Troubleshooting QoS Issues

The Cisco Nexus 5000 Series NX-OS quality of service (QoS) provides the most desirable flow of traffic through a network. QoS uses policies and flow control to classify the network traffic, police and prioritize the traffic flow, and provide congestion avoidance.

This chapter describes how to identify and resolve problems that can occur with QoS in the Cisco Nexus 5000 Series switch.

This chapter includes the following sections:

- Policy Maps
- Improper Configurations
- PFC
- Registers and Counters

Policy Maps

The Nexus 5000 QoS implementation follows the Cisco Modular QoS CLI model. It takes three steps to configure the QoS:

- Define the class map.
- Create a policy map to define the action taken for each class map.
- Apply the policy-map.

The Nexus 5000 implements three different types of policy maps:

- Policy-map type qos
- Policy-map type queuing
- Policy-map type network-qos

Additionally, the Nexus 5000 introduces a new configuration context for QoS called the System QoS. The policy-map applied under the System QoS context is applied to the entire switch.

The following table summarizes the function and attach points for these three types of policy maps.

Table 1-1 Types of Policy Maps

Policy Type	Function	Attach Point
QoS	• Define traffic classification	• System QoS • Ingress interface
Queuing	• Strict Priority queue • Deficit Weight Round Robin	• System QoS • Egress interface • Ingress interface
Network-QoS	• Define flow control mechanism (PAUSE or tail drop) • MTU per class of service • Queue size • Marking	• System QoS

With the basic process, the incoming packets are compared to the QoS classification rules that are defined by policy-map type qos. The packets are classified into 1 of 8 qos-groups.

Next, the Network-QoS and Queuing policies are applied to the packets. The Queuing policy and the Network-QoS policy define actual QoS parameters for packets belonging to each qos-group.

Note

- The Queuing and Network-QoS policies match the qos-group (identified by policy-map type qos) instead of the actual packet headers.
- When the same type of service policy is applied under the System QoS context and the interface level, the interface level service policy is preferred.
- The queuing policy that is applied under the ingress interface is not applied locally. The queueing policy is the bandwidth allocation for a different class of service that is exchanged with its peer using the DCBX protocol.

Improper Configurations

Cannot pass frame size larger than 2300 bytes through switch

Although the jumbo MTU has been configured for class-default, you cannot pass a frame size larger than 2300 bytes through the Nexus 5000 switch and the Nexus 2000 FEX.

Possible Cause

The CoS value may conflict with the existing MTU value.

Solution

CoS 7 is used internally for controlling traffic between the Nexus 5000 switch and the Nexus 2000 FEX. The MTU value for the traffic with CoS 7 is set to a fixed value. You must check that the incoming traffic is marked with CoS 7. Use any CoS value other than 7 to avoid this limitation.

MTU for "class-default" value is 1500 when jumbo MTU configured

When the configuration for the network-qos policy-map sets the class-default to jumbo MTU, the **show queuing interface** command indicates that the MTU for class-default is 1500.

Possible Cause

An incorrect startup configuration may exist after an upgrade.

Solution

If the switch has been upgraded to the 4.2(1)N1(1) release, make sure that you have used the **write erase** command to delete the startup configuration. You can save the configuration first to another file name.

After the Nexus 5000 switch boots up with an empty configuration, reapply the original configuration. You might lose your connectivity to the Nexus 5000 if you are using Telnet or SSH. It is recommended that you use the console for this procedure.

Traffic not queued or prioritized correctly on Nexus 2148, Nexus 2232, and Nexus 2248

After configuring all three types of policy maps (QoS, Network-QoS, and Queuing), the traffic is not queued or prioritized correctly on Nexus 2148, Nexus 2232, and Nexus 2248 switches.

Possible Cause

The Nexus 2148, Nexus 2232, and Nexus 2248 FEX can only support CoS-based traffic classification. The QoS service policy type configured under System QoS is populated from the Nexus 5000 to FEX only when all the matching criteria are match cos. If other match clauses exist, such as match dscp or match ip access-group in the QoS policy map, then the FEX does not accept the service policy. As a result, all the traffic is placed into the default queue.

Note Use the **show queuing interface** command to ensure that the queues have been created properly.

Solution

For the ingress traffic (from server to network) that is not marked with a CoS value, the traffic is placed into the default queue on FEX. Once the traffic is received on the Nexus 5000, it is classified based on a configured rule and are placed in the proper queue.

For the egress traffic (from Nexus 5000 to FEX, and then FEX to server), it is recommended that you mark mark the traffic with a CoS value on the Nexus 5000 so that the FEX can classify and queue the traffic properly.

