How Cisco IT Uses QoS for Critical Applications

QoS technology enables high-quality IP voice and video.

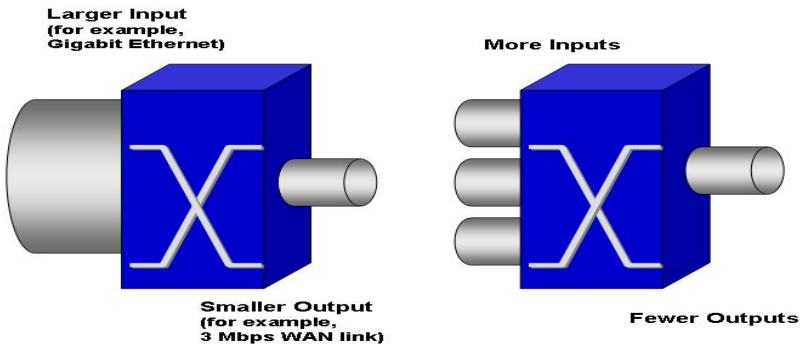
Cisco IT Case Study / Network Management / QoS for IP Voice and Video: This case study describes Cisco IT's internal deployment of quality of service within the Cisco global network, a leading-edge enterprise environment that is one of the largest and most complex in the world. Cisco customers can draw on Cisco IT's real-world experience in this area to help support similar enterprise needs.

BACKGROUND

Over the past decade, enterprises and institutions worldwide have experienced the network revolution. Using networks, these organizations can now accomplish tasks in seconds that required days, weeks, or even months just 15 or 20 years ago, and they know that networks have become the foundation for their success. Cisco Systems® has been an important contributor to this revolution, and the Cisco® IT team has learned about network behaviors and needs along with Cisco customers. We know that the Cisco enterprise network transports all types of applications and data, including such latency-sensitive data as high-quality video and real-time voice. Along with Cisco customers, we have discovered that the bandwidth-intensive applications that enhance and add value to Cisco business processes also strain our network capabilities and resources. Our goal is to ensure that the network provide secure, predictable, measurable, and—when needed—guaranteed services. Achieving the required quality of service (QoS) by managing delay, delay variation (jitter), bandwidth, and packet-loss parameters across the network has been the secret to our success in delivering a variety of end-to-end business solutions to our user community across a single converged IP network, the Cisco Enterprise Network.

Today, many individuals in enterprise IT organizations believe that networks built with high-capacity switches, multigigabit backplanes, and high-speed LAN and WAN links should never need QoS management. They believe that the more bandwidth available to an enterprise, the less QoS is needed. The Cisco WAN, MAN, and LANs—the Cisco Enterprise Network—is bandwidth-rich. Cisco IT built these networks with Cisco Catalyst® 6500 Series switches, Cisco 7600 Series routers, Gigabit Ethernet LAN links and WAN links sized up to OC-12 and STM-4. However, having a large amount of bandwidth in the network does make QoS unnecessary. All networks have congestion points where data packets can be dropped: WAN links where a larger trunk funnels data into a smaller trunk, or a place where several trunks funnel data into fewer trunks (Figure 1). QoS is not a substitute for bandwidth, nor does it create bandwidth. QoS lets network administrators control when and how data is dropped when congestion does occur. As such, QoS is an important tool that administrators use, along with adding bandwidth, as part of a coordinated capacity-planning process.

Figure 1. QoS Is Needed Wherever Congestion Occurs



Cisco IT's networks have some smaller links: traditionally the last-mile access links to remote WAN sites and, in remote parts of the world, where bandwidth is most expensive. With the increasing deployment of bandwidth-intensive uses like Cisco IP/TV® traffic and IP video traffic, these links become even more loaded and congestion increases. Even in Cisco IT's LANs, some links, no matter how well provisioned, are sometimes momentarily oversubscribed, and that can affect business. Whether it is a WAN access link, a WAN backbone link, or even a wiring closet switch with multiple lines that all share a single gigabit uplink—anywhere in a network where input is greater than output—congestion can occur and packets will be randomly delayed or lost. Sometimes the networks carry data that cannot tolerate delay. QoS is the key to managing congestion and packet delay and loss, ensuring voice and multimedia delivery all over Cisco.

