Configuring Priority Flow Control

- About Priority Flow Control, on page 1
- Prerequisites for Priority Flow Control, on page 2
- Guidelines and Limitations for Priority Flow Control, on page 2
- Default Settings for Priority Flow Control, on page 4
- Configuring Priority Flow Control, on page 4
- Enabling Priority Flow Control on a Traffic Class, on page 5
- Configuring Pause Buffer Thresholds, on page 8
- Configuring Queue Limits, on page 9
- Verifying the Priority Flow Control Configuration, on page 9
- Configuration Examples for Priority Flow Control, on page 10

About Priority Flow Control

Priority flow control (PFC; IEEE 802.1Qbb), 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.

Prerequisites for Priority Flow Control

PFC has the following prerequisites:

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

Guidelines and Limitations for Priority Flow Control

PFC on the Nexus 3500 platform has the following configuration guidelines and limitations:

- PFC is supported only on the network-qos class matching qos-group 2 and qos-group 3. Configuring it on other qos-group matching classes results in an error.
- When configuring PFC in the network-qos policy, use the **pause pfc-cos** variant of the pause command.
- The **pause no-drop** variant of the pause command is used exclusively for configuring LLFC.
- We recommend not configuring LLFC and PFC together on the same switch. Choose only one of the following configurations for flow control:

  PFC Configuration: only **pause pfc-cos** variant in network-qos config and enabling priority-flow-control on the port.

  LLFC Configuration: only **pause no-drop** variant in network-qos policy and flow control send/receive on the port.
- For PFC to function correctly, the participating entities in a PFC enabled network should honor PFC frames, as per standard. When a peer honors PFC frames, PFC frames are generated only for the queue that is congested.

  However, when a peer does not honor PFC frames, there is no provision to immediately drop the packets that have exceeded the buffer threshold. As a result, the entire no-drop buffer becomes exhausted and PFC frames are sent to other non-congested/no-traffic no-drop queues.
- PFC frames are generated for congested traffic that hit the multicast queue.
- The pause buffer threshold is configured in the network-qos policy.
- Adding pause buffer size threshold configuration is optional for cable lengths that are less than 100 meters and it need not be configured.
- For cable lengths that are greater than 100m, the pause buffer size threshold configuration is mandatory and it is required as part of the network QoS policy configuration.
- If PFC is enabled on a port or a port channel, it does not cause a port flap.
- PFC configuration enables PFC in both the send (Tx) and receive (Rx) direction.
- Configuration time quanta of the pause frames is not supported.
- The configuration does not support pausing selected streams that are mapped to a particular traffic-class queue. All flows that are mapped to the class are treated as no-drop. It blocks out scheduling for the

entire queue, which pauses traffic for all the streams in the queue. To achieve lossless service for a no-drop class, Cisco recommends that you have only the no-drop class traffic on the queue.

- When a no-drop class is classified based on 802.1p CoS x and assigned a internal priority value (qos-group) of y, Cisco recommends that you use the internal priority value x to classify traffic on 802.1p CoS only, and not on any other field. The packet priority assigned is x if the classification is not based on CoS, which results in packets of internal priority x and y to map to the same priority x.
- The PFC feature supports up to two no-drop classes of any maximum transmission unit (MTU) size. However, there is a limit on the number of PFC-enabled interfaces based on the following factors:
  - MTU size of the no-drop class
  - Buffer-size of the pause threshold
  - Number of 10G and 40G ports
  - When configuring the pause threshold, with the default MTU and 10G ports, the maximum number of interfaces that can be enabled with PFC is approximately 20 interfaces.
- You can define the upper limit of any MTU in the system using the systemjumbomtu command. The MTU range is from 1500 to 9216 bytes, and the default is 9216 bytes.
- The interface QoS policy takes precedence over the system policy. PFC priority derivation also happens in the same order.
- Ensure that you apply the same interface-level QoS policy on all PFC-enabled interfaces for both ingress and egress.

Caution Irrespective of the PFC configuration, Cisco recommends that you stop traffic before applying or removing a queuing policy that has strict priority levels at the interface level or the system level.

