Understanding Queuing and Scheduling QoS on Catalyst 4000 Supervisor III and IV

Document ID: 21389

Contents

Introduction

Catalyst 4000 with Supervisor III (WS–X4014) or Supervisor IV (WS–X4515) supports advanced Quality of Service (QoS) features including classification, policing, marking, queuing, and scheduling. This document addresses the queuing and scheduling features, including traffic shaping, sharing, and strict priority /low latency queuing. Queuing determines how packets are queued in various queues in the egress interface, and scheduling determines how (in times of congestion) high–priority traffic is given preference over low–priority traffic.

Before You Begin

Conventions

For more information on document conventions, see the Cisco Technical Tips Conventions.

Prerequisites

Readers of this document should be knowledgeable of the following:

- Layer 2 (L2) prioritization of frames is based on a Class of Service (CoS) value, which is available in the InterSwitch Link (ISL) header (three least significant bites in 4–bit user field) and 802.1Q header (three most significant bits in 2–byte tag control information field).
- Layer 3 (L3) prioritization of packets is based on Differentiated Services Code Point (DSCP) value, which is available in the Type of Service (ToS) byte in the IP header (six most significant bits) or IP precedence value in the ToS byte (three most significant bits).
- Refer to the software configuration guide for additional configuration assistance.

Components Used

The information in this document is based on the following software versions on a Supervisor III (WS–X4014):

- Cisco IOS® Software Release 12.1(8)EW

Note: Supervisor IV is first supported on Cisco IOS Software Release 12.1(12c)EW. The features described in this document apply to Supervisor IV as well, unless otherwise explicitly differentiated.

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

Queuing

The Catalyst 4000 Supervisor III and IV use a shared memory switching architecture and is able to provide queuing and scheduling features to the existing line cards. Since the Supervisor provides non−blocking switching architecture, there is no input queuing. The packets are forwarded through the backplane to the output or egress port. The output side of the interface provides four transmit queues. Queue size is currently fixed at 240 packets for FastEthernet ports, and 1920 packets for non−blocking Gigabit Ethernet interfaces. Non−blocking means the ports are not over−subscribed in the connection to the backplane. A list of non−blocking Gigabit Ethernet ports is as follows:

- uplink ports on Supervisor Engine III (WS−X4014) and IV (WS−X4515)
- ports on the WS−X4306−GB linecard
- two 1000BASE−X ports on the WS−X4232−GB−RJ linecard
- first two ports on the WS−X4418−GB linecard
- two 1000BASE−X ports on the WS−X4412−2GB−TX linecard

Blocking (over−subscribed) Gigabit Ethernet port queue size is currently fixed at 240 packets as well. Blocking ports are listed as follows:

- 10/100/1000 T ports on the WS−X4412−2GB−TX linecard
- ports on WS−4418−GB linecard, except for the first two ports
- ports on the WS−X4424−GB−RJ45 linecard
- ports on the WS−X4448−GB−LX linecard
- ports on the WS−X4448−GB−RJ45 linecard

Note: The queue size is based on the number of packets and not the size of the packets. Currently, Supervisor III does not support any congestion avoidance mechanism such as Weighted Random Early Detection (WRED) for the transmit queues.

Note: Supervisor IV supports the Active Queue Management (AQM) feature in Cisco IOS release 12.1(13)EW and later. AQM is a congestion avoidance technique that acts before buffer overflow occurs. AQM is achieved through Dynamic Buffer Limiting (DBL). DBL tracks the queue length for each traffic flow in the switch. When the queue length of a specific flow exceeds its limit, DBL will drop packets or set the Explicit Congestion Notification (ECN) bits in the packet headers. For more information on how to configure DBL, refer to Configuring QoS.

When QoS is disabled, the packets are trusted for incoming DSCP on the ingress ports and queued to the appropriate queues. These queues are serviced round−robin.