CHALLENGE

The Cisco QoS story really started in 1995, when the Cisco IT team began implementing QoS policy link by link. Originally, administrators looked at traffic patterns case by case and applied QoS functions to individual routers. By 1999, the emergence of multimedia, and especially voice traffic, created a new urgency to control network traffic. The real-time nature of the traffic generated by these new, next-generation network applications made QoS even more critical. Voice applications are the most demanding because the packets have to get to their destinations on time. A late voice packet, unlike a late data packet, is worthless. Craig Huegen, Cisco IT chief network architect, explains what factors created IT's comprehensive approach to QoS: "From a business value perspective, the network needed to support the user demand for greater productivity. Videoconferencing and collaboration applications running across the LAN and WAN needed to respond to our users in real or near-real time. We had to deliver a higher level of service than we had been delivering, and we had to guarantee packet delivery. But voice was even tougher. We didn't want to lose voice packets as we started deploying IP telephony because losing part of a voice call is not acceptable."

SOLUTION

In the mid-1990s, the Cisco team had deployed basic priority queuing and Weighted Fair Queuing (WFQ), but as the network grew in size and complexity, mirroring the growth of the company, two things happened: Better QoS technologies were needed, and it became a great time to partner with the Cisco development organizations, who were developing ways to implement the best QoS capabilities across all Cisco products. Today, the Cisco IT team uses these basics, part of the Cisco IOS® Software feature set, that grew out of these collaborations:

- Classes of service and service markings
- Defining a "trusted edge"
- Class-Based Weighted Fair Queuing
- Low Latency Queuing

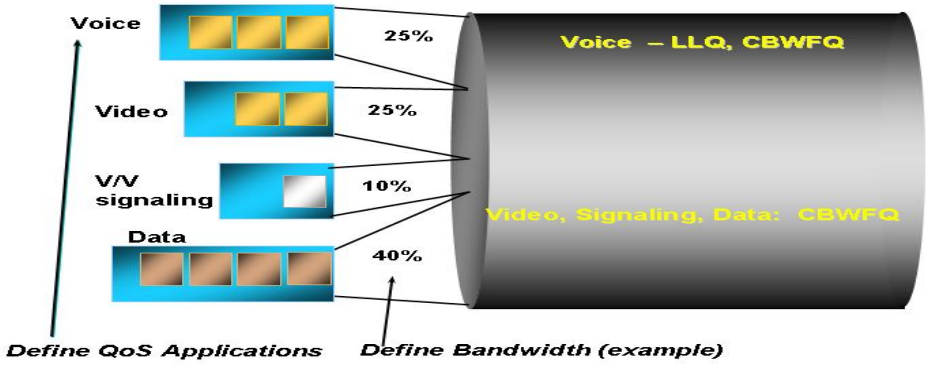
- Network-Based Application Recognition

Before applying QoS, a network administrator needs to set a policy dictating which applications need special treatment in the network. Voice traffic needs to be kept separate because it is especially sensitive to delay. Video traffic is also delay-sensitive and is often so bandwidth-intensive that care needs to be taken to make sure that it doesn't overwhelm low-bandwidth WAN links. After these applications are identified, traffic needs to be marked in a reliable way to make sure that traffic is given the correct classification and QoS treatment within the network.

After the network administrator determines QoS policies, he or she needs to define the "trusted edge"—the place where traffic is marked in a trustworthy way. It would be useless to take special care in transporting different classes of network traffic if traffic markings could be accidentally or maliciously changed. Cisco IT has determined that the trusted edge should be as close to the originating device (the IP phone, video camera, or data center application server) as possible. IT-managed routers and switches are also sources of trusted traffic. Traffic from other devices is marked as "default" traffic (class 0) by the first LAN edge switch. Lab traffic and Internet traffic are generally considered to be "untrusted," and care is taken before applying any special class of service (CoS) to these traffic types.

