RESEARCH FEATURE

End-to-End Testing of IP QoS Mechanisms

Voice quality provides a valid metric for testing the relative effectiveness of quality-of-service mechanisms in preserving the end-to-end subjective quality of voice streams in the presence of multiservice traffic and IP network congestion.

Fayaz A. Shaikh
Novellus Systems

Stan McClellan
Compaq Telecommunications

Manpreet Singh
iBasis

Sannedhi K. Chakravarthy
University of Alabama at Birmingham

At this point in technology's evolution, the simplicity, elegance, extensibility, and broad compatibility of the Internet protocol suite has made it the automatic choice for most forms of communication. From Web browsing to enterprise computing, from electronic mail to next-generation wireless telephony, the common denominator is "everything over IP."

All is not necessarily well in IP land, however. Despite IP-based communications' huge potential, the mass deployment of converged networks carrying data, voice, and video has progressed neither as smoothly nor as quickly as projected. Performance optimization for these complex distributed systems is far from straightforward. From the digital signal processing techniques used in media compression to the management, integration, and policy enforcement technologies for the overall network, end-to-end network performance depends on many critical factors.

In particular, as the IP phenomenon has expanded, a new breed of traffic has emerged that challenges some fundamental assumptions about IP-design philosophy. These real-time multimedia streams not only demand higher bandwidth but also have peculiar timing requirements such as lower delay tolerance and higher end-to-end delivery guarantees than other data.

In contrast, IP's developers based its original design on best-effort packet forwarding, an approach that doesn't distinguish explicitly between the needs of particular streams. So far, the balanced overhead of computations involved in routing lookup and packet forwarding has achieved an acceptable level of performance for most network applications. Thus, the introduction of effective per-stream or per-packet considerations into this process must not incur a corresponding increase in computational overhead.

THE QOS SPECTRUM

The attempts at resolution of this apparent dichotomy consist of a collection of technologies and philosophies known as *quality of service*. In an IP network, QoS defines the ability to compensate for traffic characteristics without compromising average throughput. QoS thus lets network elements discriminate between particular traffic streams and then treat those streams in a particular manner, subject to broad constraints on forwarding performance.

QoS requires the cooperation of all logical layers in the IP network—from application to physical media—and of all network elements, from end to end. Clearly, optimizing QoS performance for all traffic types on an IP network presents a daunting challenge.

To partially address this challenge, several Internet Engineering Task Force groups have been working on standardized approaches for IP-based QoS technologies. The IETF's approaches fall into four categories:

0018-9162/02/$17.00 © 2002 IEEE

- prioritization using differentiated services,[1]
- reservation using integrated services,[2]
- label switching using multiprotocol label switching,[3] and
- bandwidth management using the subnet bandwidth manager.[4]

Differentiated services classifies per-hop behaviors on the basis of a Diffserv code point[5] attached to the type of service byte in each packet's IP header. This DSCP approach represents a form of soft QoS that rather coarsely classifies services through packet marking. Depending on the actual queuing and forwarding implementation, the expedited forwarding (EF) class, typically DSCP value 46, minimizes delay and jitter and provides the highest level of aggregate QoS. The transport effectiveness of EF-marked packets depends heavily on the per-hop implementation. The system administrator can use another differentiated services class, assured forwarding (AF), to assign a preset drop precedence to different traffic.

Integrated services emulates the resource allocation concept of circuit switching to apportion resources according to requests each host makes. The initiating host sends the upper-bound specification of bandwidth, delay, and jitter, and intervening nodes forward that request until it reaches the receiving host.

Integrated services define two kinds of reservation. A guaranteed reservation prescribes firm bounds on delay and jitter, determined from the network path, and allocates bandwidth as requested by the initiating host. Controlled-load reservations depend on the network's congestion state, thus they work only slightly better than best-effort service.

Label switching is a traffic engineering approach in which a router determines the next hop in a packet's path without looking at the packet header or referring to routing lookup tables. The IETF's multiprotocol label switching architecture assigns short, fixed-length labels to packets as they enter the network.[3] The network uses these labels to make forwarding decisions, usually without recourse to the original packet headers.

Bandwidth management provides the QoS standard for layer 2 technologies such as Ethernet and token rings. The subnet bandwidth manager signaling protocol allows communication and coordination between network nodes and switches in the SBM framework and enables mapping to higher-layer QoS protocols.

