QoS in VoIP

Parijat Garg* **Rahul Singhai†**

Abstract

The Internet is under rapid growth and continuous evolution in order to accommodate an increasingly large number of applications with diverse service requirements. In particular, Internet telephony, or voice over IP is one of the most promising services currently being deployed. Besides the potentially significant cost reduction, Internet telephony can offer many new features and easier integration with widely adopted Web-based services.

Despite these advantages, there still exist a number of barriers to the widespread deployment of Internet telephony such as the lack of control architectures and associated protocols for managing calls, a security mechanism for user authentication, and proper charging schemes. The most prominent one, however, is how to ensure the Quality of Services (QoS) needed for voice conversation.

Keywords: Internet, VoIP, QoS.

1 Introduction

Todays Internet does far more then email and file transfers. Initially designed for non-real-time (NRT) data applications, the Internet has matured far beyond these tasks. In addition to web browsing, online imaging, and chat rooms, we now expect the Internet to deliver such real-time (RT) media as streaming music, video, and Internet phone calls directly to our homes and offices.

Presently, Voice-over-Internet Protocol (VoIP) is all the buzz, but it's more than just talk. VoIP technology has matured. Not only is it the latest hot new Internet application. Today, VoIP has emerged as a reliable technology that is commercially viable, competing (and winning) against traditional phone services in business and consumer-class markets.

As a real-time application, VoIPalso known as packet voice, packet telephony, or IP telephony-places increased demands on the evolving Internet. VoIP users expect the Internet to deliver toll-quality voice with the same clarity as the traditional Public Switched Telephone Network (PSTN). Even though VoIP applications do not require much bandwidths (voice calls over IP can be supported at as little as 8 Kbps), but they demand low delay and jitter for an aesthetically pleasing experience. In the current architecture of the Internet, no such guarantees can be provided to the application.

To meet VoIP users' expectations, the Internet connection must be more than merely reliable, it must be time-sensitive. Each and every voice packet must be delivered without significant delay and with consistent time intervals between packets. Quality of Service (QoS) technology is the key to achieving voice quality that measures up to todays high standards. The Internet Engineering Task Force (IETF), together with Internet backbone equipment providers, is addressing this with technologies like Resource Reservation Protocol (RSVP), which will let bandwidth be reserved.

The objective of this article is to review the recent developments and key enabling technologies in providing QoS supporting for voice communications in the next-generation Internet. The rest of the article is organized as follows. We first review the existing technologies in supporting VoIP networks, especially the basic mechanisms in the IETF Internet telephony architecture. We describe International Telecommunication Union Telecommunication Standardization Sector (ITU-T) H.323-related Recommendations for enabling multimedia communications in packet-based networks. We then discuss the IETF QoS framework, specifically the integrated services model (Intserv) and differentiated services (Diffserv) architecture. We present Cisco's solution, in offering IP telephony services as examples to illustrate how the real systems are implemented. We then conclude the article.

2 The Setting

As we have said before, VoIP is fast becoming the killer-application of the Internet. This is more so because VoIP calls cost much less than regular telephone calls. The biggest concern is that of providing quality of service over the Internet. At the highest level, the situation is as shown in Figure 1. In the quickly fading telephone networks, the guarantee of service was achieved by setting up a dedicated circuit for each call. However, we do not have such luxury in packet-switched networks.

