



Signalling Overview

In the most general sense, QoS signalling is a form of network communication that allows an end station or network node to communicate with, or signal, its neighbors to request special handling of certain traffic. QoS signalling is useful for coordinating the traffic handling techniques provided by other QoS features. It plays a key role in configuring successful overall end-to-end QoS service across your network.

True end-to-end QoS requires that every element in the network path—switch, router, firewall, host, client, and so on—deliver its part of QoS, and that all of these entities be coordinated with QoS signalling.

Many viable QoS signalling solutions provide QoS at some places in the infrastructure; however, they often have limited scope across the network. To achieve end-to-end QoS, signalling must span the entire network.

Cisco IOS XE QoS software takes advantage of IP to meet the challenge of finding a robust QoS signalling solution that can operate over heterogeneous network infrastructures. It overlays Layer 2 technology-specific QoS signalling solutions with Layer 3 IP QoS signalling methods of the Resource Reservation Protocol (RSVP) and IP Precedence features.

An IP network can achieve end-to-end QoS, for example, by using part of the IP packet header to request special handling of priority or time-sensitive traffic. Given the ubiquity of IP, QoS signalling that takes advantage of IP provides powerful end-to-end signalling. Both RSVP and IP Precedence fit this category.

Either in-band (IP Precedence, 802.1p) or out-of-band (RSVP) signalling is used to indicate that a particular QoS is desired for a particular traffic classification. IP Precedence signals for differentiated QoS, and RSVP for guaranteed QoS.

IP Precedence

As shown in [Figure 1](#), the IP Precedence feature utilizes the three precedence bits in the type of service (ToS) field of the IP version 4 (IPv4) header to specify class of service for each packet. You can partition traffic in up to six classes of service using IP precedence. The queueing technologies throughout the network can then use this signal to provide the appropriate expedited handling.



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

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

Figure 1 *IP Precedence ToS Field*

IP precedence enables service classes to be established using existing network queueing mechanisms, such as weighted fair queueing (WFQ) and Weighted Random Early Detection (WRED), with no changes to existing applications and with no complicated network requirements.

Resource Reservation Protocol

RSVP is the first significant industry-standard protocol for dynamically setting up end-to-end QoS across a heterogeneous network. RSVP, which runs over IP, allows an application to dynamically reserve network bandwidth. Using RSVP, applications can request a certain level of QoS for a data flow across a network.

The Cisco IOS XE QoS implementation allows RSVP to be initiated within the network using configured proxy RSVP. Using this capability, you can take advantage of the benefits of RSVP in the network even for non-RSVP enabled applications and hosts. RSVP is the only standard signalling protocol designed to guarantee network bandwidth from end-to-end for IP networks.

RSVP does not perform its own routing; instead it uses underlying routing protocols to determine where it should carry reservation requests. As routing changes paths to adapt to topology changes, RSVP adapts its reservation to the new paths wherever reservations are in place. This modularity does not prevent RSVP from using other routing services. RSVP provides transparent operation through router nodes that do not support RSVP.

RSVP works in conjunction with, not in place of, current queueing mechanisms. RSVP requests the particular QoS, but it is up to the particular interface queueing mechanism, such as WFQ or WRED, to implement the reservation.

You can use RSVP to make two types of dynamic reservations: controlled load and guaranteed rate services, both of which are briefly described in the chapter “[Quality of Service Overview](#)” in this book.

A primary feature of RSVP is its scalability. RSVP scales well using the inherent scalability of multicast. RSVP scales to very large multicast groups because it uses receiver-oriented reservation requests that merge as they progress up the multicast tree. Although RSVP is designed specifically for multicast applications, it may also make unicast reservations. However, it does not scale as well with a large number of unicast reservations.

RSVP is an important QoS feature, but it does not solve all problems addressed by QoS, and it imposes a few hindrances, such as the time required to set up end-to-end reservation.

How It Works

Hosts and routers use RSVP to deliver QoS requests to the routers along the paths of the data stream and to maintain router and host state to provide the requested service, usually bandwidth and latency. RSVP uses a mean data rate—the largest amount of data the router will keep in the queue—and minimum QoS (that is, guarantee of the requested bandwidth specified when you made the reservation using RSVP) to determine bandwidth reservation.

A host uses RSVP to request a specific QoS service from the network on behalf of an application data stream. RSVP requests the particular QoS, but it is up to the interface queueing mechanism to implement the reservation. RSVP carries the request through the network, visiting each node the network uses to carry the stream. At each node, RSVP attempts to make a resource reservation for the stream using its own admission control module, exclusive to RSVP, which determines whether the node has sufficient available resources to supply the requested QoS.

**Note**

For RSVP, an application could send traffic at a rate higher than the requested QoS, but the application is guaranteed only the minimum requested rate. If bandwidth is available, traffic surpassing the requested rate will go through if sent; if bandwidth is not available, the exceeding traffic will be dropped.

If the required resources are available and the user is granted administrative access, the RSVP daemon sets arguments in the packet classifier and packet scheduler to obtain the desired QoS. The classifier determines the QoS class for each packet and the scheduler orders packet transmission to achieve the promised QoS for each stream. If either resource is unavailable or the user is denied administrative permission, the RSVP program returns an error notification to the application process that originated the request.

WFQ or WRED sets up the packet classification and the scheduling required for the reserved flows. Using WFQ, RSVP can deliver an integrated services Guaranteed Rate Service. Using WRED, it can deliver a Controlled Load Service.

For information on how to configure RSVP, see the chapter “[Configuring RSVP](#)” in this book.

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 used in this document are not intended to be actual addresses. Any examples, command display output, and figures included in the document are shown for illustrative purposes only. Any use of actual IP addresses in illustrative content is unintentional and coincidental.

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

