



Quality of Service Overview

This chapter explains quality of service (QoS) and the service models that embody it. It also suggests benefits that you can gain from implementing Cisco IOS XE QoS in your network. Then it describes the Cisco IOS XE QoS feature categories.

This chapter contains the following sections:

- [What Is Quality of Service?, page 1](#)
- [About QoS Architecture, page 2](#)
- [Who Could Benefit from Using QoS?, page 2](#)
- [Why Deploy QoS?, page 3](#)
- [End-to-End QoS Models, page 3](#)

What Is Quality of Service?

QoS refers to the ability of a network to provide improved service to selected network traffic over various underlying technologies including ATM, Ethernet and 802.1 networks, SONET, and IP-routed networks. In particular, QoS features provide improved and more predictable network service by implementing the following services:

- Supporting guaranteed bandwidth
- Improving loss characteristics
- Avoiding and managing network congestion
- Shaping network traffic
- Setting traffic priorities across the network



Americas Headquarters:
Cisco Systems, Inc., 170 West Tasman Drive, San Jose, CA 95134-1706 USA

© 2007–2009 Cisco Systems, Inc. All rights reserved.

About QoS Architecture

You configure QoS features throughout a network to provide for end-to-end QoS delivery. The following three components are necessary to deliver QoS across a heterogeneous network:

- QoS within a single network element, which includes queuing, scheduling, and traffic shaping features.
- QoS signaling techniques for coordinating QoS for end-to-end delivery between network elements.
- QoS policing and management functions to control and administer end-to-end traffic across a network.

Not all QoS techniques are appropriate for all network routers. Because edge routers and backbone routers in a network do not necessarily perform the same operations, the QoS tasks that they perform might differ as well. To configure an IP network for real-time voice traffic, for example, you would need to consider the functions of both edge and backbone routers in the network and then select the appropriate QoS feature or features.

In general, edge routers perform the following QoS functions:

- Packet classification and marking
- Admission control
- Congestion management

In general, backbone routers perform the following QoS functions:

- Congestion management
- Congestion avoidance

Who Could Benefit from Using QoS?

All networks can take advantage of aspects of QoS for optimum efficiency, whether the network is for a small corporation, an enterprise, or an Internet service provider (ISP). Different categories of networking users—such as major enterprises, network service providers, and small- and medium-sized businesses—have their own QoS requirements; in many areas, however, these requirements overlap. The Cisco IOS XE QoS features described in the [“WRED obviates this situation proactively by providing congestion avoidance. That is, instead of waiting for buffers to fill before dropping packets, the router monitors the buffer depth and performs early discards on selected packets sent over selected connections.” section on page 4](#) address these diverse and common needs.

Enterprise networks, for example, must provide end-to-end QoS solutions across the various platforms that comprise the network. Providing solutions for heterogeneous platforms often requires that you take a different QoS configuration approach for each technology. As enterprise networks carry more complex, mission-critical applications and experience increased traffic from web multimedia applications, QoS serves to prioritize this traffic to ensure that each application gets the service that it requires.

ISPs require assured scalability and performance. For example, ISPs that have long offered best-effort IP connectivity now also transfer voice, video, and other real-time critical application data. QoS answers the scalability and performance needs of these ISPs to distinguish different kinds of traffic, thereby enabling them to offer service differentiation to their customers.

In the small- and medium-sized business segment, managers are experiencing firsthand the rapid growth of business on the Internet. These business networks must also handle increasingly complex business applications. QoS lets the network handle the difficult task of utilizing an expensive WAN connection in the most efficient way for business applications.

Why Deploy QoS?

The Cisco IOS XE QoS features enable networks to control and predictably service a variety of networked applications and traffic types. Implementing Cisco IOS XE QoS in your network has the following advantages:

- Control over resources. You have control over which resources (bandwidth, equipment, wide-area facilities, and so on) are being used. For example, you can limit bandwidth consumed over a backbone link by FTP transfers or give priority to an important database access.
- Tailored services. If you are an ISP, the control and visibility provided by QoS enables you to offer carefully tailored grades of service differentiation to your customers.
- Coexistence of mission-critical applications. Cisco IOS XE QoS features ensure the following conditions:
 - That your WAN is used efficiently by mission-critical applications that are most important to your business.
 - That bandwidth and minimum delays required by time-sensitive multimedia and voice applications are available.
 - That other applications using the link get their fair service without interfering with mission-critical traffic.