The methods employed for providing QoS guarantees in packet-based networks are in some sense a sim-

*B.Tech. Student, Department of Computer Science and Engineering, Indian Institute of Technology Bombay, Powai, Mumbai-400076. email: parijat@iitb.ac.in

†M.Tech. Student, Kanwal Rekhi School of Information Technology, Indian Institute of Technology Bombay, Powai, Mumbai-400076. email: rahul.singhai@iitb.ac.in

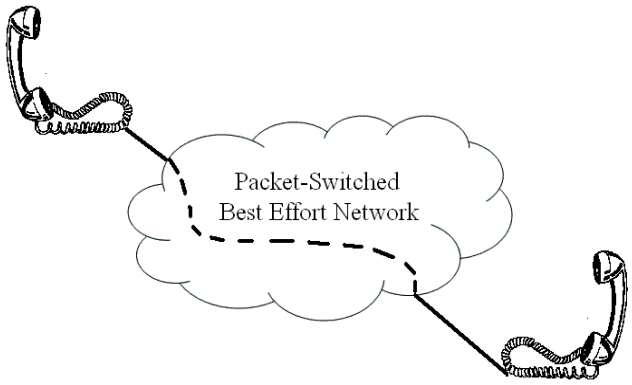

Figure 1: General VoIP Setting

ulation of dedicated circuits of the Plain Old Telephone System (POTS) on top of a packet-switched network. In the following sections, we describe how this is achieved.

3 Internet Telephony Standards

To support Internet telephony and other related applications, standards are being recommended and developed to insure interoperability. In particular, the ITU H.323 specification for Internet telephony is gaining widespread acceptance among software vendors. In addition, the IETF has also developed protocols such as Session Initiation Protocol (SIP) for multimedia session initiation, and Real Time Streaming Protocol (RTSP) for controlling multimedia servers on the Internet that can work together with H.323.

Interwoven with all of the above protocols is the Real-Time Transport Protocol (RTP). It is used by H.323 terminals as the transport protocol for multimedia; both SIP and RTSP were designed to control multimedia sessions delivered over RTP. Its main function is to carry real-time services, such as voice and video, over an IP network. It provides payload type identification so that the receiver can determine the media type contained in the packet. Sequence numbers and timestamps are also provided so that packets can be reordered, losses detected, and data played out at the right speeds. RTP was designed to easily be used in multicast conferences. To this end, it guarantees that each participant in a session has a unique identifier, providing applications a way to demultiplex packets from different users.

RTP also contains a control component, called the Real-Time Control Protocol (RTCP). It is multicast to the same multicast group as RTP, but on a different port number. Both data senders and receivers periodically multicast RTCP messages. RTCP packets provide many services. First, they are used to identify the users in a session. One RTCP packet type, the Source Descriptor (SDES), contains the name, e-mail address, telephone number, fax, and location of the participant. Another, the receiver report, contains reception quality reporting. This information can be used by senders to adapt their transmission rates or encodings dynamically during a session. It can also be used by network administrators to monitor network quality. It could potentially be used by receivers to decide which multicast groups to join in a layered multimedia session.

One of the key components supporting VoIP is a signaling protocol, which has to provide the following functions: user location, session establishment, session negotiation, call participant management, and feature invocation [SR99]. Within the IETF, two protocols are defined to implement these tasks: SIP [HSSR99] and Session Description Protocol (SDP) [HJ98].

SIP is used to initiate a session between users. It provides user location services, call establishment, call participant management, and limited feature invocation. SIP is a client-server protocol. This means that requests are generated by one entity (client), and sent to a receiving entity (the server), which process them. Since a call participant may either generate or receive requests, SIP-enabled end systems include both client and server. There are three types of servers. SIP requests can traverse many proxy servers, each of which receives a request and forwards to the next-hop server, which may be another proxy server or the final user agency server. A server may also act as a redirect server, informing the client of the next-hop server so that the client can contact it directly. SIP defines several methods, the client requests invoke method on severs. A client sets up a call by issuing an INVITE request. This request contains header fields used to convey call information. Following the header fields, there exists the body of the message that contains a description of the session to be established.

SDP is used to describe multimedia sessions for both telephony and distributed applications. The protocol includes several kinds of information, as follows. Media streams convey the type for each media stream. For each media stream, the destination address (unicast or multicast) is indicated by Address; Ports define the UDP port numbers for each sending or/and receiving stream. Payload type conveys the media formats that can be used during the session. For a broadcast- style session such as a television program, start and stop times convey the start, stop, and repeat times of the session, and Originator names the originator of the session and how that person can be contacted.

