Understanding How Routing Updates and Layer 2 Control Packets Are Queued on an Interface with a QoS Service Policy

Document ID: 18664

Contents

Introduction

This document explains how routing protocol messages, such as hellos and database descriptors, as well as other important control traffic are queued when an outbound router interface is configured with a service–policy using the commands of the modular quality of service command–line interface (MQC).

Specifically, this document reviews these two mechanisms used by Cisco IOS® routers to prioritize control packets:

Field	Location	Where the Priority Is Considered
IP Precedence Bits	Type of service (TOS) byte in IP header	Provides priority through the network
pak_priority	Internal packet label inside the router, assigned by interface driver	Provides priority through the router (per–hop)

Both mechanisms are designed to ensure that key control packets are not dropped or are dropped last by the router and the queuing system when an outbound interface is congested.

Prerequisites

Requirements

There are no specific requirements for this document.

Components Used

The information in this document is based on Cisco IOS Software Release 12.2.

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, make sure that you understand the potential impact of any command.

Conventions

Refer to Cisco Technical Tips Conventions for more information on document conventions.

External Packet Prioritization Tags

Request for Comments (RFC) 791 defines the TOS byte in the header of an IP packet. Although RFC 2474 and RFC 2475 redefine this byte as Differentiated Services Code Point (DSCP) values, a Cisco IOS router still uses the original IP precedence bits of the TOS byte, as per RFC 791. Notice how the RFC defines the TOS byte:

"The Type of Service provides an indication of the abstract parameters of the quality of service desired. These parameters are to be used to guide the selection of the actual service parameters when transmitting a datagram through a particular network. Several networks offer service precedence, which somehow treats high−precedence traffic as more important than other traffic (generally by accepting only traffic above a certain precedence at time of high load)."

As illustrated in the diagram, the IP precedence field occupies the three most significant bits of the TOS byte. Only the three IP precedence bits reflect the priority or importance of the packet, not the full value of the TOS byte.
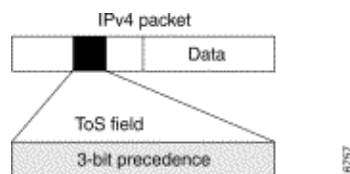
This table lists the values of the precedence bits:

Number	Bit Value	Name
0	000	Routine
1	001	Priority
2	010	Immediate
3	011	Flash
4	100	Flash Override
5	101	CRITIC/ECP
6	110	Internetwork Control

7	111	Network Control

Cisco IOS assigns an IP precedence of 6 to routing protocol packets on the control plane. As noted by RFC 791, "The Internetwork Control designation is intended for use by gateway control originators only." Specifically, Cisco IOS marks these IP−based control packets: Open Shortest Path First (OSPF), Routing Information Protocol (RIP), Enhanced Interior Gateway Routing Protocol (EIGRP) hellos, and keepalives. Telnet packets to and from the router also receive an IP precedence value of 6. The assigned value remains with the packets when the output interface transmits them into the network.

Internal Packet Prioritization Tags

While the IP precedence value specifies treatment of a datagram within its transmission *through the network*, the pak_priority mechanism specifies treatment of a packet during its transmission *inside the router*.

In addition to the core of a router CPU, every interface uses a network controller or local CPU, which runs a special piece of software called a driver. The driver code provides interface−specific instructions.

When it receives a packet, the interface driver copies the packet from a small first−in, first−out (FIFO) buffer to a data buffer in input/output (I/O) memory. It then attaches a small packet header to the buffer. The packet header, referred to in Cisco IOS terminology as the paktype structure, contains key information about the data block in the buffer. Dependent upon the contents of the packet, the packet header can point to the address in memory where the Ethernet encapsulation header, Internet Protocol (IP) header, and Transmission Control Protocol (TCP) header starts.

The Cisco IOS software uses the fields in the packet header to control the treatment of the packet in interface queues. The packet header includes the pak_priority flag, which indicates the relative importance of marked packets to the queuing system.