When QoS is enabled, packets are queued based on the internal DSCP, which is derived either from incoming CoS/DSCP using port trust states, or a CoS/DSCP default configuration on the input port or Access List (ACL)/ Class−Based marking. The queue is selected based on the global DSCP − tx−queue mapping, which is fully configurable. The mapping can be displayed as follows:

```
Switch#show qos maps dscp tx-queue
```

```
        DSCP-TxQueue Mapping Table (dscp = d1d2)
        d1 : d2 0  1  2  3  4  5  6  7  8  9
        ------------------------------------
        0 :    01 01 01 01 01 01 01 01 01 01
        1 :    01 01 01 01 01 01 02 02 02 02
        2 :    02 02 02 02 02 02 02 02 02 02
        3 :    02 02 03 03 03 03 03 03 03 03
        4 :    03 03 03 03 03 03 03 03 04 04
        5 :    04 04 04 04 04 04 04 04 04 04
        6 :    04 04 04 04
```

The above mapping is the default mapping. If needed, mapping can be changed by issuing the **qos map dscp** *dscp−values* **to tx−queue** *queue−id* command. For example, to map a DSCP value of 50 to tx−queue 2, the following configuration is made in global configuration mode:

```
        Switch(config)#qos map dscp 50 to tx-queue 2
```

!--- You can verify to make sure the changes have been made.

```
        Switch #show qos maps dscp tx-queue
        DSCP-TxQueue Mapping Table (dscp = d1d2)
        d1 : d2 0  1  2  3  4  5  6  7  8  9
        ------------------------------------
        0 :    01 01 01 01 01 01 01 01 01 01
        1 :    01 01 01 01 01 01 02 02 02 02
        2 :    02 02 02 02 02 02 02 02 02 02
        3 :    02 02 03 03 03 03 03 03 03 03
        4 :    03 03 03 03 03 03 03 03 04 04
        5 :    02 04 04 04 04 04 04 04 04 04
        6 :    04 04 04 04
```

For further information on the configuration steps for changing the mapping, refer to the following document:

- Configuration Guide

Due to switch Application−Specific Integrated Circuit (ASIC) limitation, if the ingress port is set to trust−cos, the transmit CoS is equal to either the incoming packet CoS or default CoS (for untagged packets) configured on the port. If a policy is configured to set the DSCP for the packet by issuing the **set ip dscp value** command for such packets, they will be used as the source for internal DSCP instead of the default/ packet CoS, and queued in the appropriate queues. If the port is not trusted for CoS, the outgoing CoS would be based on the internal DSCP value.

Strict Priority Queuing /Low Latency Queuing

Transmit queue 3 can be configured as a strict priority queue if required so that the packets queued in that queue would be scheduled to be transmitted ahead of the packets queued in the rest of the queues, as long they do not exceed the configured share value. This is explained in the following section.

The strict priority feature is disabled by default. The default mapping would queue packets with CoS 4 and 5 and DSCP 32 through 47 in the transmit queue 3. DSCP to tx−queue mapping can be modified as desired so that the desired packets are queued in the high−priority queue.

In order not to starve the low−priority packets, this queue needs to be configured primarily for low volume, but high−priority traffic, such as voice traffic, and not for bulk low−priority TCP/IP traffic. It is also recommended to configure shaping/sharing for the high−priority queue if one needs to prevent starvation of

the other non–strict priority queues. By configuring shaping/sharing, the other low–priority packets will be scheduled once the shape/share value for strict queue has been met.

```
Switch#show run interface gigabitEthernet 1/1
interface GigabitEthernet1/1
 no switchport
 ip address 10.1.1.1 255.255.255.0
 tx-queue 3
 priority high
 end
```

Sharing Link Bandwidth

Catalyst 4000 Supervisor III and IV support the **bandwidth** command, which is a sub–command under the **tx–queue** command. This command allows a guaranteed minimum bandwidth to each of the four transmit queues. This command should not be confused with the interface level **bandwidth** command that is used for routing protocol purposes. This, along with DSCP–tx–queue mapping, provides for granular control of how much bandwidth is guaranteed for each class of traffic queued in each of the four queues. Typically, high–priority traffic such as voice traffic is guaranteed a certain minimum amount of traffic during times of congestion through the strict priority queuing, with a share configured for the transmit queue 3. Sharing of the link bandwidth is only supported on the non–blocking Gigabit Ethernet ports. This feature is currently not available on blocking Gigabit Ethernet ports or 10/100 FastEthernet interfaces.