4 QoS Issues in the Internet

4.1 VOIP Qos Requirements

4.1.1 Latency

Callers usually notice roundtrip voice delays of 250ms or more. ITU-T G.114 recommends a maximum of a 150 ms one-way latency. Since this includes the entire voice path, part of which may be on the public Internet, your own network should have transit latencies of considerably less than 150 ms.

Most network SLAs specify maxium latency :

- Qwest SLA 50ms maximum latency
- Axiowave SLA 65ms maximum latency
- Verio SLA 55ms maximum latency
- Internap SLA 45ms maximum latency

The SLA numbers above are for backbone providers, the total latency for a VOIP call may also include additional latency in the VOIP provider's and the user's local ISP networks.

4.1.2 Jitter

Jitter can be measured in several ways. There are jitter measurement calculations defined in :

- IETF RFC 3550 RTP: A Transport Protocol for Real-Time Applications
- IETF RFC 3611 RTP Control Protocol Extended Reports (RTCP XR)

But, equipment and network vendors often don't detail exactly how they are calculating the values they report for measured jitter. Most VOIP endpoint devices (e.g. VOIP phones and ATAs) have jitter buffers to compensate for network jitter. Jitter buffers (used to compensate for varying delay) further add to the end-to-end delay, and are usually only effective on delay variations less than 100 ms. Jitter must therefore be minimized. Several network providers have specified maximum jitter in their SLAs.

- Qwest SLA 2ms maximum jitter
- Viterla SLA 1ms maximum jitter
- Axiowave SLA 0.5ms maximum jitter
- Verio SLA 0.5ms average, not to exceed 10ms maximum jitter more than 0.1% of time
- Internap SLA 0.5ms maximum jitter

The SLA numbers above are for backbone providers, the total jitter for a VOIP call may also include additional jitter in the VOIP provider's and the user's local ISP networks.

4.1.3 Packet Loss

VOIP is not tolerant of packet loss. Even 1% packet loss can "significantly degrade" a VOIP call using a G.711 codec and other more compressing codecs can tolerate even less packet loss. The default G.729 codec requires packet loss far less than 1 percent to avoid audible errors. Ideally, there should be no packet loss for VoIP. Most network SLAs specify maxium packet loss :

- Qwest SLA 0.5% maximum packet loss
- Axiowave SLA 0% maximum packet loss
- Verio SLA 0.1% maximum packet loss
- Internap SLA 0.3% maximum packet loss

The SLA numbers above are for backbone providers, the total packet loss for a VOIP call may also include additional packet loss in the VOIP provider's and the user's local ISP networks.

4.2 QoS Service Models

The existing Internet service (i.e., the best-effort service of IP) cannot satisfy the QoS requirements of emerging multimedia applications, primarily caused by the variable queuing delays and packet loss during network congestion. There has been a significant amount of work in the past decade to extend the Internet architecture and protocols to provide QoS support for multimedia applications. This has led to the development of a number of service models and mechanisms. In this section we discuss two key models: Intserv and Diffserv.

4.2.1 The Integrated Service Model

The Intserv model was proposed as an extension to support real-time applications. The key is to provide some control over the end-to-end packet delays in order to meet the real-time QoS. Specifically, the Intserv model proposes two service classes in addition to best-effort service. They are:

- Guaranteed service for applications requiring a fixed delay bound
- Controlled-load service for application requiring reliable and enhanced best-effort service

The fundamental assumption of the Intserv model is that resources (e.g., bandwidth and buffer) must be explicitly managed for each real-time application. This requires a router to reserve resources in order to provide specific QoS for packet streams, or flows, which in turn requires flow-specific state in the router. The challenge is to ensure that this new service model can work seamlessly with the existing best-effort service in one common IP infrastructure.