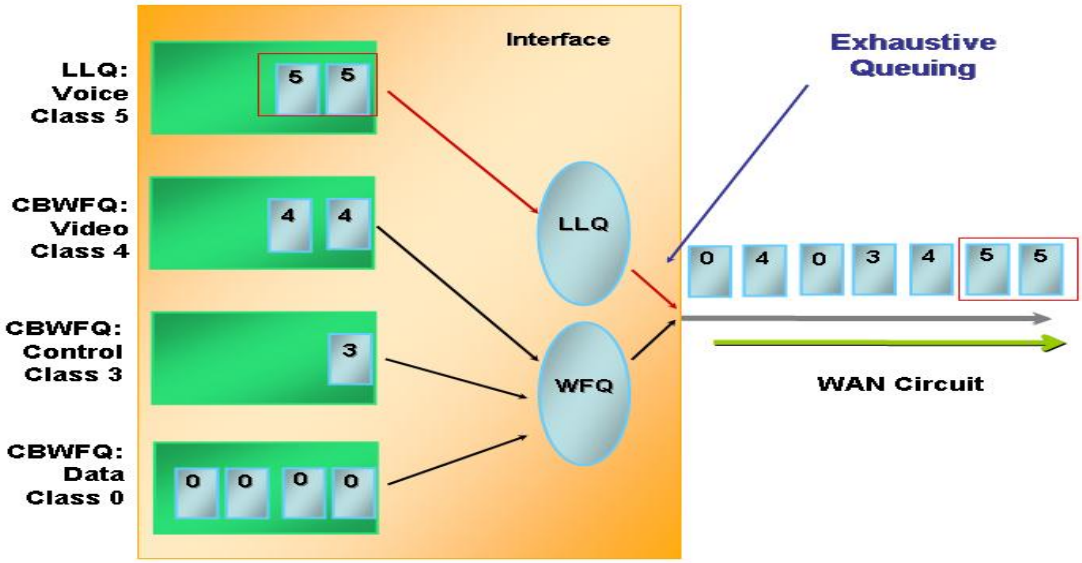
WFQ, an important Cisco queuing technique, is a flow-based queuing algorithm that fairly shares network bandwidth between high-bandwidth flows. Class-Based Weighted Fair Queuing (CBWFQ) extends the standard WFQ function to provide support for user-defined traffic classes (Figure 2). For CBWFQ, Cisco administrators define traffic classes based on match criteria, including protocols, access control lists (ACLs), and input interfaces, and then defines the minimum amount of bandwidth (often in terms of percentage of available bandwidth) that will be supported during congestion. When the network links are not congested, all traffic is allowed to flow freely. But when any traffic link is congested, the routers at that link will use CBWFQ classes to drop the least-valuable traffic first.

Figure 2. Allocating Bandwidth for Class-Based Weighted Fair Queuing



Low Latency Queuing (LLQ) brings strict priority queuing to CBWFQ (Figure 3). Even if voice packets are given priority during congestion, these latency-sensitive packets can still get trapped in the queue behind much larger data packets, and that would delay voice packet delivery. Using LLQ ensures that the smaller voice packets are sent to the front of the queue and not delayed by larger data packets.

Figure 3. Low Latency Queuing for Voice Traffic



Network-Based Application Recognition (NBAR) addresses another aspect of QoS within the Cisco network. The network has to provide Cisco enterprise applications with the appropriate services. It also has to make sure that noncritical applications don't hamper the performance of critical ones. NBAR is an intelligent classification engine in Cisco IOS Software that recognizes many kinds of applications, including Web-based, client-server, multimedia, and IP telephony applications. When a specific application is recognized using NBAR capabilities, the network can invoke the required services for that particular application. Sometimes, services are considered malicious, and traffic coming from them is unwelcome. Therefore, this traffic needs to be limited or dropped. NBAR helps limit or drop this traffic.