The following example is a complete Nexus 5000 and Nexus 2232/Nexus 2248 configuration that classifies the traffic and configures the proper bandwidth for each type of traffic. This example applies only to the Nexus 5000 and Nexus 2248. The configuration for the Nexus 2148 is slightly different due to the fact that Nexus 2148 has only two queues for user data. The Nexus 2232/Nexus 2248 has six hardware queues for user data, which is the same as Nexus 5000.

Example:

```
//class-map for global qos policy-map, which will be used to create CoS-queue mapping.//
class-map type qos voice-global
match cos 5
class-map type qos critical-global
match cos 6
class-map type qos scavenger-global
```

```
match cos 1
class-map type qos video-signal-global
match cos 4

//This qos policy-map will be attached under "system qos". It will be downloaded to 2248
to create CoS to queue mapping.//
policy-map type qos classify-5020-global
class voice-global
set qos-group 5
class video-signal-global
set qos-group 4
class critical-global
set qos-group 3
class scavenger-global
set qos-group 2
class-map type qos Video
match dscp 34
class-map type qos Voice
match dscp 40,46
class-map type qos Control
match dscp 48,56
class-map type qos BulkData
match dscp 10
class-map type qos Scavenger
match dscp 8
class-map type qos Signalling
match dscp 24,26
class-map type qos CriticalData
match dscp 18

//This qos policy-map will be applied under all N5k and 2248 interfaces to classify all
incoming traffic based on DSCP marking. Please note that even the policy-map will be
applied under Nexus 2248 interfaces the traffic will be classified on N5k//
policy-map type qos Classify-5020
class Voice
set qos-group 5
class CriticalData
set qos-group 3
class Control
set qos-group 3
class Video
set qos-group 4
class Signalling
set qos-group 4
class Scavenger
set qos-group 2
class-map type network-qos Voice
match qos-group 5
class-map type network-qos Critical
match qos-group 3
class-map type network-qos Scavenger
match qos-group 2
class-map type network-qos Video-Signalling
match qos-group 4

//This policy-map type network-qos will be applied under "system qos" to define the MTU,
marking and queue-limit(not configured here).//
policy-map type network-qos NetworkQoS-5020
class type network-qos Voice
set cos 5
class type network-qos Video-Signalling
set cos 4
mtu 9216
class type network-qos Scavenger
```

```
set cos 1
mtu 9216
class type network-qos Critical
set cos 6
mtu 9216
class type network-qos class-default
mtu 9216
class-map type queuing Voice
match qos-group 5
class-map type queuing Critical
match qos-group 3
class-map type queuing Scavenger
match qos-group 2
class-map type queuing Video-Signalling
match qos-group 4

//The queuing interface will be applied under "system qos" to define the priority queue
and how bandwidth is shared among non-priority queues.//
policy-map type queuing Queue-5020
class type queuing Scavenger
bandwidth percent 1
class type queuing Voice
priority
class type queuing Critical
bandwidth percent 6
class type queuing Video-Signalling
bandwidth percent 20
class type queuing class-fcoe
bandwidth percent 0
class type queuing class-default
bandwidth percent 73

//The input queuing policy determines how bandwidth are shared for FEX uplink in the
direction from FEX to N5k. The output queueing policy determines the bandwidth allocation
for both N5k interfaces and FEX host interfaces.//
system qos
service-policy type qos input classify-5020-global
service-policy type network-qos NetworkQoS-5020
service-policy type queuing input Queue-5020
service-policy type queuing output Queue-5020

//Apply service-policy type qos under physical interface in order to classify traffic
based on DSCP. Please note that for portchannel member the service-policy needs to be
configured under interface port-channel.//
interface eth1/1-40
service-policy type qos input Classify-5020
interface eth100/1/1-48
service-policy type qos input Classify-5020
```

The **show queuing interface** command can be used to ensure that the CoS-to-queue mapping is properly configured under the FEX interfaces. It can also be used to check the bandwidth and MTU configuration.

This same command can be used to check the QoS configuration for the Nexus 5000 interfaces.