Intserv is implemented by four components: flow specification, the signaling protocol (e.g., RSVP), admission control routine, and packet classifier and scheduler. Applications requiring guaranteed or controlled-load service must set up path and reserve resources before transmitting their data. *Flowspec*, describing the source traffic characteristics, has to be provided to the network. Under the Intserv framework, two separate parts of the Flowspec are defined: one describes the flows traffic characteristics (the Tspec), and the other specifies the service requested from the network (the Rspec). *Admission control* routines determine whether a request for resources can be granted. If a new call is accepted without a particular limit, QoS for calls in progress may be degraded below an acceptable level, because total bandwidth required for the calls exceeds the network capacity. Therefore, call admission control is necessary to reject a new call when enough network spare capacity is not available. The call admission control mechanisms were established for the traditional telephone networks based on circuit switching technology as well as ATM networks. Traditionally, the Internet has provided the best effort services, and has not supported call admission control. However, admission control is necessary for guaranteeing QoS for real-time applications such as telephone service in the Internet. When a router receives a packet, the *packet classifier* will perform a classification and put the packet in the appropriate queue based on the classification result. The *packet scheduler* will then schedule the packet accordingly to meet its QoS requirement.

The problem with IntServ is that many states must be stored in each router. As a result, IntServ works on a small-scale, but as you scale up to a system the size of the Internet, it is difficult to keep track of all of the reservations. As a result, IntServ is not very popular.

4.2.2 The IETF Differentiated Services Framework

The Diffserv architecture as specified by IETF offers a framework within which service providers can offer each user a range of network services which are differentiated on the basis of performance [BBC+98]. The Diffserv architecture is based on a simple model where traffic entering a network is classified and possibly conditioned at the boundaries of the network, and assigned to different behavior aggregates (BAs), with each BA being identified by a single Diffserv code-point (DSCP). Users request a specific performance level on a packet-by-packet basis, by marking the Diffserv field of each packet with a specific value. This value specifies the per-hop behavior (PHB) to be allotted to the packet within the providers network. Within the core of the network, packets are forwarded according to the PHB associated with the DSCP.

Sophisticated classification, marking, policing, and shaping operations need only be implemented at network boundaries or hosts (Figure 1). Network resources are allocated to traffic streams by service provisioning policies which govern how traffic is marked and conditioned upon entry to a Diffserv-capable network, and how this traffic is forwarded within that network. A wide variety of services can be implemented on top of these building blocks.

A salient feature of the Diffserv framework is its scalability, which allows it to be deployed in very large networks. This scalability is achieved by forcing much complexity out of the core of the network into boundary devices which process smaller volumes of traffic and fewer flows, and by offering services for aggregated traffic rather than on a per-microflow basis. That is, complex traffic classification and conditioning functions are only implemented at network boundary nodes; inside the core network, PHBs are applied to aggregates of traffic which have been appropriately marked using the Diffserv field in the IPv4 or IPv6 headers. PHBs are defined to permit a reasonably granular means of allocating buffer and bandwidth resources at each node among competing traffic streams. Per-application flow or per-user forwarding state need not be maintained within the core of the network.

A Diffserv architecture can be specified by defining or implementing the following four components:

- The services provided to a traffic aggregate
- The traffic conditioning functions and PHBs used to realize the services
- The Diffserv field value (DSCP) used to mark packets to select a PHB
- The particular node mechanism to realize a PHB

Services A service defines some significant characteristics of packet transmission in one direction across a set of one or more paths within a network. There are two approaches to provide Diffserv:

- The first approach specifies the QoS in deterministically or statistically quantitative terms of throughput, delay, jitter, and/or loss. Such approach is called quantitative Diffserv.
- The second approach specifies the services in terms of some relative priority of access to network resources and is called priority-based Diffserv.

Conditioning Functions and PHB In order for a user to receive Diffserv from its Internet service provider (ISP), it must have a service-level agreement (SLA) with its ISP. A SLA basically specifies the service classes supported and the amount of traffic allowed in each class, respectively.

