffered and should not be dropped. To absorb the large numbers of in-flight packets, the buffer size needs to be larger.

The ingress buffer is governed by these two thresholds:

- Lower threshold—Pause: The lower threshold is to trigger the generation of pause packets. When the buffer reaches the pause threshold, pause frames are generated.
- High threshold—Stop: This threshold is to stop generating the pause frames. When the buffer reaches the stop threshold, the pause frames stop getting generated.

To absorb a large number of in-flight packets, the difference between the pause threshold and the stop threshold should to be more. The pause threshold should be lower and the stop threshold should be higher. If the difference between the pause and stop thresholds are higher, the number of in-flight packets are absorbed in ingress buffer. If the pause threshold is lower, a pause frame is generated even for small buffer utilization.

On Cisco MDS 9000 switches, there is only one queue, the 7e template and all buffers are allocated to this queue. The pause threshold and the stop threshold are marked at same level. With this limitation, it cannot handle the long distance functionality.

With introduction of the long distance feature for Cisco MDS 9000 Family, you can enable the long distance feature on certain ports where peer switches are at a longer distance.

Guidelines and Limitations for Long Distance

The long distance feature has the following guidelines:

- If you use the **long-distance** command, a pause frame is generated even for a small buffer utilization.
- We recommend that you enable the long distance feature only on those ports where this feature is required.
- The internal buffer is currently set at 33%.

Configuring Long Distance

SUMMARY STEPS

1. **configure terminal**
2. **interface ethernet** [*slot/port-number*] or **interface ethernet-port-channel** [*port-number*]
3. **[no] priority-flow-control long-distance**
4. **show system internal eth-qos port-node ethernet** [*slot/port-number*]

DETAILED STEPS

	Command	Purpose
Step 1	**configure terminal** **Example:** switch# configure terminal switch(config)#	Enters global configuration mode.
Step 2	**interface ethernet** [*slot/port-number*] **Example:** switch(config)# interface ethernet-port-channel 2/5 switch(config-if)#	Enters interface mode on the interface specified.
Step 3	**priority-flow-control long-distance** **Example:** switch(config-if)# priority-flow-control fcoe long-distance switch(config-if)# **no priority-flow-control long-distance** switch(config-if)# no priority-flow-control fcoe long-distance	Enables or disables the long-distance feature.
Step 4	**show sys internal eth-qos port-node ethernet 1/1** **Example:** switch# *show system internal eth-qos port-node ethernet 1/1*	Displays the status of long-distance feature.

Default Settings for Long-Distance

Table Table 3-1 lists the default setting for long-distance.

Table 3-2 Default PFC Setting

Parameter	Default
long-distance	Disabled.

Configuration Examples for Priority Flow Control

The following example shows how to configure PFC:

```
configure terminal
 interface ethernet 5/5
 priority-flow-control mode on
```