The following is the output from the **show queuing interface** command for the Nexus 2248 interfaces when the above configrations are applied:

```
switch# sh queuing interface ethernet 100/1/1
Ethernet100/1/1 queuing information:
  Input buffer allocation:
  Qos-group: 0  2  3  4  5  (shared)
  frh: 2
  drop-type: drop
  cos: 0 1 2 3 4 5 6
```

```
xon       xoff      buffer-size
  ------+---------+-----------
  21760     26880     48640
Queueing:
  queue qos-group cos            priority  bandwidth mtu
  -----+----------+------------+-----------+-------------+------
  2        0            0 2 3           WRR        73      9280
  4        2            1               WRR         1      9280
  5        3            6               WRR         6      9280
  6        4            4               WRR        20      9280
  7        5            5               PRI         0      1600
Queue limit: 64000 bytes

  Queue Statistics:
  queue  rx              tx
  ------+-------------+--------------
  2      113822539041    1
  4      0               0
  5      0               0
  6      417659797       0
  7      0               0
Port Statistics:
  rx drop         rx mcast drop   rx error        tx drop
  ------------+-----------------+---------------+---------------
  0               0               0               0

  Priority-flow-control enabled: no
  Flow-control status:
  cos     qos-group   rx pause  tx pause  masked rx pause
  -------+-----------+---------+---------+---------------
  0              0    xon       xon       xon
  1              2    xon       xon       xon
  2              0    xon       xon       xon
  3              0    xon       xon       xon
  4              4    xon       xon       xon
  5              5    xon       xon       xon
  6              3    xon       xon       xon
  7            n/a    xon       xon       xon
switch#
```

The Nexus 2148 has two queues in both the ingress and egress directions. One queue is mapped to the no-drop system class and another queue is mapped to the drop system class. For the ingress direction, the two queues are scheduled using WRR (Weight Round Robin). For the egress direction, the queue for the no-drop system class is the priority queue.

In order to separate traffic for the two queues, the user has to create a no-drop system class. All no-drop system classes created on the Nexus 5000 are mapped to the no-drop queue on the Nexus 2148.

The **pause no-drop** command is added to the Network-QoS in order for the Nexus 2148 to place voice in the priority queue at the FEX egress direction.

Example:

```
policy-map type network-qos NetworkQoS-5020
  class type network-qos Voice
    set cos 5
    pause no-drop
  class type network-qos Video-Signalling
    set cos 4
    mtu 9216
  class type network-qos Scavenger
    set cos 1
    mtu 9216
  class type network-qos Critical
```

```
    set cos 6
    mtu 9216
  class type network-qos class-default
    mtu 9216
```

The configuration classifies the incoming voice traffic based on DSCP and marks the voice traffic to CoS 5. At the Nexus 2148 egress direction, the FEX assigns voice traffic to the priority queue.

The following is example output from the **show queuing interface** command for the Nexus 2148 with the above configuration.

Example:

```
switch# sh queuing interface ethernet 199/1/1
Ethernet199/1/1 queuing information:
  Input buffer allocation:
  Qos-group: 0  2  3  4  (shared)
  frh: 3
  drop-type: drop
  cos: 0 1 2 3 4 6 7
  xon       xoff      buffer-size
  ---------+---------+-----------
  16640     33280     56320

  Qos-group: 5
  frh: 2
  drop-type: no-drop
  cos: 5
  xon       xoff      buffer-size
  ---------+---------+-----------
  8960      19200     34560

  Queueing:
  queue    qos-group    cos            priority  bandwidth mtu
  ------+------------+--------------+---------+----------+----
3         0 2 3 4      0 1 2 3 4 6     WRR       100      9280
  2         5            5               PRI         0      1600

  Buffer threshold: 271360 bytes
  Queue limit: Disabled

  Queue Statistics:
  queue  rx
  ------+---------------
  3      241439087
  2      0

  Port Statistics:
  tx queue drop
  ---------------
  0

  Priority-flow-control enabled: no
  Flow-control status:
  cos     qos-group   rx pause  tx pause  masked rx pause
  -------+-----------+---------+---------+---------------
  0              0    xon       xon       xon
  1              2    xon       xon       xon
  2              0    xon       xon       xon
  3              0    xon       xon       xon
  4              4    xon       xon       xon
  5              5    xon       xon       xon
  6              3    xon       xon       xon
  7            n/a    xon       xon       xon
```

```
switch#
```

PFC

Link pause (flow control) not enabled on back to back Nexus 5000 switch links

When link pause (flow control) is not enabled on back-to-back Nexus 5000 switch links, packets are dropped while sending traffic on a no-drop class.

Possible Cause

If the peer Nexus 5000 switch supports PFC TLV with DCBX, then configuring **flowcontrol send on** and **flowcontrol receive on** will not enable the link pause. You have to disable the PFC TLV sent by DCBX on that interface.

Use one of the following commands to verify:

- Use the **show interface ethx/y flowcontrol** command and check to see if the operating state is off.
- Use the **show interface ethx/y priority-flow-control** command and check to see if the operating state is on.

Solution

Configure the following commands under **interface ethx/y** to enable link pause instead of PFC on back-to-back switch links.