The "Quality-of-Service Strategies" sidebar lists a selection of IP-based QoS mechanisms that pertain primarily to the differentiated services and inte-

Quality-of-Service Strategies

The following IP-based QoS mechanisms pertain primarily to the differentiated services and integrated services categories. Under the proper circumstances, these mechanisms, which are widely available in conventional packet forwarding systems, can differentiate and appropriately handle isochronous traffic.

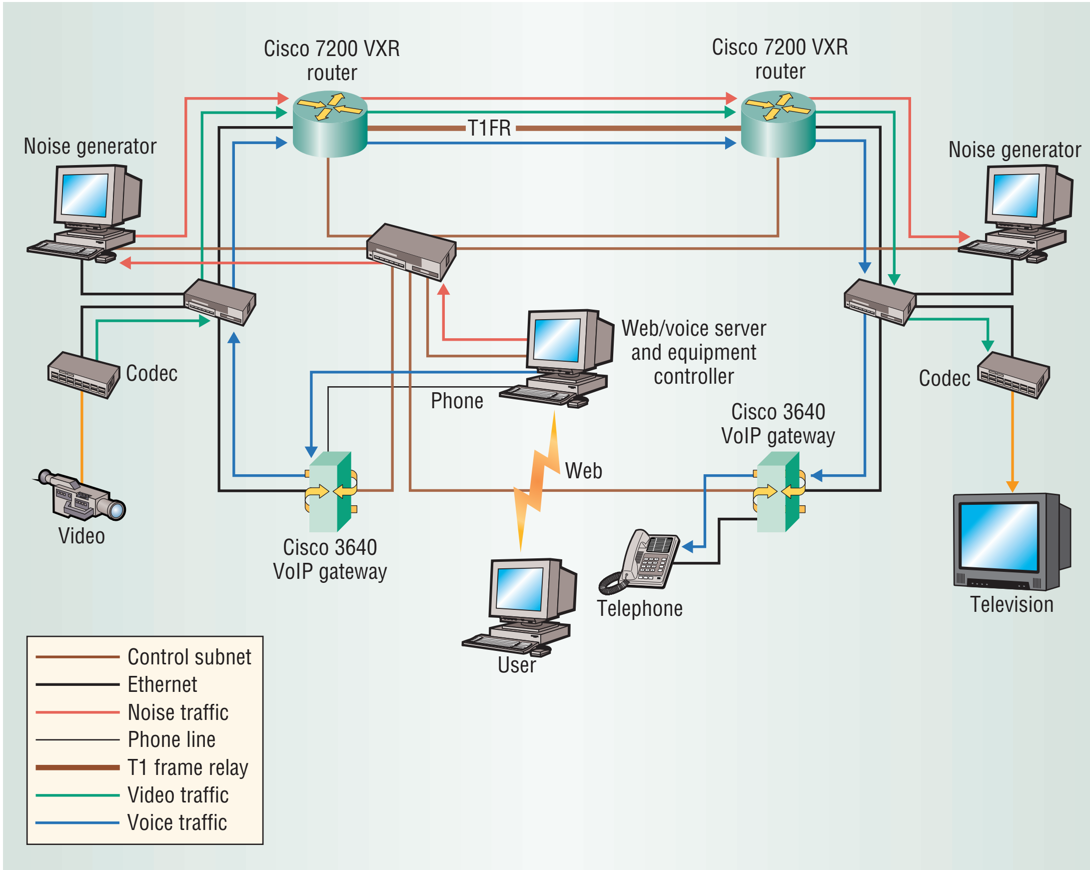
Queuing

- *First-In, First-Out.* FIFO, also known as the best-effort service class, simply forwards packets in the order of their arrival.
- *Priority Queuing.* PQ allows prioritization on some defined criteria, called policy. Four queues—high, medium, normal, and low—are filled with arriving packets according to the policies defined. DSCP packet marking can be used to prioritize such traffic.
- *Custom Queuing.* CQ allows allocating a specific amount of a queue to each class while leaving the rest of the queue to be filled in round-robin fashion. It essentially facilitates prioritizing multiple classes in queuing.
- *Weighted Fair Queuing.* WFQ schedules interactive traffic to the front of the queue to reduce response time, then fairly shares the remaining bandwidth among high-bandwidth flows.
- *Class-Based Weighted Fair Queuing.* CBWFQ combines custom queuing and weighted fair queuing. This strategy gives higher weight to higher-priority traffic, defined in classes using WFQ processing.
- *Low-Latency Queuing.* LLQ brings strict priority queuing to CBWFQ. It gives delay-sensitive data—such as voice—preferential treatment over other traffic. This mechanism forwards delay-sensitive packets ahead of packets in other queues.

