Configure QoS on a UCS and Nexus 5000

Contents

Introduction

This document describes the configuration of Quality of Service (QoS) within the Unified Computing System (UCS) and Nexus devices.

Prerequisites

Requirements

There are no specific requirements for this document.

Components Used

The information in this document is based on these software and hardware versions:

- UCS Fabric Interconnect (FI) 6100, and 6200
- Nexus 5000 and 5500

The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. If your network is live, ensure that you understand the potential impact of any command.

Background Information

This document is about UCS(6100 and 6200 Fabric Interconnects) and Nexus(5000 and 5500) QoS specifically related to FlexPod and vBlock.

Terminology used in this documentation that relates to QoS.

CoS = Class of Service = 802.1p = 3 bits in .1q header on each packet to tell the switch how to classify.

QoS = Quality of Service = How the switch handles each Class of Service.

MTU = Maximum Transmission Unit = Maximum size of a frame/packet allowed on the switch. The most common and default (normal is what the below UCS screenshot shows) is 1500.

Configure

UCS QoS out-of-the-box

The UCS QoS settings for reference(UCSM / LAN / QoS System Class):

Priority	Enabled	CoS	Packet Drop	Weight	Weight (%)	MTU	Multicast Optimized
Platinum	☐	5	☐	10	N/A	normal	☐
Gold	☐	4	☑	9	N/A	normal	☐
Silver	☐	2	☑	8	N/A	normal	☐
Bronze	☐	1	☑	7	N/A	normal	☐
Best Effort	☑	Any	☑	5	50	normal	☐
Fibre Channel	☑	3	☐	5	50	fc	N/A

Note: Best Effort and Fibre Channel are grayed-out and cannot be disabled within UCS.

Default QoS Configuration

```
P10-UCS-A(nxos)# show running-config ipqos
logging level ipqosmgr 2
class-map type qos class-fcoe
class-map type queuing class-fcoe
  match qos-group 1
class-map type queuing class-all-flood
  match qos-group 2
class-map type queuing class-ip-multicast
  match qos-group 2
policy-map type qos system_qos_policy
  class class-fcoe
    set qos-group 1
  class class-default
policy-map type queuing system_q_in_policy
  class type queuing class-fcoe
    bandwidth percent 50
  class type queuing class-default
    bandwidth percent 50
policy-map type queuing system_q_out_policy
  class type queuing class-fcoe
    bandwidth percent 50
  class type queuing class-default
    bandwidth percent 50
class-map type network-qos class-fcoe
  match qos-group 1
class-map type network-qos class-all-flood
  match qos-group 2
class-map type network-qos class-ip-multicast
  match qos-group 2
policy-map type network-qos system_nq_policy
  class type network-qos class-fcoe
    pause no-drop
    mtu 2158
  class type network-qos class-default
system qos
  service-policy type qos input system_qos_policy
  service-policy type queuing input system_q_in_policy
  service-policy type queuing output system_q_out_policy
  service-policy type network-qos system_nq_policy
```

Relevant Information:

- qos-group is how the switch internally treats a given CoS. Think of qos-group as a bucket or lane which each packet goes into.
- Best Effort does not get an explicit qos-group, so it default to qos-group 0
- Fibre Channel over Ethernet (FCoE) has CoS 3 and gets put into qos-group 1

CoS <=> qos-group cheat sheet

	CoS	qos-group
Platinum	5	2
Gold	4	3
Silver	2	4
Bronze	1	5
Best Effort	Any	0
Fibre Channel	3	1

CoS can be changed to CoS 6 on UCS. CoS 7 is reserved for internal UCS communications.

show queuing interface command

```
P10-UCS-A(nxos)# show queuing interface
Ethernet1/1 queuing information:
  TX Queuing
    qos-group  sched-type  oper-bandwidth
        0       WRR             50
        1       WRR             50

  RX Queuing
    qos-group 0
    q-size: 360640, HW MTU: 1500 (1500 configured)
    drop-type: drop, xon: 0, xoff: 360640
    Statistics:
        Pkts received over the port             : 27957
        Ucast pkts sent to the cross-bar        : 0
        Mcast pkts sent to the cross-bar        : 27957
        Ucast pkts received from the cross-bar  : 0
        Pkts sent to the port                   : 347
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)

    qos-group 1
    q-size: 79360, HW MTU: 2158 (2158 configured)
    drop-type: no-drop, xon: 20480, xoff: 40320
    Statistics:
        Pkts received over the port             : 0
        Ucast pkts sent to the cross-bar        : 0
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 0
        Pkts sent to the port                   : 0
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)

  Total Multicast crossbar statistics:
    Mcast pkts received from the cross-bar      : 347
```

This output shows how this interface queues each class.