The RIP and OSPF routing processes that run on the core CPU of a router mark all traffic they originate with both IP precedence 6 and pak_priority. In contrast, the Border Gateway Protocol (BGP) instructs TCP to mark its traffic with IP precedence 6, but does not set pak_priority.

Cisco IOS must also ensure a low drop probability for several types of non−IP control packets. These packet types include these:

- Intermediate System−to−Intermediate System (IS−IS) routing protocol messages
- Enhanced Interior Gateway routing protocol (EIGRP) messages
- Point−to−Point Protocol (PPP) and high−level data link control (HDLC) keepalives on serial and packet over SONET (POS) interfaces
- Operations, administration, and maintenance (OAM) cells and address resolution protocol (ARP) messages on ATM interfaces

Since such traffic is not IP, Cisco IOS cannot match on the IP precedence value to provide prioritization. Instead, it uses only the internal pak_priority value in the packet buffer header.

Note: The Cisco Catalyst 6000 / Cisco 7600 Series initially supported the pak_priority mechanism on the FlexWAN only. Enhancements to the prioritization of IP and non−IP control packets was subsequently implemented.

Packet Prioritization Tags and Queuing

Routers such as the Cisco 7500 Route/Switch Processor (RSP) and lower−end routers (such as the Cisco 7200 and 3600 Series) use a different mechanism to route and control traffic than the Cisco 7500 Versatile Interface

Processor (VIP). This table summarizes the two approaches and assumes that a service–policy configured with the MQC is applied to the outbound interface.

Platform	Queuing of pak_priority Messages
Cisco 7500 Series (with distributed QoS and VIPs)	• Places pak_priority traffic in the class–default class queue by default or in a specifically configured (separate) queue. • When queued to class–default, packets go to the tail end of the queue. The pak_priority flag is used to avoid dropping the high–priority packets.
RSP–based QoS and other platforms, which include the Cisco 7200, 3600, 2600 Series	• Places pak_priority traffic in a separate set of queues other than class–default. (See the Understand Special Queues with Non–RSP Platforms section.) • Marks such messages with a special weight value (currently 1024).

In other words, on the Cisco 7500 Series, if an output service–policy is attached to the interface, the packets are classified with respect to the classes in that policy, and the pak_priority packet is placed at the end of the chosen class queue. If the pak_priority packet does not match any user defined class, it is placed at the tail of the class–default queue.

Note: With legacy queuing methods such as priority queuing and custom queuing or with a default interface FIFO queue, non–RSP routers enqueue pak_priority messages to the head of the queue to ensure both minimal latency and minimal drop probability.

Understand Special Queues With Non–RSP Platform

As noted in the Packet Prioritization Tags and Queuing table, Cisco router platforms like the Cisco 7200, 3600 and 2600 series place pak_priority messages into a separate set of queues and not the class–default set of queues.

There are three sets of queues on an interface:

- 2^n set of flow–based queues that consider such header values as the source and destination IP addresses. The actual number of queues is based on the bandwidth of the interface or virtual circuit. Refer to the description of the **fair–queue** command in the Cisco IOS Command Reference.
- Queues for user–created classes.
- Queues accessed on the basis of a hash of the linktype. For example, IP microflows are classified by the fair queuing system into queues based on a hash of the source and destination addresses and ports, TOS bits, and IP protocol number. Frame Relay local management interface (LMI) messages are queued based on a hash of the magic number that indicates that the message is LMI. Messages with the pak_priority flag go into these separate linktype queues.

This table lists the various queues and their conversation IDs (as seen in the output of the show policy−map interface or show queue commands) for an interface with greater than 512 Kbps of bandwidth.