When QoS is enabled globally on the switch, all four queues by default are assigned a minimum bandwidth of 250 Mbps on all ports. It may be necessary to change the default settings to make sure that they match the desired settings for the application or the network in question.

```
Switch#show run interface gigabitEthernet 1/1
interface GigabitEthernet1/1
 no switchport
 ip address 10.1.1.1 255.255.255.0
 tx-queue 1
   bandwidth 500 mbps
 tx-queue 2
   bandwidth 25 mbps
 tx-queue 3
   bandwidth 50 mbps
   priority high
 tx-queue 4
   bandwidth 200 mbps
 end

Switch#show qos interface GigabitEthernet 1/1

QoS is enabled globally
Port QoS is enabled
Port Trust State: 'untrusted'
Default DSCP: 0 Default CoS: 0
tx-Queue Bandwidth ShapeRate Priority QueueSize
         (bps)     (BPS)             (packets)
    1    500000000 disabled    N/A     1920
    2          250000000 disabled    N/A       1920
    3           50000000  disabled    high      1920
    4    200000000 disabled    N/A      1920
```

The switch currently does not validate that the sum of the bandwidth share per queue <= 1 Gbps. For example, if Q1 = 300 Mbps, Q2 = 200 Mbps, Q3= 100 Mbps, and Q4 = 500 Mbps, we are exceeding the 1 Gbps total bandwidth available for that interface. To understand how the switch would behave in this oversubscribed

scenario, we need an understanding of how the scheduling works.

When a transmit queue output rate is below its configured share and shape values, it is considered a high–priority queue. Initially, all the queues will be high–priority since none of them have been granted their share, and hence will be serviced in round–robin (note that a queue configured as high–priority will always be serviced first if it is not empty until it meets is share). Once some of the queues meet their share, if there are any more queues with high–priority, they will be serviced. If there are no high–priority queues, all the low–priority queues (queues which already have met their share) are serviced in round–robin.

Based on this above description of operation, in our example scenario, Q1, Q2, and Q3 would get their share, but not Q4 in times of congestion, as the interface can't allocate bandwidth more than its available physical bandwidth. Care should be exercised in choosing the share values according to the user/application requirements.

Traffic Shaping

Catalyst 4000 Supervisor III and IV support other traffic shaping features besides the policing feature. Shaping features can be configured per transmit queue on FastEthernet as well as Gigabit Ethernet. Shaping limits the bandwidth transmitted per queue per second to the configured maximum value configurable from 16 Kbps to 1 Gbps (100 Mbps for FastEthernet port). The shaping has very low variance from the configured value as the decision to transmit a packet from a specific queue is made per packet.

```
Switch#show run interface FastEthernet 5/9
interface FastEthernet5/9
 no switchport
 no snmp trap link-status
 ip address 10.1.1.1 255.255.255.0
 tx-queue 1
   shape 50 mbps
 tx-queue 2
   shape 35 mbps
 tx-queue 3
   priority high
   shape 5 mbps
 tx-queue 4
   shape 10 mbps

Switch#show qos interface FastEthernet 5/9

QoS is enabled globally
Port QoS is enabled
Port Trust State: 'untrusted'
Default DSCP: 0 Default CoS: 0
tx-Queue Bandwidth ShapeRate Priority QueueSize
        (BPS)      (BPS)            (packets)
  1       N/A      50000000   N/A      240
  2       N/A      35000000   N/A      240
  3       N/A       5000000high  240
  4       N/A      10000000   N/A      240
```

Transit Queue Scheduling

Packets are queued based on internal DSCP on one of the four queues described earlier. Internal DSCP can be derived from ingress DSCP, ingress port DSCP, or class–based marking. Transmit queue scheduling happens as follows. If shaping is configured, the packet in the transmit queue is checked whether it is within the configured maximum shape value. If it exceeds the value, it is queued and is not transmitted.

If the packet is eligible, the sharing/strict priority feature is considered. First, strict priority queued packets are given preference as long as they are below the configured shape parameter for the queue. After the strict priority queue is serviced (that is, no packets in strict priority queue or it has met its share), packets queued in the non−strict priority queue are serviced in round−robin. Since there are three such queues, sharing configured for those queues is again considered. For example, if the transmit queue 1 has not met its share, it has higher priority than transmit queue 2, which has met its share. Once such higher priority queue packets are dequeued, packets in the queues which have already met their share are considered.

Note: Higher priority in this context does not mean better DSCP, CoS, or IP precedence value. It is solely based on whether a particular queue has met its share or not. If the particular non−strict priority queue has not met its share, it is considered higher priority queue among non−strict priority queue which has met its share.

Related Information

- **Understanding and Configuring QoS**
- **Quality of Service on Catalyst 4000 Frequently Asked Questions**
- **QOS Policing and Marking with Catalyst 4000 Supervisor Engine 3**
- **LAN Product support**
- **LAN Switching Technology Support**
- **Technical Support – Cisco Systems**

Updated: Mar 24, 2005   Document ID: 21389