Information about the switchport Ethernet 1/1:

- Best Effort gets qos-group 0 and a q-size of 360640 bytes of buffers and a MTU of 1500.
- This port has ingressed/received 27957 packets of Best Effort and egressed/sent 347 packets.
- "Pkts discarded on ingress" is the number of packets which have been received but during that instant the buffer(q-size) was full and the switch decided to discard, this is also known as tail drop.

IOM port que

Show queuing interface for the Input and Output Modules (IOM) ports in the UCS chassis:

```
Ethernet1/1/1 queuing information:
```

```
  Input buffer allocation:
  Qos-group: 1
  frh: 3
  drop-type: no-drop
  cos: 3
  xon       xoff      buffer-size
  ---------+---------+-----------
  8960      14080     24320

  Qos-group: 0
  frh: 8
  drop-type: drop
  cos: 0 1 2 4 5 6
  xon       xoff      buffer-size
  ---------+---------+-----------
  0         117760    126720

  Queueing:
  queue    qos-group    cos            priority  bandwidth mtu
  --------+------------+--------------+---------+---------+----
  2        0            0 1 2 4 5 6     WRR        50      1600
  3        1            3               WRR        50      2240

  Queue limit: 66560 bytes

  Queue Statistics:
  queue  rx              tx
  ------+---------------+---------------
  2      18098           28051
  3      0               0

  Port Statistics:
  rx drop         rx mcast drop   rx error        tx drop         mux ovflow
  --------------+---------------+---------------+---------------+--------------
  0               0               0               0                InActive

  Priority-flow-control enabled: yes
  Flow-control status:
  cos     qos-group   rx pause  tx pause  masked rx pause
  -------+-----------+---------+---------+---------------
  0              0    xon       xon       xon
  1              0    xon       xon       xon
  2              0    xon       xon       xon
  3              1    xon       xon       xon
  4              0    xon       xon       xon
  5              0    xon       xon       xon
  6              0    xon       xon       xon
  7            n/a    xon       xon       xon
```

There are qos-group 0 and qos-group 1, qos-group 0 gets packets marked with cos 0 1 2 4 5 6, and qos-group 1 get cos 3. The buffer-size on Fabric Extender (FEX)/IOMs is a bit smaller and is only 126720 bytes. The FEX does QoS slightly differently and takes multiple qos-groups and bundles them into a queue. The rx and tx counters for each queue can be seen.

show interface priority-flow-control

The last output to check out is: **show interface priority-flow-control**

```
P10-UCS-A(nxos)# show interface priority-flow-control
================================================================
Port               Mode Oper(VL bmap)  RxPPP      TxPPP
```

```
================================================================

Ethernet1/1        Auto Off           0          0
Ethernet1/2        Auto Off           0          0
Ethernet1/3        Auto Off           0          0
Ethernet1/4        Auto Off           6          0
Ethernet1/5        Auto Off           0          0
Ethernet1/6        Auto Off           0          0
Ethernet1/7        Auto Off           0          0
Ethernet1/8        Auto Off           0          0
Ethernet1/9        Auto Off           0          0
Ethernet1/10       Auto Off           2          0
..snip..
Vethernet733       Auto Off           0          0
Vethernet735       Auto Off           0          0
Vethernet737       Auto Off           0          0
Ethernet1/1/1      Auto On  (8)       0          0
Ethernet1/1/2      Auto Off           0          0
Ethernet1/1/3      Auto On  (8)       0          0
Ethernet1/1/4      Auto Off           0          0
```

This shows on what interfaces Priority Flow Control (PFC) negotiates(Auto On) and what interfaces PFC does not negotiate(Auto Off). PFC is a way for a switch to ask a neighbor switch to not send packets of a specific CoS for a short amount of time. PFC pauses(PPP,per priority pause) occur when the buffers are full/almost full. The output of `show cdp neighbors` and `show fex details` tells us this Ethernet 1/1-4 is down to the FEX/IOM of Chassis 1 and Ethernet 1/9-10 is up to the Nexus 5000. In this output 6 pauses were sent down to the FEX/IOM on Ethernet 1/4 and 2 pauses have been sent out Ethernet1/10 to the upstream Nexus 5000.