Conversation / Queue Number	Type of Traffic
1 − 256	General flow−based traffic queues. Traffic that does not match to a user−created class matches to class−default and one of the flow−based queues.
257 − 263	**Reserved for Cisco Discovery Protocol and for packets marked with an internal high−priority flag.**
264	Reserved queue for the priority class (classes configured with the **priority** command). Look for the "Strict Priority" value for the class in the **show policy−map interface** output. The priority queue uses a conversation ID equal to the number of dynamic queues plus 8.
265 and higher	Queues for user−created classes.

Note: The values in this table are implementation−dependent and subject to change.

Prioritize IS−IS Packets

The Intermediate System to Intermediate System (IS−IS) routing control packets are a special case with respect to queuing and packet prioritization.

The IS−IS is the routing protocol for the Connectionless Network Protocol (CLNP) of the International Organization for Standardization (ISO). The developers of CLNP viewed TCP/IP as an interim protocol suite that the Open System Interconnection (OSI) suite eventually would replace. In order to support this predicted transition, Integrated IS−IS (or dual IS−IS) was created as an extension to IS−IS to provide a single routing protocol capable of routing both Connectionlesss−mode Network Service (CLNS) and IP. The protocol was designed to operate in a pure CLNS environment, pure IP environment, or a dual CLNS/IP environment.

Even when IS−IS is used to route only TCP/IP, IS−IS is still an ISO CLNP protocol. The packets by which IS−IS communicates with its peers are CLNS protocol data units (PDUs), which in turn means that even in an IP−only environment, the queuing system and Cisco IOS cannot use IP precedence to prioritize CLNS control messages. Instead, IS−IS packets receive priority through the pak_priority mechanism inside the router.

Configure a Queuing Strategy for Routing Packets

This section considers the three general approaches to designing a queuing strategy specifically to minimize the chances of dropped control packets under heavy congestion conditions on the Cisco 7500 series and VIPs. *(Recall that non−RSP platforms place control packets in separate queues by default.)*

Strategy	When to Use	Description of How to Configure
Match to a separate	Most conservative strategy. Ensures	Use the modular QoS CLI to configure a separate

queue.	little or no drops.	class and use the **bandwidth** command to assign a minimum bandwidth allocation to the matching traffic during periods of congestion. A class configured with the **bandwidth** command uses a scheduling "weight" based on the bandwidth and not on IP precedence. Refer to Understanding Class Based Weighted Fair Queuing on ATM.
Match to class–default with fair queuing.	Sufficient for most configurations. Some control packets can be dropped in the presence of congestion.	Use the IP precedence 6 automatically assigned by Cisco IOS to the packet to influence its weight and thus its share of the bandwidth. See Understanding Weighted Fair Queuing on ATM.
Match to class–default with FIFO queuing.	Not recommended for congested links. Some control packets can be dropped in the presence of congestion.	This approach does not consider IP precedence. With VIP–based QoS, the pak_priority messages are queued to the tail end of the FIFO queue.

This is an example of how to create a separate queue for RIP control packets.

```
 class-map match-all rp
   match access-group 104
!
 access-list 104 permit udp any eq rip any eq rip

!--- Create a class-map that matches an ACL permitting RIP.

!
 policy-map bandwidth
   class voip
     priority 64
   class bus
    bandwidth 184
   class RP
    bandwidth 8

!--- Create a policy-map (named "bandwidth") and specify
!--- class RP.

!
interface Serial1/0:0.1 point-to-point
  bandwidth 256
  ip unnumbered Loopback0
  ip accounting precedence input
  no cdp enable
```

```
        frame-relay class sample
        frame-relay interface-dlci 100 IETF

    !--- Apply the map-class named "sample" to the PVC.

    !
    map-class frame-relay sample
      frame-relay cir 256000
      frame-relay bc 2560
      frame-relay mincir 256000
      no frame-relay adaptive-shaping
      service-policy output bandwidth
      frame-relay fragment 160

    !--- Create a frame relay map-class and apply the service
    !--- policy inside the map-class.
```

Consider these factors when you choose one of these approaches:

- The particular routing protocol used and the configured timer values for hellos and database refresh
- The size of the database that needs to be exchanged and whether only updates/changes or full tables are refreshed periodically
- The amount of congestion expected at the interface or virtual circuit

In other words, consider the chances of actually queuing high−priority packets in the presence of congestion.