Reservation, allocation, and policing

- *Resource Reservation Protocol.* RSVP, a signaling protocol, provides reservation setup and control to enable the resource reservation that integrated services prescribe. Hosts and routers use RSVP to deliver QoS requests to routers along data stream paths and to maintain router and host state to provide the requested service—usually bandwidth and latency.
- *Real-Time Protocol.* RTP offers another way to prioritize voice traffic. Voice packets usually rely on the user datagram protocol with RTP headers. RTP treats a range of UDP ports with strict priority.
- *Committed Access Rate.* CAR, a traffic-policing mechanism, allocates bandwidth commitments and limitations to traffic sources and destinations while specifying policies for handling traffic that exceeds the bandwidth allocation. Either the network's ingress or egress can use CAR policies, while access ports, IP addresses, or application flows can apply CAR thresholds.

Figure 1. Programmable network testbed. The state-of-the-art networking equipment was configured to collect voice quality measurements for IP-based QoS mechanisms.



grated services categories. The sidebar separates these QoS mechanisms into either queuing strategies or reservation, allocation, and policing techniques. Under the proper circumstances, these mechanisms, which are widely available in conventional packet-forwarding systems—IP routers—can differentiate and appropriately handle time-sensitive isochronous traffic.

We tested the relative effectiveness of these mechanisms in preserving the end-to-end subjective quality of voice streams in the presence of multiservice traffic and network congestion. We chose this metric for evaluating these QoS mechanisms because the telecommunications industry has generated a wealth of techniques for testing voice quality that have a long history of producing reliable results.

Our testing revealed that QoS mechanisms that approximate per-stream assurances through classification and careful forwarding techniques tend to perform better than mechanisms that simply reserve bandwidth, ensure low latency for forwarding without classification, or prioritize based on generic traffic classifications. Although this isn't a particularly surprising result, it is useful to compare the technologies and examine some of the ancillary factors that weighed heavily on our testing.

END-TO-END QOS EVALUATION

The four most important network parameters for the effective transport of multiservice traffic are

- bandwidth,
- delay,
- jitter, and
- packet loss.

Measuring and validating these parameters is difficult, particularly in the context of a highly subjective and variable phenomenon such as voice quality.

In many cases, the network designers who guide the technical implementations of voice-over-IP (VoIP) technology have little experience in subjective testing methodologies and the nuances of testing voice quality. Thus, these designers may optimize VoIP networks according to ineffective parameters or they may overengineer the implementations in an attempt to ensure sufficient resources for multiservice, real-time traffic.

Unfortunately, many of the popular remedies for packet-based network issues simply are not directly applicable to voice-grade transport. These issues may require a more abstract or indirect form of optimization because of the complex interrelationships between network configuration, packet-forwarding technologies, and effective end-to-end results.

One potential approach indirectly optimizes network configurations by testing the effectiveness of QoS technologies in transporting several kinds of real-time traffic. Using voice quality as a performance metric for network configurations provides appropriate mechanisms for obtaining and correlating both subjective and objective data.

The correlation between subjective results and objective measurements can yield significant insights into network performance. We can collect these subjective rankings and objective measurements, categorize them by the network configuration under which they were collected, sort them according to the user who produced the data, and store them for later use in comparing the relative overall performance of the various network parameters.

For this approach to be effective, the test administrator must be able to access, modify, and compare the effect of specific network parameters on end-to-end subjective voice quality for a large user population.

PROGRAMMABLE NETWORK TESTBED

To isolate specific variables, designers and administrators of subjectively oriented tests must be able to completely control the test environment. For VoIP testing, this level of control requires the ability to invoke particular network configurations across a wide variety of heterogeneous network elements.

All aspects of network configuration play essential roles in such an environment, including periodic variations in network parameters, call generation, and data collection, as well as a capacity for remotely administrating the test environment. A programmable network testbed can provide a platform for deploying various network scenarios and testing them under controlled conditions.