- PPPs themselves ARE NOT A BAD THING!

  Note: Since the FEX/IOM are not really switches PFC does NOT negotiate between them on Ethernet1/1-4 but can negotiate to the endpoint Ethernet1/1/1. The PPPs sent to a FEX/IOM are sent along out the remote switchport Ethernet1/1/1.

That's what UCS QoS looks like out of the box....

What if Silver is enabled?

This results in configuration:

```
class-map type qos class-fcoe
class-map type qos match-all class-silver
  match cos 2
class-map type queuing class-silver
  match qos-group 4
class-map type queuing class-all-flood
  match qos-group 2
class-map type queuing class-ip-multicast
  match qos-group 2
policy-map type qos system_qos_policy
  class class-silver
    set qos-group 4
policy-map type queuing system_q_in_policy
class type queuing class-silver bandwidth percent 44
  class type queuing class-fcoe
    bandwidth percent 29 class type queuing class-default bandwidth percent 27 policy-map type
```

```
queuing system_q_out_policy class type queuing class-silver bandwidth percent 44
  class type queuing class-fcoe
    bandwidth percent 29 class type queuing class-default bandwidth percent 27 policy-map type
queuing org-root/ep-qos-Default-Qos class type queuing class-fcoe class type queuing class-
default bandwidth percent 50 shape 40000000 kbps 10240 class-map type network-qos class-silver
match qos-group 4class-map type network-qos class-all-flood match qos-group 2 class-map type
network-qos class-ip-multicast match qos-group 2 policy-map type network-qos system_nq_policy
class type network-qos class-silver
  class type network-qos class-fcoe
    pause no-drop
    mtu 2158
  class type network-qos class-default
system qos
  service-policy type qos input system_qos_policy
  service-policy type queuing input system_q_in_policy
  service-policy type queuing output system_q_out_policy
  service-policy type network-qos system_nq_policy

Ethernet1/1 queuing information:
  TX Queuing
    qos-group  sched-type  oper-bandwidth
        0       WRR             27
        1       WRR             29
        4       WRR             44

  RX Queuing
    qos-group 0
    q-size: 308160, HW MTU: 9216 (9216 configured)
    drop-type: drop, xon: 0, xoff: 301120
    Statistics:
        Pkts received over the port             : 12
        Ucast pkts sent to the cross-bar        : 12
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 17
        Pkts sent to the port                   : 17
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)

    qos-group 1
    q-size: 79360, HW MTU: 2158 (2158 configured)
    drop-type: no-drop, xon: 20480, xoff: 40320
    Statistics:
        Pkts received over the port             : 7836003
        Ucast pkts sent to the cross-bar        : 7836003
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 4551954
        Pkts sent to the port                   : 4551954
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)

    qos-group 4
    q-size: 22720, HW MTU: 1500 (1500 configured)
    drop-type: drop, xon: 0, xoff: 22720
    Statistics:
        Pkts received over the port             : 0
        Ucast pkts sent to the cross-bar        : 0
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 0
        Pkts sent to the port                   : 0
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)
```

Notice the **Best Effort (qos-group 0)** q-size went from **360640 to 308160** because **Silver (qos-**

group 4) was allocated **22720** of buffers space.

What if Silver is made Jumbo?