QoS and Locally Generated Packets

Traffic generated by the router represents a special case for outbound QoS service policies. Some locally generated traffic must be treated as any other user traffic, and the QoS system must apply the configured QoS mechanisms to this traffic. An example of such traffic is performance probes that are designed to measure the behavior incurred by packets of a given class. Other locally generated traffic, particularly Layer 2 keepalives and routing protocol messages, are vital to the basic functioning of the router and must not be subject to some QoS features. For example, weighted random early detection (WRED) must not drop Layer 2 keepalives when the average queue depth reaches a high watermark

In addition, packets destined to the router must be handled carefully. For example, remember that a service−policy that applies class−based policing must not apply to packets destined to the router to avoid dropping important control messages.

Note: As per design, RP generated packets are not accounted in Modular QoS CLI counters even though those packets are properly classified/queued. Those packets are not accounted in the **show policy−map interface** command output.

This table lists how packets destined to and from the router currently interact with key QoS features.

QoS Feature	Description
Class−Based Marking	- Originally worked only on Cisco Express Forwarding (CEF)−switched packets. - Support for process switching and fast switching methods is introduced in Cisco IOS Software Release 12.2(5) (CSCdt74738).

Policing	• *Inbound* – Rate limiting can be applied. Inbound interface must be configured with CEF if Committed Access Rate (CAR) (and not class–based policing) is used. On a Cisco 7500 series router, traffic policing can monitor CEF switching paths only. • *Outbound* – Rate limiting with CAR or class–based policing works.

Prioritize Packets on the Catalyst 6000

When you run Cisco IOS on both the supervisor and the MultiLayer Switch Feature Card (MSFC) in the Catalyst 6000, the RP marks routing control packets with IP precedence 6. This remarked value can be used with output scheduling to map the routing control packets to the high queue, high threshold in the weighted round robin (WRR) system. Such mapping of routing control packets sourced by the MSFC happens automatically as long as QoS is enabled globally with the **mls qos** command. If you enable QoS, it causes the system to set up all queuing parameters, such as WRED drop thresholds, WRR bandwidths, and queue limits. With QoS disabled globally, all packets are mapped to the low queue, low threshold for output scheduling, of WRR.

As noted in the Configuring QoS chapter of the Catalyst 6000 Configuration Guide, QoS supports classification, marking, scheduling, and congestion avoidance using Layer 2 class of service (CoS) values at Ethernet ingress ports. Classification, marking, scheduling, and congestion avoidance at Ethernet ingress ports do not use or set Layer 3 IP precedence or DSCP values. In addition, with any switching engine, QoS supports Ethernet egress port scheduling and congestion avoidance with Layer 2 CoS values. As a result, crucial IP and non–IP packets must be mapped to a CoS value, even if such values are used only internally as part of the data bus header. Crucial IP packets have their IP precedence value of 6 mapped to an equivalent CoS value of 6. Crucial non–IP packets, which include IS–IS packets that originate from the MSFC, are marked with the pak_priority flag and then such flagged packets are mapped to a CoS value of 6. This mapping happens automatically in current Cisco IOS releases.

Neither ingress policers nor egress policers mark packets sourced by the MSFC and destined for transmission through a physical Ethernet interface.

QoS configuration on the Catalyst 6000 is outside the scope of this document. Refer to Configuring QoS and the Catalyst LAN and ATM Switches Support Page for more information.

Related Information

- **QoS Support Pages**
- **Technical Support & Documentation – Cisco Systems**

Updated: Feb 15, 2008   Document ID: 18664