Users can mark Diffserv (DS) fields of individual packets to indicate the desired service at hosts or have

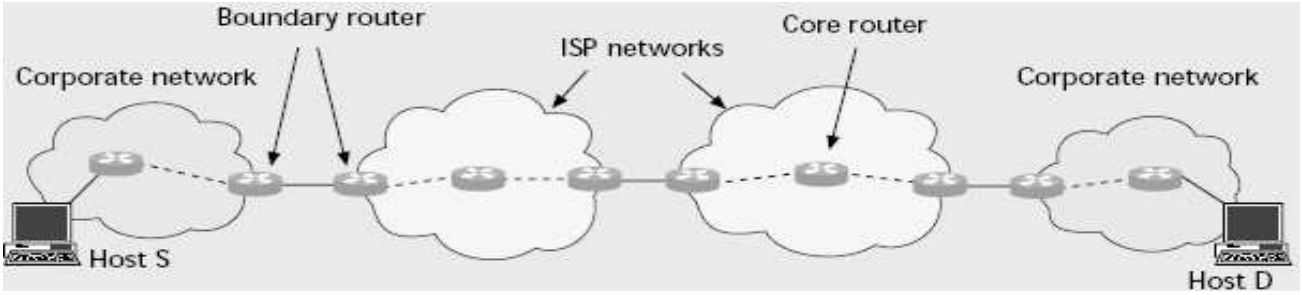

Figure 2: End-to-end transport from host S to host D under the Diffserv architecture.

them marked by the access or boundary router (Figure 2). At the ingress of the ISP networks, packets are classified, policed, and possibly shaped. The classification, policing, and shaping rules used at the ingress routers are derived from the SLAs. When a packet enters one domain from another, its DS field may be remarked, as determined by the SLA between the two domains. Such traffic control functions at hosts, or access or boundary routers are generically called traffic conditioning [BBC+98].

PHB refers to the externally observable forwarding behavior applied to a Diffserv behavior aggregate at a Diffserv-compliant node. PHBs are defined to permit a reasonably granular means of allocating buffer and bandwidth resources at each node among competing traffic streams.

DS Codepoint An IPv4 header contains a type of service (ToS) field, while an IPv6 header contains a traffic class byte. The IETF Differentiated Services Working Group has defined the layout of this byte (the DS field). By marking the DS field of packets differently and handling packets based on their DS fields, various Diffserv classes can be created. Six bits of the DS field are used as a codepoint (DSCP) to select the PHB a packet experiences at each node, while the other two are currently unused (CU).

A Node Mechanism for Achieving PHB PHBs are implemented in nodes by means of some buffer management and packet scheduling mechanisms. PHBs are defined in terms of behavior characteristics relevant to service provisioning policies, not in terms of particular implementation mechanisms. In general, a variety of implementation mechanisms may be suitable for implementing a particular PHB group.

5 Providing Quality of Service

There are two different dimensions along which quality of service management can be done in the VoIP setting. On the one hand, resource reservation techniques (RSVP), call admission control, etc. can be used. This is what we call the control plane. On the other hand, methods like loss recovery and error concealment can be usedin the data plane.

5.1 Techniques in Control Plane

Methods for providing QoS by working in the Control Plane usually involve procedures done before a call is setup. This could include route selection, resource reservation along the route and call admission control. Call admission control itself can be based on several metrics, including availability of bandwidth, availability of path capable of meeting required delay requirements, etc.

It is accepted that the performance of a VoIP application deteriorates very fast once the delay exceeds about 300 ms. A delay of greater than this duration essentially turns the channel into a half-duplex channel which is not acceptable for voice calls. Hence, any route selected for a call has to be able to provide a delay guarantee of less than 300 ms. Since delay is an additive metric, ordinary constraint shortest path first (CSPF) is not trivial to implement. Algorithms such as the Delay Scaling Algorithm (DSA) have been developed to efficiently find routes that can provide requisite QoS guarantees. MPLS can then be used to fix the path to be taken for a particular call.