Set MTU to 9216.

```
Ethernet1/1 queuing information:
  TX Queuing
    qos-group  sched-type  oper-bandwidth
        0       WRR             27
        1       WRR             29
        4       WRR             44

  RX Queuing
    qos-group 0
    q-size: 301120, HW MTU: 9216 (9216 configured)
    drop-type: drop, xon: 0, xoff: 301120
    Statistics:
        Pkts received over the port             : 3
        Ucast pkts sent to the cross-bar        : 3
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 0
        Pkts sent to the port                   : 0
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)

    qos-group 1
    q-size: 79360, HW MTU: 2158 (2158 configured)
    drop-type: no-drop, xon: 20480, xoff: 40320
    Statistics:
        Pkts received over the port             : 7842224
        Ucast pkts sent to the cross-bar        : 7842224
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 4555791
        Pkts sent to the port                   : 4555791
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)

    qos-group 4
    q-size: 29760, HW MTU: 9216 (9216 configured)
    drop-type: drop, xon: 0, xoff: 29760
    Statistics:
        Pkts received over the port             : 0
        Ucast pkts sent to the cross-bar        : 0
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 0
        Pkts sent to the port                   : 0
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)
```

Silver(**qos-group 4**) now gets **29760** q-size, up from 22720.

What if Silver is made no-drop?

Uncheck the **Packet Drop** setting?

```
Ethernet1/1 queuing information:
```

```
  TX Queuing
    qos-group  sched-type  oper-bandwidth
        0       WRR             27
        1       WRR             29
        4       WRR             44

  RX Queuing
    qos-group 0
    q-size: 240640, HW MTU: 9216 (9216 configured)
    drop-type: drop, xon: 0, xoff: 240640
    Statistics:
        Pkts received over the port             : 20
        Ucast pkts sent to the cross-bar        : 20
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 1
        Pkts sent to the port                   : 1
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)

    qos-group 1
    q-size: 79360, HW MTU: 2158 (2158 configured)
    drop-type: no-drop, xon: 20480, xoff: 40320
    Statistics:
        Pkts received over the port             : 7837323
        Ucast pkts sent to the cross-bar        : 7837323
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 4552726
        Pkts sent to the port                   : 4552726
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)

    qos-group 4
    q-size: 90240, HW MTU: 9216 (9216 configured)
    drop-type: no-drop, xon: 17280, xoff: 37120
    Statistics:
        Pkts received over the port             : 0
        Ucast pkts sent to the cross-bar        : 0
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 0
        Pkts sent to the port                   : 0
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)
```

Notice the Silver (**qos-group 4**) **q-size** increases to **90240, drop-type** changes to **no-drop**, and **Best Effort qos-group 0** is reduced to **240640.**

The Best Effort qos-group 0 buffer space is reallocated to other QoS classes.

Upstream Nexus 5000

Nexus 5000 default qos configs are similar but not exact.

show running-config ipqos

```
Ethernet1/1 queuing information:
  TX Queuing
    qos-group  sched-type  oper-bandwidth
        0       WRR             27
        1       WRR             29
```

```
        4       WRR             44

  RX Queuing
    qos-group 0
    q-size: 240640, HW MTU: 9216 (9216 configured)
    drop-type: drop, xon: 0, xoff: 240640
    Statistics:
        Pkts received over the port             : 20
        Ucast pkts sent to the cross-bar        : 20
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 1
        Pkts sent to the port                   : 1
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)

    qos-group 1
    q-size: 79360, HW MTU: 2158 (2158 configured)
    drop-type: no-drop, xon: 20480, xoff: 40320
    Statistics:
        Pkts received over the port             : 7837323
        Ucast pkts sent to the cross-bar        : 7837323
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 4552726
        Pkts sent to the port                   : 4552726
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)

    qos-group 4
    q-size: 90240, HW MTU: 9216 (9216 configured)
    drop-type: no-drop, xon: 17280, xoff: 37120
    Statistics:
        Pkts received over the port             : 0
        Ucast pkts sent to the cross-bar        : 0
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 0
        Pkts sent to the port                   : 0
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)
```

The Nexus 5000 hides default options so **show running-config ipqos all** is required to see the whole configuration.