- **no priority-flow-control mode on**
- **flowcontrol receive on**
- **flowcontrol send on**

Cannot enable “pause no-drop” on more than one ethernet class

Cannot enable pause no-drop on more than one Ethernet class.

CLI commands fail with the following error when trying to enable pause no-drop.

```
ERROR:   Module 1 returned status "Not enough buffer space available. Please change your
configuration and re-apply"
```

Possible Cause

Nexus 5000 supports a maximum of three no drop classes (including FCoE). If five Ethernet classes are created, then there will be insufficient buffers to enable twoof the five Ethernet no-drop classes.

You will get an error if not enough buffers exist to enable the no-drop.

Example:

```
class type network-qos s4
pause no-drop
ERROR:   Module 1 returned status "Not enough buffer space available. Please change your
configuration and re-apply"
```

Solution

If you create five **ethernet** classes, then there will be an insufficient number of buffers to configure two of the five Ethernet no-drop classes. If you delete two Ethernet classes and configure the remaining three Ethernet classes (including class-default), then no-drop can be enabled on two of the Ethernet classes.

Changing no-drop configuration causes VPC peer-link to go down and FEX to go offline

Changing the QoS no-drop configuration causes the VPC MCT peer-link to go down and FEX to go offline.

Possible Cause

The network QoS policy parameters, such as MTU and pause, are treated as type1 parameters and should match between the VPC primary and secondary nodes. If a mismatch exists between the VPC primary and secondary nodes, then the VPC peer-link does not come up and FEX goes offline. Only CoS based class no-drop/MTU parameters are considered as type 1 consistency checked for VPC. If you configure an ACL based class, then it is not treated as a vtype 1 parameter for VPC.

Use one of the following commands to verify:

- **show vpc brief**
- **show vpc consistency-parameters global**

Solution

Configure the similar no-drop class configuration between the VPC primary and secondary nodes. Any mismatch of no-drop policy on nqos CoS-based class parameters causes a type1 inconsistency.

Pause enabled on all cos values when no-drop enabled on class-ip-multicast

Priority flow control enables pause on all CoS values when no-drop is enabled on the class-ip-multicast class.

Possible Cause

When you create a class-ip-multicast class and no-drop is enabled, then pause is enabled on all of the CoS values.

Use the **show interface ethx/y priority-flow-control** command and check that the VL bitmap is enabled for all CoS values (ff).

Solution

Use the following commands to enable PFC on CoS 4 only, instead of on all CoS values under the class-ip-multicast class.

- **Policy-map type network-qos system**
- **Class type network-qos class-ip-multicast**
- **Pause no-drop pfc-cos 4**

No drop class not created on N2K-C2148T/N2K-C2248TP-1GE based FEX with default QoS configuration

The no-drop class is not created on the N2K-C2148T/N2K-C2248TP-1GE based FEX with the default QoS configuration.

The show queuing interface is different for the switchport and HIF port on N2K-C2248TP and N2K-C2148T.

Possible Cause

FCoE is not supported on the N2K-C2148T and N2K-C2248TP-1GE based FEX and the no drop class is not created with the default QoS configuration.

Use the following command to verify (check for no-drop class):

show queuing interface eth100/1/1

Solution

If you want an ethernet no-drop class on a N2K-C2148T/N2K-C2248TP-1GE FEX, then you have to create an ethernet no-drop class with the following:

- **Policy-map type network-qos no-drop**
- **Class type network-qos class-0**
- **Pause no-drop**

How to enable link pause (flow control) on Nexus 5000 interface

Configuring “lowcontrol send on and flowcontrol receive on does not enable flowcontrol on on Nexus 5000 switch port links when connected to another Nexus 5000 interface.

Possible Cause

By default, the DCBX runs on the Nexus 5000 interface. If the peer does not run DCBX, then the interface is configured for tail-drop.

Use one of the following commands to verify:

- Use the **show interface ethx/y flowcontrol** command and check to see if the operating state is off.
- Use the **show interface ethx/y priority-flow-control** command and check to see the if operating state is off.