Resource reservation is an important tool for providing QoS in general. Resource reservation allows us to set aside certain bandwidth for particular flows so that other misbehaving flows in the network cannot deter routers from providing the minimum required guarantees for higher priority flows such as VoIP flows. Specifically, RSVP can be used to setup Fixed Filter reservations for one-to-one voice calls.

Call admission control means that the control plane in the network takes runtime decisions about whether or not to allow a certain call to go through. This could be based on different metrics. For example, the service might first attempt to create a reservation for the call using RSVP. If the reservation fails, the service can reject the call. Call admission control can also be based on a preconfigured utilization threshold. Another criteria could be the availability of a route providing some minimum delay guarantee (say 50 ms) or minimum bandwidth (for a high voice quality call).

5.2 Techniques in Data Plane

Techniques used in the data plane are usually more directly affected by the nature of VoIP applications. The basic methods are naturally applicable. These include per-flow packet scheduling, class-based packet scheduling, etc. Per-flow packet scheduling does not scale well as the number of flows increase and hence class-based scheduling can be used. However, class-based scheduling decreases the control one has over exact delay guarantees for a particular flow and hence must be used with care.

Even though active queue management (say RED) is useful in a general QoS environment, it may not be applicable to the VoIP domain where the underlying protocol is more likely to be UDP than TCP. Random Early Detection (RED) is based on the assumption that the transport mechanism is TCP (or some such) which uses dropped packets as an indication of congestion and hence decrease their flow rates. However, RED might be useful to control TCP-based applications which can potentially crowd out VoIP packets. Also, by the very nature of voice, a certain rate of packet loss can be tolerated (as we will see ahead) and hence even if RED drops some VoIP packets, it will not affect performance very much.

As we said, VoIP applications are more sensitive to delay and jitter. Hence application level techniques like jitter buffers are used to absorb the jitter introduced due to the nature of the underlying network. This is a very standard technique to handle jitter even in non-interactive streaming video/audio applications.

5.2.1 Loss Recovery

To handle packet loss, two different techniques can be used. One is that of loss recovery. Another is error concealment. Due to the nature of voice applications, error concealment can be done fairly effectively and there is active research in developing sophisticated techniques. Loss recovery, using packet retransmission is usually not useful in the VoIP setting because it adds too much latency. However, another method, that of adding redundant information in the stream is employed and is fairly effective.

One technique used is that of transmitting two copies of the same voice stream. One of the stream is in higher quality encoding and the other is a lower bandwidth encoding. In case of loss of the high quality encoding, the lower quality packets are used instead. Thus, slight voice quality deterioration is introduced but the voice stream is not broken. Since human ears are much less sensitive to voice quality that to total absence of a packet, this is a useful method.

5.2.2 Error Concealment

Error concealment is a fairly intricate field in its own right. In this technique, the receiver of the voice stream does some local manipulations to conceal the loss of a voice packet. This could range from a very simple insertion of silence to a complicated reconstruction of the stream using the nature of the coding used for the voice stream and the previous and next packets.

In the encoding agnostic methods, the best technique used is that of replaying the last correctly received voice packet. Other techniques, inserting silence or noise result in unacceptable quality.

6 Voice Quality Monitoring

An element of voice quality is the ability to predict and monitor voice quality in an IP network. In circuit-switched networks, good voice quality, for those calls admitted into the network, was generally a given. However, IP networks introduce many new sources of distortion that can degrade voice quality. Before adding new services or network components to the existing network infrastructure, a service provider needs to determine the potential impact of the changes.

The end-to-end call quality consists of voice quality, call setup time, call blocking rate, call tear down time, and other call or service related defects. After a call is properly set up, voice quality is probably the most important characteristics for the entire call duration. The end-to-end voice quality must be maintained for the entire call duration.