show queuing interface

```
Ethernet1/1 queuing information:
  TX Queuing
    qos-group  sched-type  oper-bandwidth
        0       WRR             27
        1       WRR             29
        4       WRR             44

  RX Queuing
    qos-group 0
    q-size: 240640, HW MTU: 9216 (9216 configured)
    drop-type: drop, xon: 0, xoff: 240640
    Statistics:
        Pkts received over the port             : 20
        Ucast pkts sent to the cross-bar        : 20
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 1
        Pkts sent to the port                   : 1
        Pkts discarded on ingress               : 0
```

```
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)

    qos-group 1
    q-size: 79360, HW MTU: 2158 (2158 configured)
    drop-type: no-drop, xon: 20480, xoff: 40320
    Statistics:
        Pkts received over the port             : 7837323
        Ucast pkts sent to the cross-bar        : 7837323
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 4552726
        Pkts sent to the port                   : 4552726
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)

    qos-group 4
    q-size: 90240, HW MTU: 9216 (9216 configured)
    drop-type: no-drop, xon: 17280, xoff: 37120
    Statistics:
        Pkts received over the port             : 0
        Ucast pkts sent to the cross-bar        : 0
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 0
        Pkts sent to the port                   : 0
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)
```

show interface priority-flow-control

The ports down to the UCS (Ethernet1/1 - 2) have PFC off(Auto Off).

```
Ethernet1/1 queuing information:
  TX Queuing
    qos-group  sched-type  oper-bandwidth
        0       WRR             27
        1       WRR             29
        4       WRR             44

  RX Queuing
    qos-group 0
    q-size: 240640, HW MTU: 9216 (9216 configured)
    drop-type: drop, xon: 0, xoff: 240640
    Statistics:
        Pkts received over the port             : 20
        Ucast pkts sent to the cross-bar        : 20
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 1
        Pkts sent to the port                   : 1
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)

    qos-group 1
    q-size: 79360, HW MTU: 2158 (2158 configured)
    drop-type: no-drop, xon: 20480, xoff: 40320
    Statistics:
        Pkts received over the port             : 7837323
        Ucast pkts sent to the cross-bar        : 7837323
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 4552726
        Pkts sent to the port                   : 4552726
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)
```

```
    qos-group 4
    q-size: 90240, HW MTU: 9216 (9216 configured)
    drop-type: no-drop, xon: 17280, xoff: 37120
    Statistics:
        Pkts received over the port             : 0
        Ucast pkts sent to the cross-bar        : 0
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 0
        Pkts sent to the port                   : 0
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)
```

Add FCoE to the configuration

These policies are there by default on the Nexus 5000 but not enabled, so just need to use them.

```
Ethernet1/1 queuing information:
  TX Queuing
    qos-group  sched-type  oper-bandwidth
        0       WRR             27
        1       WRR             29
        4       WRR             44

  RX Queuing
    qos-group 0
    q-size: 240640, HW MTU: 9216 (9216 configured)
    drop-type: drop, xon: 0, xoff: 240640
    Statistics:
        Pkts received over the port             : 20
        Ucast pkts sent to the cross-bar        : 20
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 1
        Pkts sent to the port                   : 1
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)

    qos-group 1
    q-size: 79360, HW MTU: 2158 (2158 configured)
    drop-type: no-drop, xon: 20480, xoff: 40320
    Statistics:
        Pkts received over the port             : 7837323
        Ucast pkts sent to the cross-bar        : 7837323
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 4552726
        Pkts sent to the port                   : 4552726
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)

    qos-group 4
    q-size: 90240, HW MTU: 9216 (9216 configured)
    drop-type: no-drop, xon: 17280, xoff: 37120
    Statistics:
        Pkts received over the port             : 0
        Ucast pkts sent to the cross-bar        : 0
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 0
        Pkts sent to the port                   : 0
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)
```

show interface priority-flow-control

The ports down to the UCS (Ethernet1/1 - 2) have PFC on(Auto On).

```
Ethernet1/1 queuing information:
  TX Queuing
    qos-group  sched-type  oper-bandwidth
        0       WRR             27
        1       WRR             29
        4       WRR             44

  RX Queuing
    qos-group 0
    q-size: 240640, HW MTU: 9216 (9216 configured)
    drop-type: drop, xon: 0, xoff: 240640
    Statistics:
        Pkts received over the port             : 20
        Ucast pkts sent to the cross-bar        : 20
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 1
        Pkts sent to the port                   : 1
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)

    qos-group 1
    q-size: 79360, HW MTU: 2158 (2158 configured)
    drop-type: no-drop, xon: 20480, xoff: 40320
    Statistics:
        Pkts received over the port             : 7837323
        Ucast pkts sent to the cross-bar        : 7837323
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 4552726
        Pkts sent to the port                   : 4552726
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)

    qos-group 4
    q-size: 90240, HW MTU: 9216 (9216 configured)
    drop-type: no-drop, xon: 17280, xoff: 37120
    Statistics:
        Pkts received over the port             : 0
        Ucast pkts sent to the cross-bar        : 0
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 0
        Pkts sent to the port                   : 0
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)
```

PFC

PFC(802.1Qbb) is how Nexus/UCS devices create a lossless fabric as part of Data Center Bridging(DCBX). FCoE requires a lossless fabric, multi-hop FCoE is especially prone to this configuration issue. The upstream switch, typically a Nexus 5000, must match QoS settings configured on UCS.