- To achieve end-to-end lossless service over the network, Cisco recommends that you enable PFC on each interface through which the no-drop class traffic flows (Tx/Rx).
- Cisco recommends that you change the PFC configuration when there is no traffic. Otherwise, packets already in the Memory Management Unit (MMU) of the system might not get the expected treatment.
- Cisco recommends that you use default buffer sizes for no-drop classes or configure different network QoS policies suitable to 10G and 40G interfaces and the no-drop class MTU size. If the buffer size is specified through the CLI, it allocates the same buffer size for all ports irrespective of the link speed and MTU size. Applying the same pause buffer-size on 10G and 40G interfaces is not supported.
- Do not enable WRED on a no-drop class because it results in egress queue drops.
- For VLAN-tagged packets, priority is assigned based on the 802.1p field in the VLAN tag and takes precedence over the assigned internal priority (qos-group). DSCP or IP access-list classification cannot be performed on VLAN-tagged frames.
- For non VLAN-tagged frames, priority is assigned based on the **set qos-group** action given by the ingress QoS policy. Classification is based on a QoS policy-allowed match condition such as precedence, DSCP, or access-list. You must ensure that the **pfc-cos** value provided in the network-qos policy for this class is the same as the **qos-group** value in this case.

Default Settings for Priority Flow Control

Table 1: Default PFC Setting

Parameter	Default
PFC	Auto

Configuring Priority Flow Control

You can configure PFC on a per-port basis to enable the no-drop behavior for the CoS as defined by the active network QoS policy. PFC can be configured in one of these modes:

- on—Enables PFC on the local port regardless of the capability of the peers.
- off—Disables PFC on the local port.

SUMMARY STEPS

1. **configure terminal**
2. **interface** *type slot/port*
3. **priority-flow-control mode** [| **off** |**on**]
4. **show interface priority-flow-control**

DETAILED STEPS

Procedure

	Command or Action	Purpose

Step 1	**configure terminal**
	Example:
	`switch# configure terminal`
	`switch(config)#`
		Enters global configuration mode.

Step 2	**interface** *type slot/port*
	Example:
	`switch(config)# interface ethernet 2/5`
	`switch(config-if)#`
		Enters interface mode on the interface specified.

Step 3	**priority-flow-control mode** [	**off** |**on**]
		Example:
		`switch(config-if)# priority-flow-control mode on`
		`switch(config-if)#`
			Sets PFC to the on mode.

	Command or Action	Purpose
Step 4	**show interface priority-flow-control** **Example:** `switch# show interface priority-flow-control`	(Optional) Displays the status of PFC on all interfaces.

Enabling Priority Flow Control on a Traffic Class

You can enable PFC on a particular traffic class.

SUMMARY STEPS

1. **configure terminal**
2. **class-map type qos** *class-name*
3. **match cos** *cos-value*
4. **exit**
5. **policy-map type qos** *policy-name*
6. **class type qos** *class-name*
7. **set qos-group** *qos-group-value*
8. **exit**
9. **exit**
10. **class-map type network-qos match-any** *class-name*
11. **match qos-group** *qos-group-value*
12. **exit**
13. **policy-map type network-qos** *policy-name*
14. **class type network-qos** *class-name*
15. **pause pfc cos-value**
16. **exit**
17. **exit**
18. **system qos**
19. **service-policy type network-qos** *policy-name*

DETAILED STEPS

Procedure

	Command or Action	Purpose

Step 1	**configure terminal**
	Example:
	`switch# configure terminal`
	`switch(config)#`
		Enters global configuration mode.

Step 2	**class-map type qos** *class-name* **Example:**	Creates a named object that represents a class of traffic. Class-map names can contain alphabetic, hyphen, or

	Command or Action	Purpose
	`switch(config)# class-map type qos c1` `switch(config-cmap-qos)#`	underscore characters, are case sensitive, and can be up to 40 characters.
Step 3	**match cos** *cos-value* **Example:** `switch(config-cmap-qos)# match cos 2`	Specifies the CoS value to match for classifying packets into this class. You can configure a CoS value in the range of 0 to 7.

Step 4	**exit**
	Example:
	`switch(config-cmap-qos)# exit`
	`switch(config)#`
		Exits class-map mode and enters global configuration mode.