Network configuration control is particularly troublesome in the case of per-hop behaviors, which are critical for effective VoIP testing. Enabling per-hop behaviors often requires a series of low-level configurations on IP routers and other network elements.

In addition to being a fairly complex working environment, an IP network is extremely vulnerable to slight misconfigurations, which could cause serious outages or performance degradation. Unfortunately, in most cases, these configuration parameters are the source of trouble in the end-to-end QoS of converged networks. As a result, a programmable network testbed that automates all activities involved in network configuration is crucial for the effective collection of subjective data from users and objective data from measurement equipment.

Figure 1 shows the programmable network testbed we used to collect voice quality measurements for the IP-based QoS mechanisms described in the "Quality-of-Service Strategies" sidebar. The test network consists of several pieces of state-of-the-art networking equipment, including Cisco 7200VXR routers at the core and Cisco 3640 VoIP gateways at the edges. These gateways use the H.323 protocol for voice-call signaling. For encoding and decoding real-time video traffic, the testbed used VBrick MPEG-1 codecs.

Network configuration control is particularly troublesome in the case of per-hop behaviors, which are critical for effective VoIP testing.

This architecture separates the control subnet from other network connections, a configuration that lets the control traffic dynamically reconfigure the network to isolate it from the primary network connection that carries *bearer* traffic—voice, video, and IP signaling. For our study, we connected the core routers via a T1 trunk with frame relay encapsulation. At the edge of the network, each of the Cisco 3640 VoIP gateways has a port directly connected to a telephone handset or modem.

To evaluate congestion—a common and significant problem in IP networks—we used a configurable tool to generate noise that congests the T1 trunk. All voice, video, IP signaling, and noise traffic passes through the same T1 link. Using QoS mechanisms implemented in the core routers, the network prioritizes traffic or reserves resources even though the link may be completely congested. The system's software architecture consists of six main functional components. Each component acts as a proxy agent that uses a dynamic Web interface integrated into the system.[6]

RANKING QOS MECHANISMS

We collected two sets of results from our network testbed environment. The first set excludes video bearer traffic, limiting the traffic streams for this case to noise, voice, and signaling. The second set includes multiservice bearer traffic: noise, voice, signaling, and video. In both cases, we configured the system to collect objective measurements such as various types of packet loss as well as subjective measurements from test participants regarding transported voice quality.

To determine the effectiveness of various QoS techniques, we compared the tabulated subjective scores for the two sets of results. To achieve a relative ranking between specific network QoS mechanisms using subjective voice quality as a rough guide, we used a pseudomean opinion score (pMOS), which slightly redefines some components of the usual mean opinion score approach.

Table 1 sorts the relative subjective ranking of QoS mechanisms without video bearer traffic on

Table 1. Relative ranking of QoS mechanisms with no video traffic.

QoS mechanism	pMOS 95 percent confidence	Percentage of missed calls
Class-based weighted fair queuing with expedited forwarding	4.4 ± 0.22	0
Real-time protocol priority with weighted fair queuing	3.9 ± 0.27	0
Weighted fair queuing	3.6 ± 0.27	0
Class-based weighted fair queuing with low-latency queuing and expedited forwarding	2.3 ± 0.27	0
Resource reservation protocol with real-time protocol priority	1.5 ± 0.28	15.4
Resource reservation protocol only	1.4 ± 0.28	7.7
Real-time protocol priority	1.4 ± 0.28	19.2
Committed access rate with expedited forwarding	1.4 ± 0.31	0
Custom queuing with expedited forwarding	1.2 ± 0.19	7.7
Priority queuing with expedited forwarding	1.2 ± 0.19	19.2
Best effort	1.1 ± 0.14	26.9

two keys: the pMOS score and the percentage of missed calls for which signaling messages didn't arrive at the destination gateway for the corresponding network scenario. As a result, the destination handset didn't ring, and the call did not complete.

The results shown in Table 1 indicate that without explicit compensation for signaling packet transport, the particular QoS technology cannot ensure the reliable delivery of call-signaling messages, thus it performs poorly in the pMOS scores. This is a major difference between conventional telecommunications networks and IP networks.[7] Generally speaking, telecommunications networks physically segregate signaling traffic on carefully calibrated, highly redundant networks. In contrast, IP networks commingle bearer and signaling traffic on a single channel. Although it does not explicitly address QoS issues, the stream control transmission protocol[8] attempts to ensure reliable signaling transport in IP networks, and it may be effective in addressing some VoIP signaling transport issues.