As stated previously PFC is a way for switches to notify neighbor switches to stop to send additional frames. Think about this in the context of a multiple switch network environment with traffic which goes many directions at once, not only does this add buffers of path1(source1/destintion1) this is multiplying buffers because the neighbor switch likely has traffic that ingress multiple ports(multiple buffers). While PFC isn't required when you use IP storage it often helps dramatically improve performance due to this buffer multiplication effect which

prevents unnecessary packet loss.

An excellent PFC/DCBX overview.

Why does PFC NOT negotiate?

No-drop QoS policy must match on each side.

If a QoS class is defined on one switch as no-drop and not as no-drop on the other, PFC does not negotiate. Since UCS configures Platinum as no-drop but disabled out-of-the-box, this often occurs when Platinum is enabled.

System qos must match on each side

If the queuing input and queuing output and qos input don't match, PFC does not negotiate.

NetApp

Gold

NetApp filers by default send ALL IP storage traffic which are VLAN tagged by the NetApp in CoS 4(Gold). As the CoS bits are in the .1q header when the NetApp is connected to an access port NetApp traffic is put into Best Effort.

Asymmetric QoS

A common configuration mistake is to choose another CoS color(Silver) to put Network File System NFS traffic from UCS into and return NFS traffic from a NetApp is put into Gold. So the traffic is something like:

Server Send	UCS Silver >	Nexus 5k Silver >	NetApp Best Effort
Receive	<Gold	<Gold	<Gold

If UCS were configured for Silver to be Jumbo but NOT Gold, this would cause problems.

Undefined QoS

When a QoS class(Platinum/Gold/Silver/Bronze) is NOT enabled, UCS and Nexus devices treat those packets as Best Effort and put them into qos-group 0.

Server Send	UCS Silver >	Nexus 5k Best Effort >	NetApp Best Effort
Receive	<Gold	<Best Effort	<Gold

Note: the CoS bits on the packet are NOT changed/remarked, but the packets are treated differently.

Virtual Computing Environment (VCE) QoS

VCE QoS design is less than ideal.

	Nexus 1k	UCS	Nexus 5k
BE / CoS 0	1500	1500	1600
FC / CoS 1	-	2158(no-drop)	-
CoS 6	mgmt	-	-
Platinum / CoS 5	-	1500(no-drop)	1500
Gold / CoS 4	vmotion	1500	1500
Silver / CoS 2	NFS	-	9216(no-drop)

If you have classes of CoS defined at one level, but ignored at another level is complicated and could make things not work the way it was intended. For instance VCE uses Silver for NFS, but if UCS doesn't have Silver defined this traffic is queued in Best Effort which isn't Jumbo and can cause NFS traffic to be dropped or fragmented. PFC isn't negotiated due to the mismatches in no-drop policies, but evidently this is OK because PFC isn't required for ethernet.

Shallow Buffers

Internet Protocol (IP) based storage protocols are all very bursty protocols and often configured with 9000 MTU. As such they perform poorly in Platinum/Gold/Silver/Bronze due to the 29760 q-size / 9000 MTU only allows 3 packets into the buffer before tail-drop is caused.