Step 5	**policy-map type qos** *policy-name*
	Example:
	`switch(config)# policy-map type qos p1`
	`switch(config-pmap-qos)#`
		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 6	**class type qos** *class-name*
	Example:
	`switch(config-pmap-qos)# class type qos c1`
	`switch(config-pmap-c-qos)#`
		Associates a class map with the policy map and enters the configuration mode for the specified system class.
		Note
		The associated class map must be the same type as the policy map type.

Step 7	**set qos-group** *qos-group-value* **Example:** `switch(config-pmap-c-qos)# set qos-group 2`	Configures one or more qos-group values to match on for classification of traffic into this class map. There is no default value.

Step 8	**exit**
	Example:
	`switch(config-pmap-c-qos)# exit`
	`switch(config-pmap-qos)#`
		Exits the system class configuration mode and enters policy-map mode.

Step 9	**exit**
	Example:
	`switch(config-pmap-qos)# exit`
	`switch(config)#`
		Exits policy-map mode and enters global configuration mode.

Step 10	**class-map type network-qos match-any** *class-name*
	Example:
	`switch(config)# class-map type network-qos match-any c1`
	`switch(config-cmap-nqos)#`
		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 11	**match qos-group** *qos-group-value* **Example:** `switch(config-cmap-nqos)# match qos-group 2`	Configures the traffic class by matching packets based on a list of QoS group values. Supported on QoS group 2 and QoS group 3.

	Command or Action	Purpose

Step 12	**exit**
	Example:
	`switch(config-cmap-nqos)# exit`
	`switch(config)#`
		Exits class-map mode and enters global configuration mode.

Step 13	**policy-map type network-qos** *policy-name*
	Example:
	`switch(config)# policy-map type network-qos p1`
	`switch(config-pmap-nqos)#`
		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 14	**class type network-qos** *class-name*
	Example:
	`switch(config-pmap-nqos)# class type network-qos c1`
	`switch(config-pmap-nqos-c)#`
		Associates a class map with the policy map, and enters the configuration mode for the specified system class.
		Note
		The associated class map must be the same type as the policy map type.

Step 15	**pause pfc cos-value** **Example:** `switch(config-pmap-nqos-c)# pause pfc-cos 2`	PFC sends a pause frame that indicates which CoS value needs to be paused. (CoS value range is 0-7.) **Note** Nexus 3500 does not support the **receive** option of the pause command. An error occurs when the **receive** option is configured.

Step 16	**exit**
	Example:
	`switch(config-pmap-nqos-c)# exit`
	`switch(config-pmap-nqos)#`
		Exits configuration mode and enters policy-map mode.

Step 17	**exit**
	Example:
	`switch(config-pmap-nqos)# exit`
	`switch(config)#`
		Exits policy-map mode and enters global configuration mode.

Step 18	**system qos**
	Example:
	`switch(config)# system qos`
	`switch(config-sys-qos)#`
		Enters system class configuration mode.

Step 19	**service-policy type network-qos** *policy-name* **Example:** `switch(config-sys-qos)# service-policy type network-qos p1`	Applies the policy map of type network-qos at the system level or to the specific interface.

Configuring Pause Buffer Thresholds

Pause buffer thresholds are configured in the network-qos policy. It is shared by all the ports in the system.

Note Configuring pause thresholds in the ingress queuing policy is not supported for the Nexus 3500.

SUMMARY STEPS

1. **configure terminal**
2. **policy-map type queuing** *policy-map-name*
3. **class-map type network-qos** *class-map-name*
4. **pause buffer-size** *buffer-size* **pause threshold** *xoff-size* **resume threshold** *xon-size* **pfc-cos** *pfc-cos-value*
5. **no pause buffer-size** *buffer-size* **pause threshold** *xoff-size* **resume-threshold** *xon-size* **pfc-cos** *pfc-cos-value*

DETAILED STEPS

Procedure

	Command or Action	Purpose
Step 1	**configure terminal**	Enters global configuration mode.
Step 2	**policy-map type queuing** *policy-map-name*	Enters policy-map queuing class mode and identifies the policy map assigned to the type queuing policy map.
Step 3	**class-map type network-qos** *class-map-name*	Attaches the class map of type network-qos and then enters class network-qos mode.
Step 4	**pause buffer-size** *buffer-size* **pause threshold** *xoff-size* **resume threshold** *xon-size* **pfc-cos** *pfc-cos-value*	Specifies the buffer threshold settings for pause and resume.
Step 5	**no pause buffer-size** *buffer-size* **pause threshold** *xoff-size* **resume-threshold** *xon-size* **pfc-cos** *pfc-cos-value*	Removes the buffer threshold settings for pause and resume.