Our data indicates that QoS mechanisms employing weighted fair queuing are preferable to other mechanisms, including reservation mechanisms, bandwidth-only methods, and generic port-based methods.

A comparison of the tabulated subjective pMOS scores for the results that include video and those that do not reveals the effectiveness of particular QoS techniques in a congested network. Because different techniques incorporate different policies for routing and packet forwarding, the end-to-end nature of the pMOS scores tends to magnify the factors that most effectively preserve high-quality voice transport. To validate our results, we calculated the 95 percent confidence intervals for pMOS scores over the entire test-taker population.

In addition to observing general subjective tendencies, we can divide the QoS technologies into two categories according to the success and failure of voice signaling.

Unsuccessful voice transport

Voice packet transmission failure occurred when the H.245 call-signaling control could not establish a logical channel for some QoS mechanisms.[7] Although we could not collect the packet loss data for voice traffic for this group, Table 2 shows the average video packet drop rate in the presence of voice and noise packets. The sample size for this data was smaller than for Table 1, which included a significantly larger number of users.

We didn't explicitly collect subjective data for video quality, which can be extremely difficult, but our personal opinions tend to align with the packet drop numbers. Thus, Table 2 sorts the QoS mechanisms in best-quality-video order. These QoS technologies exhibited voice call signaling failure in the presence of video.

Because of the voice signaling failure in these scenarios, the video packets were treated with either highest-priority or best-effort service. For example, using RSVP alone reserved significant bandwidth for voice, but it did not make a reservation for video packets. As a result, the video was treated as noise, resulting in a higher average packet drop rate. Assuming the availability of a reasonable estimate for the relative quantities of voice and video traffic,

Table 2. Objective performance of QoS technologies in the presence of multiservice bearer traffic.

QoS mechanism	Video packet drop rate (percent)
Priority queuing with expedited forwarding	0.36
Custom queuing with expedited forwarding	16.26
Resource reservation protocol only	56.15
Best effort only	72.95
Resource reservation protocol with priority	73.29

Table 3. Subjective and objective results for QoS technologies that exhibited voice call signaling success in the presence of video.

QoS mechanism	pMOS 95 percent confidence	Voice packet drop rate (percent)	Video packet drop rate (percent)
Class-based weighted fair queuing with expedited forwarding	4.0 ± 0.24	0.00	14.17
Weighted fair queuing	3.5 ± 0.30	0.00	27.79
Resource reservation protocol with real-time protocol priority	1.4 ± 0.43	13.66	50.49
Class-based weighted fair queuing with low-latency queuing and expedited forwarding	3.8 ± 0.22	0.00	23.62
Committed access rate with expedited forwarding	3.3 ± 0.28	1.20	28.65
Real-time protocol priority with weighted fair queuing	2.0 ± 0.24	56.37	28.58

network engineering could easily deal with this problem through a combination of call admission control and bandwidth allocation.

These configurations could also be made time-dependent, but any modifications must then acknowledge that the system is not forwarding voice-signaling packets properly, which obfuscates the situation somewhat. For example, the RTP priority mechanism treated voice packets as the only ones having forwarding priority, which meant that the network transported video packets with best-effort forwarding. Thus, the drop rate for these cases is similar—best effort and RTP priority both exhibit about a 73 percent packet drop rate.

For priority queuing, the network assigns video packets to the second-highest-priority queue and voice packets to the highest-priority queue. When call signaling failure occurs, no voice packets are present, effectively promoting the video packets. This prioritization causes a negligible average video packet drop rate, but it presents a misleading statistic for a properly configured network, which would handle voice signaling deterministically.

Custom queuing assigns the highest number of byte-counts to the voice queue, which means the network will dequeue the maximum number of voice packets before serving the next class of traffic—video. Because video packets received fewer byte counts than voice, the results show a slightly higher average video-packet drop rate than if there was no voice traffic. Even so, it's likely the drop rate is significantly lower than it would have been if voice signaling had been successful.

Successful voice transport

Table 3 shows the pMOS ratings and the average voice and video packet drop percentages for the successful transport of voice signaling packets using a slightly different set of QoS mechanisms. These QoS mechanisms produced subjective ratings consistent with those presented in Table 1 as well as subjective ratings that differ slightly from previous results.[6] In spite of some numerical differences between similar QoS mechanisms examined in the two sets of results, the relative ordering of objective and subjective results remained generally consistent throughout our testing.

