Configuring Priority Flow Control

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.

Licensing Requirements for Priority Flow Control

The following table shows the licensing requirements for this feature:

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

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 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.
- 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 three 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
  - Number of 10G and 40G ports
- 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 input queuing 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.
- PFC is not supported on the Cisco Nexus 9300 platform and the N9K-X9564TX and N9K-X9564PX linecards.

Note Since the no-drop queuing configuration requires PFC, the no-drop queuing configuration is not supported in the network-qos policy for the Cisco Nexus 9300 platform.

- Dynamic load balancing cannot be enabled for internal links with PFC. You must disable DLB and enable RTAG7 load-balancing for internal links with the port-channel load-balance internal rtag7 command.
- The dynamic load balancing (DLB) based hashing scheme is enabled by default on all internal links of a linecard. When DLB is enabled, no-drop traffic might experience out-of-order packet delivery when congestion on internal links occurs and PFC is applied. If applications on the system are sensitive to out-of-order delivery, you can adjust for this by disabling DLB at the qos-group level. Disable DLB by using the **set dlb-disable** action in the QoS policy-maps and the **set qos-group** action for no-drop classes.

  In the following example assume that qos-group 1 is a no-drop class. DLB is disabled for this no-drop class by adding the **set dlb-disable** action and the **set qos-group** action.

```
switch(config)# policy-map p1
```

```
switch(config-pmap-qos)# class c1
switch(config-pmap-c-qos)# set qos-group 1
switch(config-pmap-c-qos)# set dlb-disable
switch(config-pmap-c-qos)# end
switch# show policy-map p1


  Type qos policy-maps
  ====================

  policy-map type qos p1
    class  c1
      set qos-group 1
      set dlb-disable
```

- 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** {**on**}
4. **show interface priority-flow-control**

DETAILED STEPS

	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** {**on**}
	
	Example:
	`switch(config-if)# priority-flow-control mode on`
	`switch(config-if)#`
		Sets PFC to the on mode.

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 buffer-size** *buffer-size* **pause-threshold** *xoff-size* **resume-threshold** *xon-size* **pfc-cos** *cos-value*
16. **exit**
17. **exit**
18. **system qos**
19. **service-policy type network-qos** *policy-name*

DETAILED STEPS

	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:
	`switch(config)# class-map type qos c1`
	`switch(config-cmap-qos)#`
		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	**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.

	Command or Action	Purpose

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. Values can range from 0 to 5. QoS group 0 is equivalent to class-default.

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.

	Command or Action	Purpose

Step 14	**class type network-qos** *class-name*
	
	Example:
	`switch(config-pmap-nqos)# class type network-qos c-nq1`
	`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 buffer-size** *buffer-size* **pause-threshold** *xoff-size* **resume-threshold** *xon-size* **pfc-cos** *cos-value*
	
	Example:
	`switch(config-pmap-nqos-c)# pause buffer-size 20000 pause-threshold 100 resume-threshold 1000 pfc-cos 1`
		Specifies the buffer threshold settings for pause and resume.
		
		• **buffer-size** *buffer-size*—Specifies the buffer size for ingress traffic, in bytes. Valid values are from 10240 to 490880.
		
		Note You can configure a maximum buffer size of 143680 bytes.
		
		• **pause-threshold** *xoff-size*—Specifies the buffer limit at which the port pauses the peer, in bytes. Valid values are from 0 to 490880.
		
		Note You can configure a maximum pause threshold value of 58860 bytes.
		
		• **resume-threshold** *xon-size*—Specifies the buffer limit at which the port resumes the peer, in bytes. Valid values are from 0 to 490880.
		
		Note You can configure a maximum resume threshold value of 38400 bytes.
		
		• **pfc-cos** *cos-value*—Specifies the CoS values on which to assert PFC. Valid values are from 0 to 7.
		
		Note When you configure the buffer size, ensure the following:
		
		• The buffer size must be greater than the pause threshold value, and the pause threshold value must be greater than the resume threshold value. Otherwise, the following message appears:
		
		`ERROR: buffer-size can't be less then pause/resume-threshold`
		
		• The minimum difference between the pause threshold value and the resume threshold value must be 20480 bytes. Otherwise, the following message appears:
		
		`Warning: The recommended difference between pause and resume threshold is 20480 bytes`
		
		• The pause threshold value must be greater than the resume threshold value. Otherwise, the following message appears:
		
		`ERROR: pause-threshold can't be less then resume-threshold`

	Command or Action	Purpose

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 and Queue Limit using Ingress Queuing Policy

The pause buffer thresholds specified in the network-qos policy are shared by all the ports in the system. However, there are situations where a few ports may need different thresholds (such as long distance connections). An ingress queuing policy can be used for this purpose.

An ingress queuing policy also allows the configuration of the queue-limit to restrict the amount of shared buffer that can be used in addition to the reserved pause buffer by the no-drop class.

Each no-drop class is mapped internally to one of the port's priority-group in the ingress direction. The configured pause buffer thresholds and queue-limit are applied to the priority-group associated with the class.

SUMMARY STEPS

1. **configure terminal**
2. **policy-map type queuing** *policy-map-name*
3. **class type queuing** *c-in-q1*
4. **pause buffer-size** *buffer-size* **pause threshold** *xoff-size* **resume threshold** *xon-size*
5. **no pause buffer-size** *buffer-size* **pause threshold** *xoff-size* **resume threshold** *xon-size*
6. **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** *policy-map-name*	Enters policy-map queuing class mode and identifies the policy map assigned to the type queuing policy map.
Step 3	**class type queuing** *c-in-q1*	Attaches the class map of type queuing and then enters policy-map class queuing mode. Class queuing names are listed in the System-Defined Type queuing Class Maps table. **Note** The qos-group associated with the class must be defined as a no-drop class in the network-qos policy applied in the system qos.
Step 4	**pause buffer-size** *buffer-size* **pause threshold** *xoff-size* **resume threshold** *xon-size*	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*	Removes the buffer threshold settings for pause and resume.
Step 6	**queue-limit** *queue size* [**dynamic** *dynamic threshold*]	(Optional) Specifies either the static or dynamic shared limit available to the ingress priority-group. The static queue limit defines the fixed size to which the priority-group can grow. The dynamic queue limit allows the priority-group's threshold size to be decided depending on the number of free cells available, in terms of the alpha value.

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.

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
```

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

```
switch(config)# class-map type qos c1
switch(config-cmap-qos)# match cos 3
switch(config-cmap-qos)# exit
switch(config)# policy-map type qos p1
switch(config-pmap-qos)# class type qos c1
switch(config-pmap-c-qos)# set qos-group 3
switch(config-pmap-c-qos)# exit
switch(config-pmap-qos)# exit
switch(config)# class-map type network-qos match-any c1
switch(config-cmap-nqos)# match qos-group 3
switch(config-cmap-nqos)# exit
switch(config)# policy-map type network-qos p1
switch(config-pmap-nqos)# class type network-qos c-nq1
switch(config-pmap-nqos-c)# pause buffer-size 20000 pause-threshold 100 resume-threshold
1000 pfc-cos 3
switch(config-pmap-nqos-c)# exit
switch(config-pmap-nqos)# exit
switch(config)# system qos
switch(config-sys-qos)# service-policy type network-qos p1
```