QoS: What Is It?  Why Do We Need It?
Introduction

"My applications are sometimes slow."

"My video stream is jerky and the voice is out of sync."

"My voice calls sound bad at times."

Are these familiar complaints? Do your users face intermittent application problems that are hard to pin down?

This is an issue that many companies face at one time or another. It is also an issue that many companies are dealing with through the implementation of Quality of Service policies.

The Drivers Behind QoS

There are two primary drivers behind the need for QoS policies.

- Convergence of Voice and Data Networks
- Differentiation between Data Applications

The most common driver for QoS implementation is the convergence of voice and data onto an existing data network. The corporate PBX has reached end-of-life, and the decision has been made to replace it with a Voice over IP (VoIP) telephony service.

The mistake many companies make is to perform a very basic network analysis (if any at all) and add voice to the network without proper planning. Once the VoIP service is up and running, users call in to the helpdesk if service is not up to expected standards, and troubleshooting begins.

Without QoS, all packets on the network vie for the same pool of resources and when congestion occurs, any packet can be dropped without regard as to what the packet contains.

QoS may be implemented for differentiation of data applications only, even where there is no VoIP present. Some companies require allocation of bandwidth for critical applications.

If the company is a Service Provider, they will want to allocate bandwidth per customer and per customer class based on Service Level Agreements signed with that customer.

QoS policies can be implemented to provide resources based on a specific application.
What Is QoS?

QoS is the ability to treat packets differently as they transit a network device, based on the packet contents.

QoS configuration performs different tasks based on the direction of traffic flow and location of the device performing the QoS functionality.

At the access layer, where the ip packet first enters the network, QoS policies classify and mark each packet. This type of policy is applied in the inbound direction on an access layer interface.

Once a packet has been marked on the inbound path, the marking can be used on the outbound path to give each packet access to the appropriate amount of resources.

As the packet travels through the network, each device simply applies policies based on existing markings and does not need to do an in-depth analysis of the content of each packet.

Why Do We Need QoS?

Without QoS policies, each packet is given equal access to resources. If we cannot tell a voice packet from a data packet, we cannot give voice priority.

In order for a company to efficiently utilize its network resources, it must identify which network traffic is critical traffic and allocate appropriate resources to support those traffic streams. If voice is present in the network, it must get priority over all data streams; otherwise, the result could be intermittent voice quality complaints. Voice and video applications are delay- and jitter-sensitive. A good QoS policy will give the voice packet priority access to the interface queue.

For example, a voice packet and an FTP packet both arrive at an outbound router interface at the same time. Without QoS present, the voice packet may need to wait in a queue until the FTP packet has been processed out the interface. This may inject an unacceptable amount of delay into the voice path (depending on the interface speed).

With QoS configuration, the voice packet could be given priority over the FTP packet and be processed first.

If the interface is of a speed less than T1, the FTP packet may be fragmented to ensure that the voice packet does not have an excessive delay.
Another example will illustrate how QoS is used for application differentiation. In this example, a company relies on customer access to its web server so that customers can place orders from the on-line catalog. While on break, a staff member decides to watch a video trailer on an Internet movie web site, which requires significant bandwidth resources. Without QoS, a customer may experience delays in reaching the on-line catalog because of congestion caused by the video download. With QoS configuration, priority could be given to http packets that are sent to and from the corporate web server, while other web packets could be given lower priority. This would maintain a high level of resources for customer access while still providing some bandwidth to staff web traffic.

With the mix of traffic present in today’s networks, QoS ensures that the right applications get access to network resources first.

**Connecting To A Service Provider**

Special care must be taken when transiting corporate traffic through a service provider network. If the corporate traffic requires packet differentiation, then a Service Level Agreement (SLA) must be purchased from the service provider to assure access to proper resources across the provider network.

An SLA will usually define between three and six classes of data. As well, an SLA will define expected service levels for parameters such as packet loss, delay, and jitter per data class.

If the corporate network has differentiated its data streams into 6 classes, while the SLA only provides for 4 classes, the data must be re-classified at the customer edge prior to being sent into the provider network. Decisions will need to be made as to how to re-mark the packets based on a 4 class policy.

The re-classification should occur on the customer edge router. The packets will need to be re-marked according to the classes supported in the SLA. Once a packet is re-marked, then that packet will receive the appropriate treatment transiting the provider network.

**Service Provider SLA example**

When the traffic re-enters the customer network at the other end, it will need to be re-classified and re-marked back into the 6 data classes to normalize the markings for consistent internal QoS policy application.

**WAN Connection Considerations**

Additional mechanisms may be required on slower speed WAN links. A WAN link is considered “slower speed” if it has a speed lower than T1.

In this case, there are three mechanisms that may be required: header compression; data packet fragmentation; and interleaving.
Header Compression
Header compression will take a 40 byte RTP/UDP/IP header and compress it down to 2-4 bytes. In environments using the lower bandwidth G.729 codec, this could result in over 50% bandwidth savings.

Data Packet Fragmentation
Voice packets may need to wait for a data packet to clear the queue before being provided priority access to the interface. In order to minimize the wait time for the voice packet, the larger data packets may need to be fragmented into smaller pieces. The size of the fragment will depend on the speed of the interface. It is recommended that a data packet should take no more than 10ms to clear out of the queue, giving the voice packet access to the queue next.

Interleaving
Once the data packets are fragmented into smaller pieces, the scheduler must allow voice packets to interleave with the data fragments. This will ensure that the voice packet gets priority access to the queue.

The specific parameters for fragmentation and interleaving are configured differently for Frame Relay than for PPP links.

Summary
QoS can be used in converged networks to provide voice packets priority access to resources, or it can be used to differentiate data packets from different application stream and provide access to resources according to policy.

If voice is present in the network, it must be given priority over data packets to provide an acceptable level of voice service quality.
Learn More

Learn more about how you can improve productivity, enhance efficiency, and sharpen your competitive edge. Check out the following Global Knowledge courses:

QOS—Implementing Cisco Quality of Service
AQOS—Advanced Cisco Quality of Service
CVOICE/QOS Mini Camp

For more information or to register, visit www.globalknowledge.com or call 1-800-COURSES to speak with a sales representative. Our courses and enhanced, hands-on labs offer practical skills and tips that you can immediately put to use. Our expert instructors draw upon their experiences to help you understand key concepts and how to apply them to your specific work situation. Choose from our more than 700 courses, delivered through Classrooms, e-Learning, and On-site sessions, to meet your IT and management training needs.

About the Author

Berni Gardiner is a Certified Cisco Instructor and has taught for Global Knowledge since 1998. Berni’s 30+ years of technical expertise spans software development, network design and implementation and Voice over IP design and implementation. Over the past nine years, Berni has focused on converged network design, integrating voice technologies into data networks. Berni teaches a variety of Cisco courses including Implementing Cisco QoS, CVoice, Cisco IP Telephony I and II, and IP Telephony Express. Berni has worked with Global Knowledge and Cisco as the Subject Matter Expert (SME) developing material for multiple Cisco Instructor Led Training courses (ILT) and Global Knowledge Self-Paced eLearning courses (SPeL) in the voice arena. Her real-world experience includes working with local and national ISPs for network provisioning, as well as consulting on VoIP implementations across North America.