Because routers play a large role in the Cisco network's ability to deliver the end-to-end performance that each application needs, Cisco engineering created the Cisco Modular QoS Command-Line Interface (MQC). CBWFQ, LLQ, and NBAR are all part of MQC. MQC, a high-level language that represents a common configuration and command structure, lets Cisco administrators create networkwide policies for the treatment of different types of traffic based on application, protocol, user, or other criteria. MQC is part of the full QoS feature set that Cisco developed as part of Cisco IOS Software. Because it is a single framework for QoS that enables the network, MQC simplifies and lowers the cost of provisioning policies across the LAN and WAN and is consistent across all Cisco platforms, making QoS administration all over Cisco much easier.

Liem Nguyen, a Cisco IT network engineer, describes the advantages of MQC for Cisco network administrators: "When you have high-end platforms and low-end platforms like we do in the Cisco network, there are differences in how you implement QoS. In the high-end switches, there's a limited number of hardware-based queues for certain types of traffic, but on the low-end platforms, everything is done in software—they're very, very flexible. The MQC umbrella builds a baseline that connects all the platforms. That baseline is feature functionality in what we want to offer to users. The same commands are used for all products, but the implementation in the products is different, depending on the platform. With MQC, we can implement the same policy across network and all the platforms."

Cisco supports six classes of service. The network administration team has created a model that specifies all classes of service, assigning differentiated services based on application behaviors. Table 1 shows the classes and behavior assignments.

Table 1. Service Class and Packet Assignments

Service Class	Packet Assignments
Class 6	Network control traffic. This traffic is critical to maintaining the infrastructure. No user traffic is ever assigned to this class.
Class 5	Real-time voice. This class bears voice traffic.
Class 4	IT-sanctioned videoconferencing. This class bears any IT-sanctioned video traffic, and it carries only traffic from one videoconferencing node to another but at this time excludes the desktop. It is a real-time class, and it already claims large portions of bandwidth. Sometimes, it gets more bandwidth than things that are more critical.
Class 3	Voice and video and other signaling and control traffic. This class of traffic supports phone calls by sending signals to IP phones to ring. It has high priority against data applications because this signaling traffic will be retransmitted if lost, but it doesn't need the same expedited guarantees that voice traffic requires.
Class 2	High-priority data.
Class 1	Scavenger class. This is low-priority traffic, transporting high-volume bulk data with no interactivity demands. It uses remaining bandwidth after the other classes take what they need.
Class 0	Default traffic. All traffic enters the network as Class 0. In the interests of operational simplicity, this traffic as a matter of course is not remarked. Class 0 has higher priority than Class 1 traffic.

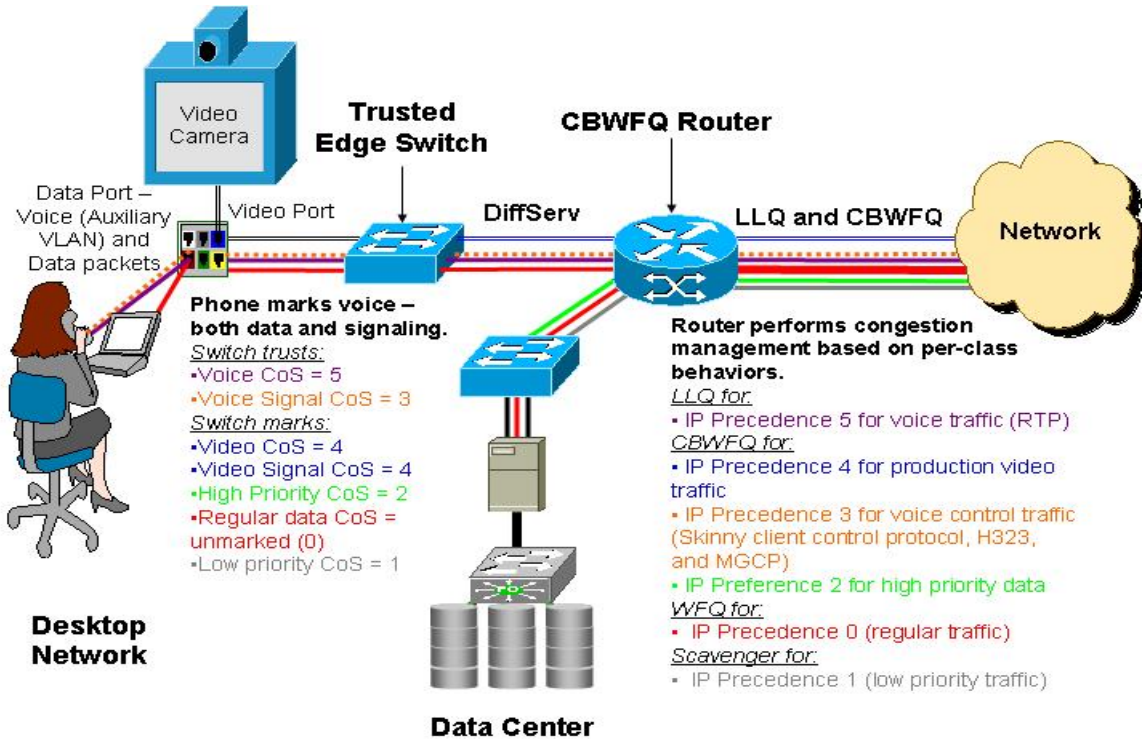
Huegen discusses the question of traffic class assignment: "Early on, we establish a trusted edge that we controlled at the edge of our network. It ensures that traffic is set to the right levels. We trust voice traffic from phones, but we don't trust packets from engineering labs." He laughs as he explains: "Here at Cisco, we've got 35,000 experts who know that we have QoS on the network. Sometimes they go in and reconfigure QoS assignments for business that isn't business-critical. We want traffic entering our network to be at the expected level for the application it will run. We put congestion management into the core of the network. Voice traffic goes into the carpool lane if there's congestion. The network doles out classes and numbers of packets, enforcing our QoS desires."