Solution

Use the following commands under **interface ethx/y** to enable link pause:

- **flowcontrol receive on**
- **flowcontrol send on**

Registers and Counters

The following are the commands to access various registers and counters:

Nexus 5000 10G PFC

Use the following command:

```
show hard in gatos asic <gatos_num> registers match mm_CFG_pause$
```

Nexus 5000 1G storm control

Use the following commands:

```
show plat fwm info lif eth1/1
show plat fwm info pif eth1/1
debug hardware internal gatos asic 0 dump-mem 0x3b9000 20
```

Nexus 5000 10G storm control

Use the following commands:

```
show plat fwm info lif eth1/5
show plat fwm info pif eth1/5
debug hardware internal gatos asic 1 dump-mem 0x3b9000 20
```

Nexus 5000 storm control counter

Use the following command:

```
show hardware internal gatos asic 1 counters rx_db 2 | grep storm
```

afm-related CLI commands and tools

Commands	Purpose
`show platform afm in att br`	Shows which features or groups are attached to which interface.
`show platform afm in att global`	Shows the IDs of policies including QoS Policies (printed as NP Policies) attached on the global interface.
`show platform afm in att interface ethernet x/y`	Shows the IDs of policies including QoS Policies for an interface or PC.
`show platform afm in group id X asic Y`	Shows the TCAM entries for a particular group on a particular ASIC/GATOS.
`show platform afm in map-tbls`	Shows the internal mapping tables, such as the ext-cos to qos-group, qos-group to int-cos, and int-cos to class_id maps.

FEX qosctrl debug commands

Command	Purpose
show platform software qosctrl port 0 0 nif <0-48> [sat|switch]	Displays the PI information for every port. (Useful if port level configuration exists.)
show platform software qosctrl port 0 0 hif <0-48> [sat|switch]	Displays the PI information for every port. (Useful if port level configuration exists.)
show platform software qosctrl policy hif	Displays the global network-qos and queueing configurations.
show platform software qosctrl global	Global PI level configurations.
show platform software qosctrl pss	Stores PSS information.
show platform software qosctrl asic *<mod> <asic>*	Displays per asic level port details.
show platform software qosctrl default port *<mod> <asic>*	Displays default port settings on FEX ports.
show platform software qosctrl port *<mod> <asic> <port-type> <port>*	Displays per-port level PI and PD data structures.

N2K-C2148T FEX counters

Note Use the following commands (in the FEX shell) in preparation to display the statistics of MAC level traffic and pause statistics:
- **show plat soft fex info satport** <*fex-interface-id*> (for mapping except in the case of NIF in RW6)
- **show plat soft redwood sts**
- **show plat soft redwood ss**

Command	Purpose
show platform software qosctrl port 0 6 hif 1 counters	Displays counters.
show plat soft redwood rmon 6 nif0	Displays statistics of MAC level traffic and pause statistics of NIF of eth103/1/37.
show plat soft redwood rmon 6 hif5	Displays statistics of MAC level traffic and pause statistics of iHIF for eth103/1/37.

Command	Purpose
show plat soft redwood rmon 4 nif1	Displays statistics of MAC level traffic and pause statistics of iNIF for eth103/1/37.
show plat soft redwood rmon 4 hif5	Displays statistics of MAC level traffic and pause statistics of HIF for eth103/1/37.
show plat soft redwood ss	Displays mapping of HIF/NIF to SS.
show plat soft redwood ss 4 3	Displays statistics of RW4 SS3 - Host Receive from HIF4-7 to NIF0-3
show plat soft redwood ss 4 2	Displays statistics of RW4 SS2 - Host Receive from HIF0-3 to NIF0-3
show plat soft redwood rate	Displays overall statistics for non-zero traffic.
show plat soft redwood rmon 6 cif0	Helps debug traffic going from CIF to CPU.
show plat soft qosctrl port 0 6 cif 0 counters	Helps debug traffic going from CIF to CPU.

Nexus 5000 multicast-optimization

Use the following commands:

```
show plat fwm in mco-info
show plat fwm in vlan 1 all_macgs
```

Nexus 5000 FCoE classification

- For the FCoE interface, use the following commands:

```
show plat fwm info pif ethernet 1/1 | grep gatos
debug platform hardware peek lu 7 index 5 pifTable
```

- For the FC interface, use the following commands.
(The first command is used to get the gatos number and the fc number.)

show platform fwm info pif fc *<id>*
debug peek lu *<gatos>* **index** *<fc num>* **pifTable**

Nexus 5000 MTU programming

Use the following command:

```
show hardware internal gatos asic 0 registers match bm_port_CFG.*_max
```

Nexus 5000 interrupt

Use the following commands:

```
debug hardware internal gatos asic 0 clear-interrupt
```

```
show hardware internal gatos asic 0 interrupt
show hardware internal gatos event-history errors
```

Untagged COS

Use the following commands:

```
sh platform afm info attachment interface eth3/1
sh system internal ipqos port-node eth3/1
```

Buffer usage and packet drop debugging on N2K-C2232P FEX

Use the following command:

```
show platform software qosctrl asic 0 0
```