With weighted fair queuing, which keeps packets containing interactive data at the front of the queue to be forwarded first, the network does not drop voice packets, demonstrating that dropped packets are not the only factor affecting perceived voice quality. The network drops video packets at a relatively higher rate than voice packets, but in our experience, getting the video right is noticeably less important, subjectively, than getting the voice right.

The situation is similar with variations of class-based weighted fair queuing, except that CBWFQ performs relatively better than weighted fair queuing alone for both voice and video. Because it is a class-based technique, CBWFQ uses weighted fair queuing mechanisms within each class, and a test administrator defines the weights and other specifications for the classes. Thus, CBWFQ provides a better approximation of explicit per-stream QoS guarantees.

A notable difference is that the results for RTP priority with WFQ are subjectively worse in Table 3 than in Table 1. The performance of the combined RTP priority and WFQ approach is subjectively worse in the presence of video because WFQ cannot use the differentiated services packet markings. As a result, it eventually mixes real-time voice and video traffic, dropping a large proportion of voice packets.

The performance distinction between QoS technologies is also interesting in regard to objective behavior in the presence of video traffic. The left side of Figure 2 uses the data from Table 2 to show the QoS mechanisms for which H.245 signaling call control could not establish a logical channel, resulting in voice call failure. Therefore, the subjective information for this group of QoS technologies could not be collected. In contrast, the right side of Figure 2 uses data from Table 3 to show QoS technologies for which the H.245 signaling call control established a channel that successfully transported and transmitted the voice signaling packets. In both cases, video packet transmission statistics were available.

Correlating these results with available subjective indicators provides additional insights into net-

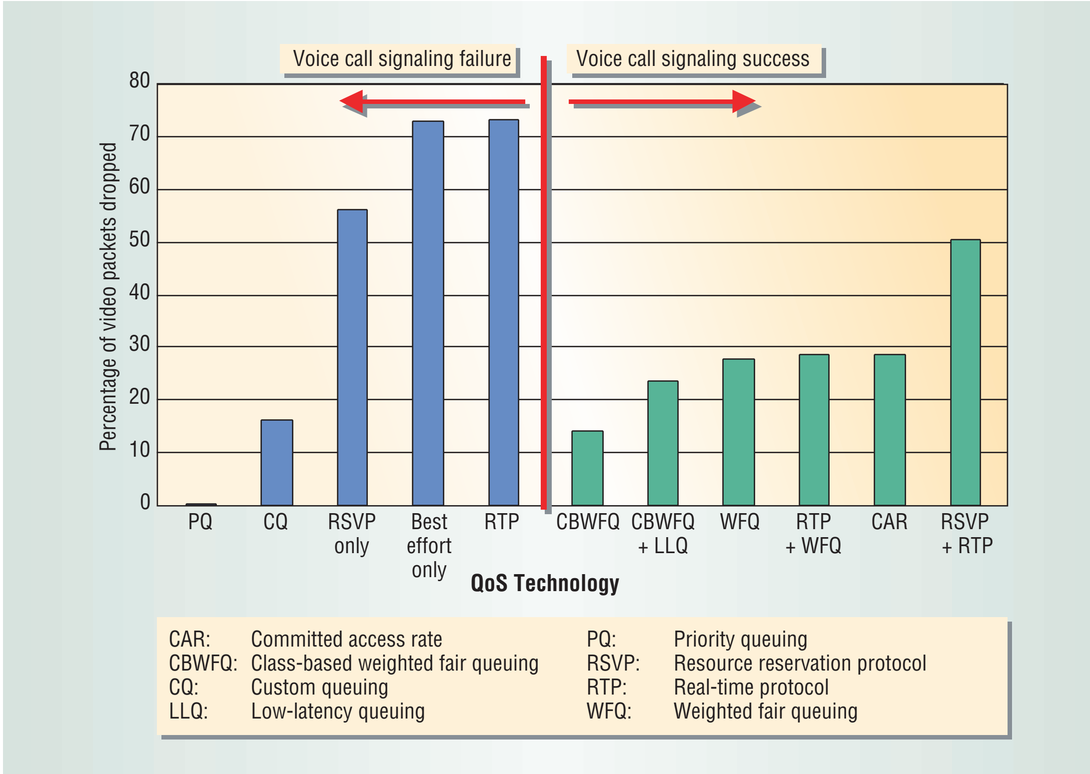


Figure 2. Performance of QoS mechanisms for transporting video packets. Voice packet transmission failure occurred when the signaling call control could not establish a logical channel.