Nguyen describes the QoS process on the Cisco network: "We built an architecture which had the overall trust and classes of service. Then, we mapped applications to classes. Once that framework was in place, we designed it for our network, and then, we implemented it." Lab testing led the team to select a Cisco IOS Software release that supported class mapping with matching, using IP Precedence.

The architecture dictated how equipment across the network would be used to support QoS (Figure 4). It was important to configure edge switches, edge routers, and backbone routers differently because each had a different function:

- **Edge switches** mark all traffic that can't be marked by access control lists (ACLs). That included all voice traffic based on VLAN, all video traffic based on physical port, and all GeoTel and Cisco IP Contact Center (IPCC) traffic based on physical port. This LAN QoS requirement is also important for WAN QoS: without it, voice and video traffic would be marked as default data, defeating the purpose of WAN QoS.
- **Edge routers** trust all edge-switch marking, and perform all other packet classification based on ACLs (port numbers and endpoint addresses). Edge routers also perform Call Admission Control based on the amount of WAN bandwidth available to them.
- **Backbone routers** manage congestion and congestion avoidance functions.

Figure 4. QoS at the Switch and Router



Deployment and Implementation

The Cisco IT QoS deployment shares common themes with other comprehensive Cisco IT deployments. The Cisco IT group examines the problem and develops an architectural strategy and an architectural document first because it gives them a clearer picture of what is really needed. Getting there included lots of discussion and a prioritization of needs. When the strategy is agreed upon, it is easy to plan and deploy the technology solution. The QoS process at Cisco began by determining which applications needed special QoS treatment and then defining a common set of service classes that met application requirements. Cisco then established a trusted edge around the entire network, marking traffic to expected levels. For example, IP phone traffic was defined as traffic on particular voice-only VLANs and IP video traffic as traffic originating on specific switch ports dedicated to IP video cameras. Other applications were defined as traffic going to or from particular servers. Then the WAN and LAN links were classified and standard QoS rules were defined, for example, in smaller WAN access links where only one IP video connection would be allowed during periods of congestion.

To implement the architecture, making consistent QoS policies comprehensive across Cisco, the team created “cookbooks.” They published the cookbooks to the company’s traditional operational theaters and split up the implementation over three quarters. They deployed classification and marking rules at the trusted edge for the LAN, and then they deployed LLQ and CBWFQ rules at the WAN edge separately.