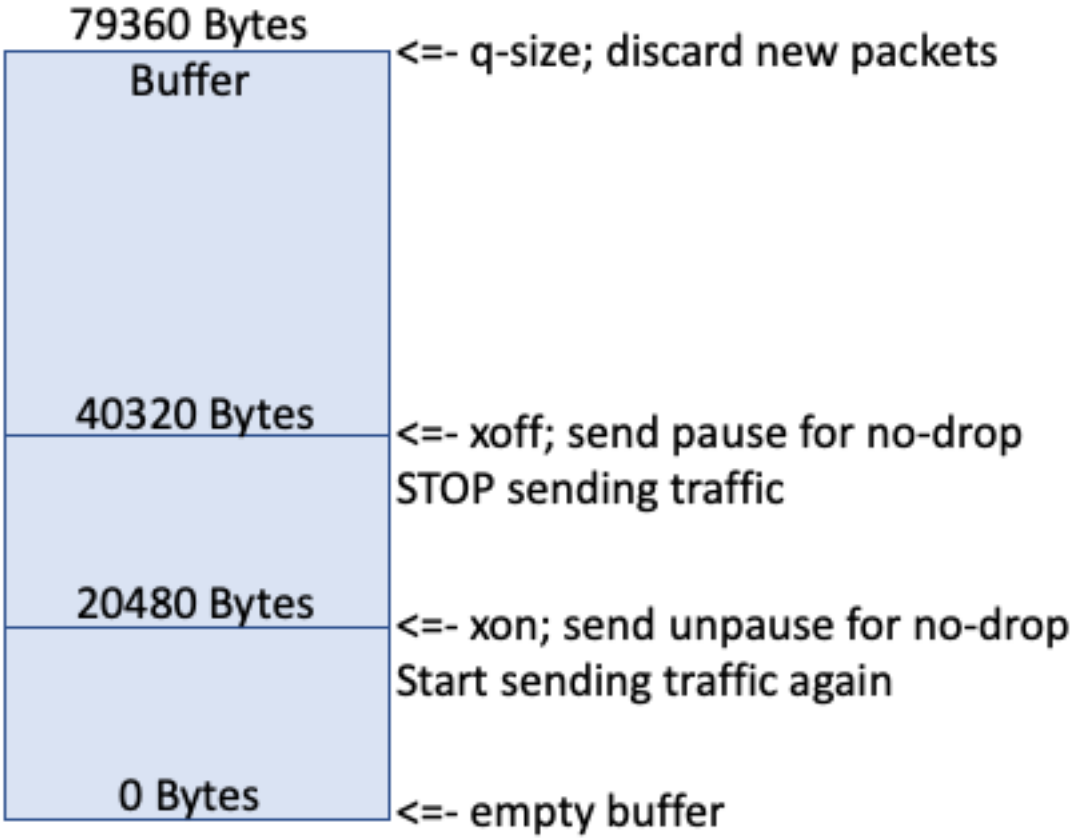
Bigger Buffers

UCS Ethernet policy allows the vNIC buffers (ring size) to be increased. The default is 512 and the maximum is 4096. if you change this value to the maximum, the full buffer latency(##KB / 10Gbps) increases from 0.4ms to 3.2ms. So changes in this buffer allows for fewer drops, but at the expense of increased latency.

9216 MTU vs 9000 MTU

The point of the configuration of **Jumbo Frames** is to allow an endpoint device to talk to another endpoint device with 9000 byte layer 3 packets. When layer 2 encapsulation techniques are used the switches and routers between the endpoint devices need to be able to handle slightly larger layer 2 frames than 9000 MTU layer 3 packets to account for the encapsulation overhead. When in doubt allow 9216 MTU on switches.

PFC and PPP



As new packets are queued, the buffer fills.

When buffer gets to 20k, the buffer continues to fill.

When buffer gets to 40k, the switch sends a PPP pause if this queue is no-drop, whixh indicates the remote switch to stop to send traffic.

Ideally the remote side soon stops to send traffic and the remainder of the buffer (79360-40320) holds incoming in-flight packets.

"Pkts discarded on ingress" counters increments when the buffer is full.

FC and FCoE is a lossless protocol in an ideal situation where the remote switch stops to send traffic and buffer levels eventually fall and reache 20k. The switch sends another PPP unpause for this no-drop queue which tells the remote switch to start to send traffic again.

Troubleshoot

There is currently no specific troubleshoot information available for this configuration.

Related Information

- **UCS Manager Network Management Guide, Release 4.0**
- **Nexus 5000 Series Quality of Service Configuration Guide**
- **UCS with VMware Esxi End to End Jumbo MTU Configuration Example**
- **Technical Support & Documentation - Cisco Systems**