work performance. For example, comparing the results of RSVP and RTP priority, it seems that a combination of RTP priority and RSVP improves the treatment of voice signaling packets without significantly improving the subjective rating of either method alone. This occurs regardless of the presence of video, although the combination of methods appears to reduce packet loss for the video stream. A limitation of the RSVP implementation resulted in higher than normal video packet loss in our study. On the other hand, the presence of video seems to have improved the classification capabilities for differentiated CAR and CBWFQ with LLQ. The objective results are consistent with the pMOS rankings of these QoS mechanisms.

Our studies demonstrate the relative performance of various QoS implementations under specific but broadly applicable network architectures and performance conditions. Our results reinforce the observation that joint optimization of network characteristics in the presence of general application-level traffic presents an extremely complex issue.

Adequate QoS requires optimizing queuing strategies, call-admission controls, congestion-avoidance mechanisms, and traffic-shaping and policing technologies. Given an IP network system's distributed nature, administrators often must perform these often highly interrelated optimizations simultaneously. Different QoS technologies—each addressing different aspects of end-to-end performance and each implemented in particular ways by equipment vendors—must be evaluated to achieve adequate QoS for VoIP deployments. ■

Acknowledgments

This work was supported in part by Cisco Systems, Compaq Telecom, and the UAB Center for Telecommunications Education and Research.

References

1. S. Blake et al., "An Architecture for Differentiated Services," IETF, RFC 2475, 1998, http://www.faqs.org/rfcs/rfc2475.html.
2. J. Wroclawski, "The Use of RSVP with IETF Integrated Services," IETF, RFC 2210, 1997, http://www.faqs.org/rfcs/rfc2210.html.
3. E. Rosen, A. Viswanathan, and R. Callon, "Multiprotocol Label Switching Architecture," IETF, RFC 3031, 2001, http://www.faqs.org/rfcs/rfc3031.html.
4. R. Yavatkar et al., "SBM (Subnet Bandwidth Manager): A Protocol for RSVP-Based Admission Control over IEEE 802-Style Networks," IETF RFC 2814, 2000, http://www.ietf.org/rfc/rfc2814.html.
5. K. Nichols et al., "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers," IETF RFC 2474, 1998, http://www.faqs.org/rfcs/rfc2474.html.
6. F.A. Shaikh and S. McClellan, *Evaluation of End-to-*

End QoS Mechanisms in IP Networks, Lecture Notes in Computer Science 2094, Springer-Verlag, Berlin, 2001, pp. 579-589.
7. D. Collins, *Carrier Grade Voice over IP*, 1st ed., McGraw-Hill, New York, 2001.
8. R. Stewart et al., "Stream Control Transmission Protocol," IETF, RFC 2960, 2000, http://www.ietf.org/rfc/rfc2960.html.

Fayaz A. Shaikh *is a software engineer at Novellus Systems. His research interests include software engineering, network systems designing and testing, and QoS in converged networks. Shaikh received an MS in electrical and computer engineering from the University of Alabama at Birmingham. Contact him at fayaz.shaikh@novellus.com.*

Stan McClellan *is currently a principal member of the technical staff in the Telecommunications Division of Compaq Computer Corporation. Previously, he was an associate professor in the UAB Department of Electrical Engineering and a senior research engineer in the UAB Center for Telecommunications Education and Research. His research interests include performance, network and platform architecture issues for "converged" wireless and wireline, and circuit-switched and packet-switched networks. He received a PhD in electrical engineering from Texas A&M University. Contact him at s.mcclellan@ieee.org.*

Manpreet Singh *is a SIP/network systems engineer at iBasis, Burlington, Mass. His research interests include VoIP, SIP, and other related protocols. Singh received an MS in electrical and computer engineering from the University of Alabama at Birmingham. Contact him at msingh@ibasis.net.*

Sannedhi K. Chakravarthy *is a master's degree candidate at the University of Alabama at Birmingham. His interests are VoIP quality issues and network security. Chakravarthy received a BE in electrical and electronics engineering from Andhra University, Visakhapatnam, India. Contact him at c.sannedhi@ieee.org.*