RESULTS

Because QoS is a pervasive feature of the Cisco network, it is quickly reflected in improved application performance. Explaining its results, Huegen says, “Once we put QoS into the devices, what we did discover were opinions. What we hear from our users is that they are hearing better voice quality. We don’t receive nearly as many voice or video complaints as we used to. Things just run better.” He adds that “QoS is also tied to bandwidth; it only kicks in when a circuit is experiencing congestion. So the places QoS makes the most impact are in areas of low-speed bandwidth or highest utilization. We try not to operate circuits with high utilization: QoS is an insurance policy against that.”

NEXT STEPS

Huegen says the QoS work is not finished. “It’s an ongoing effort because there will always be enhancement to our QoS policies.” Here are some examples of expected enhancements:

- **Lab traffic with QoS needs—**Cisco IT connects to hundreds of labs, which are often the source of uncontrolled traffic. Some of these lab servers can generate overwhelming quantities of traffic and can be the sources of viruses or worms. Some lab traffic will require special QoS handling, and these lab QoS requests must be handled case by case.
- **IP voice-over-Internet VPN—**Cisco IT is starting to run IP voice traffic over the Internet, using VPN connections, for employees working primarily from their home offices. These IP voice streams will require QoS handling.
- **QoS-over-Multiprotocol Label Switching (MPLS) VPN—**Cisco IT is implementing QoS across MPLS VPN service providers’ networks and requires these service providers to provide special handling for QoS traffic within their networks. However, each service provider has different QoS handling capabilities, different numbers of QoS service classifications, and different policies for handling and billing QoS. Cisco IT is still working on these issues.
- **Call Admission Control—**Cisco IT continues to work with gatekeeper technology in voice and video to prevent voice or video traffic to be oversubscribed; that is, to prevent too many IP voice or video packets from being delivered over a network link that cannot carry it all. Currently, most gatekeepers are unaware of network topologies and cannot guarantee bandwidth for their services.
- **Desktop trusted edge—**Cisco IT is updating its QoS strategy to bring the trusted edge to the desktop. Cisco frequently revises its strategy as new applications come on line. Bringing the trusted edge to the desktop will allow Cisco IT to trust specific, powerful desktop applications, including desktop videoconferencing.
- **Storage networking—**Cisco IT has begun to support the transport of storage area network, host-to-storage data between data centers on the same metropolitan-area network (MAN), at first over Fibre Channel over IP. This very-high-volume traffic will require special QoS handling and could also affect other traffic on the MAN.

CONTACTS

For more information about quality of service or other Cisco production deployments in the LAN or WAN, start a conversation with the Cisco IT experts in the Cisco@Work Campus, WAN, and Metro Forum at:

http://iforums.cisco.com/iforum/servlet/SCom?page=scom&folderID=caw&CommCmd=MB?cmd=display_messages&mode=new&location=.ee77358

For more information about quality of service as a technology, visit:

http://www.cisco.com/en/US/tech/tk543/tech_topology_and_network_serv_and_protocol_suite_home.html.

FOR MORE INFORMATION

To read the entire case study or for additional Cisco IT case studies on a variety of business solutions, visit Cisco on Cisco: Inside Cisco IT www.cisco.com/go/ciscoit

NOTE

This publication describes how Cisco has benefited from the deployment of its own products. Many factors may have contributed to the results and benefits described; Cisco does not guarantee comparable results elsewhere.

CISCO PROVIDES THIS PUBLICATION AS IS WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING THE IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Some jurisdictions do not allow disclaimer of express or implied warranties, therefore this disclaimer may not apply to you.

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 527-0883

Asia Pacific Headquarters
Cisco Systems, Inc.
168 Robinson Road
#28-01 Capital Tower
Singapore 068912
www.cisco.com
Tel: +65 6317 7777
Fax: +65 6317 7799

Europe Headquarters
Cisco Systems International BV
Haarlerbergpark
Haarlerbergweg 13-19
1101 CH Amsterdam
The Netherlands
www-europe.cisco.com
Tel: +31 0 800 020 0791
Fax: +31 0 20 357 1100

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**