Moreover, in implementing QoS features in your network, you put in place the foundation for a future fully integrated network.

End-to-End QoS Models

A service model, also called a level of service, describes a set of end-to-end QoS capabilities. End-to-end QoS is the ability of the network to deliver service required by specific network traffic from one end of the network to another. Cisco IOS XE Software supports three types of QoS service models: best effort, integrated, and differentiated services.

**Note**

QoS service models differ in how they enable applications to send data and in the ways in which the network attempts to deliver that data. For instance, one service model can be used for real-time applications, such as audio and video conferencing and IP telephony, while another service model can be used for file transfer and e-mail applications.

Consider the following factors when deciding which type of service to deploy in the network:

- The application or problem that you are trying to solve. Each of the three types of service—best effort, integrated, and differentiated—is appropriate for certain applications.
- The kind of capability that you want to allocate to your resources.
- Cost-benefit analysis. For example, the cost of implementing and deploying differentiated service is certain to be more expensive than the cost for a best-effort service.

The following sections describe the service models that are supported by features in Cisco IOS XE Software:

- [Best-Effort Service, page 4](#)
- [Integrated Service, page 4](#)
- [Differentiated Service, page 4](#)

Best-Effort Service

Best effort is a single service model in which an application sends data whenever it must, in any quantity, and without requesting permission or first informing the network. For best-effort service, the network delivers data if it can, without any assurance of reliability, delay bounds, or throughput.

Best-effort service is suitable for a wide range of networked applications such as general file transfers or e-mail.

Integrated Service

Integrated service is a multiple service model that can accommodate multiple QoS requirements. In this model the application requests a specific kind of service from the network before it sends data. The request is made by explicit signaling; the application informs the network of its traffic profile and requests a particular kind of service that can encompass its bandwidth and delay requirements. The application is expected to send data only after it gets a confirmation from the network. It is also expected to send data that lies within its described traffic profile.

The network performs admission control on the basis of information from the application and available network resources. It also commits to meeting the QoS requirements of the application as long as the traffic remains within the profile specifications. The network fulfills its commitment by maintaining per-flow state and then performing packet classification, policing, and intelligent queuing based on that state.

Differentiated Service

Differentiated service is a multiple service model that can satisfy differing QoS requirements. However, unlike in the integrated service model, an application using differentiated service does not explicitly signal the router before sending data.

For differentiated service, the network tries to deliver a particular kind of service based on the QoS specified by each packet. This specification can occur in different ways, for example, using the IP Precedence bit settings in IP packets or source and destination addresses. The network uses the QoS specification to classify, mark, shape, and police traffic and to perform intelligent queuing.

The differentiated service model is used for several mission-critical applications and for providing end-to-end QoS. Typically, this service model is appropriate for aggregate flows because it performs a relatively coarse level of traffic classification.

WRED obviates this situation proactively by providing congestion avoidance. That is, instead of waiting for buffers to fill before dropping packets, the router monitors the buffer depth and performs early discards on selected packets sent over selected connections.

WRED is the Cisco implementation of the RED class of congestion avoidance algorithms. When RED is used and the source detects the dropped packet, the source slows its transmission. RED is primarily designed to work with TCP in IP internetwork environments.

WRED can also be configured to use the DSCP value when it calculates the drop probability of a packet, enabling WRED to be compliant with the DiffServ standard being developed by the Internet Engineering Task Force (IETF).

Cisco and the Cisco Logo are trademarks of Cisco Systems, Inc. and/or its affiliates in the U.S. and other countries. A listing of Cisco's trademarks can be found at www.cisco.com/go/trademarks. Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1005R)

Any Internet Protocol (IP) addresses and phone numbers used in this document are not intended to be actual addresses and phone numbers. Any examples, command display output, network topology diagrams, and other figures included in the document are shown for illustrative purposes only. Any use of actual IP addresses or phone numbers in illustrative content is unintentional and coincidental.

© 2011 Cisco Systems, Inc. All rights reserved.