Example

Example:

```
switch(config-cmap-nqos)# class type network-qos nc2
switch(config-cmap-nqos)# match qos-group 2
switch(config-cmap-nqos)#
switch(config-cmap-nqos)# policy-map type network-qos n1
switch(config-pmap-nqos)# class type network-qos nc2
switch(config-pmap-nqos-c)# pause buffer-size 30000 pause-threshold 29000 resume-threshold
 12480 pfc-cos 2
```

Configuring Queue Limits

The queue-limit is configured with the network-qos policy.

Note A queue-limit is configurable under a no-drop (PFC) enabled network-qos class. However, a queue-limit is not effective on such classes.

SUMMARY STEPS

1. **configure terminal**
2. **policy-map type network-qos** *policy-map-name*
3. **class-map type network-qos** *class-map-name*
4. **queue-limit** *queue-size* **bytes**

DETAILED STEPS

Procedure

	Command or Action	Purpose
Step 1	**configure terminal**	Enters global configuration mode.
Step 2	**policy-map type network-qos** *policy-map-name*	Enters network-qos queuing class mode and identifies the policy map assigned to the type network-qos policy map.
Step 3	**class-map type network-qos** *class-map-name*	Attaches the class map of type network-qos and then enters network-qos class queuing mode.
Step 4	**queue-limit** *queue-size* **bytes**	Specifies the queue-limit. (Range is 20480 to 6000000.)

Verifying the Priority Flow Control Configuration

To display the PFC configuration, perform the following task:

Command	Purpose
show interface priority-flow-control {**module** [*number*]}	Displays the status of PFC on all interfaces or on specific modules.
show interface priority-flow-control [*detail*] **module** [*number*]	Displays the detail status of PFC on all interfaces or on specific modules.
show interface Ethernet {*mod-number* | *port-number*} **priority-flow-control** [*detail*]	Displays the PFC status per interface.

Configuration Examples for Priority Flow Control

The following example shows how to configure PFC:

```
configure terminal
interface ethernet 1/1
priority-flow-control mode on
```

The following example shows how to enable PFC on a traffic class:

```
switch(config)# class-map type qos c2
switch(config-cmap-qos)# match cos 2
switch(config-cmap-qos)# exit

switch(config)# policy-map type qos p1
switch(config-pmap-qos)# class type qos c2
switch(config-pmap-c-qos)# set qos-group 2
switch(config-pmap-c-qos)# exit
switch(config-pmap-qos)# exit

switch(config)# class-map type queuing cq2
switch(config-cmap-que)# match qos-group 2
switch(config-cmap-que)# exit

switch(config)# policy-map type queuing pq1
switch(config-pmap-que)# class type queuing cq2
switch(config-pmap-c-que)# bandwidth percent 20
switch(config-pmap-c-que)# exit
switch(config-pmap-que)# exit

switch(config)# class-map type network-qos cn1
switch(config)# class-map type network-qos n2
switch(config-cmap-nqos)# match qos-group 2
switch(config-cmap-nqos)# exit

switch(config)# policy-map type network-qos pn1
switch(config-pmap-nqos)# class type network-qos n2
switch(config-pmap-nqos-c)# pause pfc-cos 2
switch(config-pmap-nqos-c)# exit
switch(config-pmap-nqos)# exit

switch(config)# system qos
switch(config-sys-qos)# service-policy type network-qos pn1
switch(config-sys-qos)# service-policy type qos input p1
switch(config-sys-qos)# service-policy type queuing output pq1
```

Note When there is an issue with priority-flow-control functionality, collect the output from the following commands for troubleshooting:

- **show tech-support module 1**
  - Displays internal QoS hardware buffer/configuration information.
- **show tech-support aclqos**
  - Displays PFC configuration/status commands.
- **show tech-support**
  - Displays **show running config** output along with other QoS internal commands.