Voice quality can be affected by various impairment factors such as codecs, delay, and packet loss. Such impairments are caused by the configuration of network equipment, network performance, and routing path of calls. Among them, the network performance must be monitored continuously due to its dynamic changes. A well-managed network is necessary to provide the desired level of VoIP service. If the voice quality were below the desired level, it would be necessary to perform root-cause analysis based on the measurements of network performance. Once the causes of the degraded voice quality are diagnosed, the problems must be fixed and it is important to ensure the solution really fixed the problems and did not cause any new problems.

6.1 Mean Opinion Scores

Voice quality is a subjective measure of how individual users perceive the speech quality and ease of conversing. The gold standard for measuring voice quality is specified by ITU Recommendation P.800 and is known as the Mean Opinion Score (MOS). The Mean Opinion Score (MOS) defines a method to derive a mean opinion score of voice quality after collecting scores between 1 (bad) and 5 (excellent) from human listeners. (see Figure 3) This is a form of subjective testing because human listeners are involved. In subjective testing, subjects (human listeners) are required to classify

the perceived quality into categories (excellent, good, fair, poor, bad). In each subjective experiment, the MOS scores may differ, even for the same condition, depending on the design of the experiment, the range of conditions included in the study, etc.

A rating of 4.0 or higher is often referred to as "toll quality" even though many Public Switch Telephone Network (PSTN) connections would be rated at about a 4.3. The measurements have to be done very carefully in a lab setting and require many subjects to be statistically valid. Thus this test may useful in rating specific pieces of equipment or a stable reference connection, but it is expensive, time-consuming and inappropriate for general network measurements.

6.2 Speech Quality Measures

There has been great interest in developing objective measures of voice quality that approximate the subjective human measures, and could be deployed in a network setting. In the mid 1990s, the ITU began to standardize objective speech quality measures designed to estimate subjective voice quality. A robust objective speech quality measure should correlate well with subjective speech quality. There are two types of objective speech quality measures: perceptual models and the E-Model.

The Perceptual Model : The Perceptual models [HKW04] estimates the voice quality by comparing the received speech signal to the sent speech signal in a psychoacoustic domain. The model focus on the effects of one-way speech distortion and they do not consider other impairments related to two-way interaction such as delay. The perceptual models are not scalable because they need to inject the speech samples at one end point and receive them at another end point in order to measure voice quality between two end points. If the voice quality becomes degraded, the perceptual models do not show the causes of degradations. These measures only get a snapshot of system performance by monitoring synthetic calls or average calls, not "real" calls. Additionally, by adding synthetic calls on the network, these measures can exacerbate conditions being tested by increasing load on the network. This tends to make the perceptual models more suitable for lab or prototype environments for capacity planning type activities.

The E-Model : The ITU has developed another class of objective measures, known as the E-Model and specified in Recommendation G.107 [G.100]. Thef E-model is a tool for predicting how an "average user" would rate the voice quality of a phone call with known characterizing transmission parameters. It estimates the user satisfaction of a narrowband, handset conversation, as perceived by the listener. The E-Model calculates the transmission rating factor R, using the network impairment factors, which were obtained after an extensive set of subjective experiments. Typical network impairment factors used in VoIP are codecs, delay, and packet loss. After computing the R-value based on the impairment factors, the R-value is converted into an MOS score. Since the E-Model is based on the measurements of impairments, it is appropriate for root-cause analysis in terms of impairment factors as well as network segments, and can be easily incorporated within the Network Management System. The E-Model is also scalable because it does not require the speech samples between many pairs of nodes to estimate the voice quality.

7 The Cisco Solution: Enterprise IP Telephony

The Cisco solution for IP telephony in enterprise networks [LHJC00] includes hardware, such as switches, routers, IP/PSTN gateways, desktop IP phones, and software, such as the call manager. An IP telephony system can be built by utilizing these products in the current IP infrastructure. Figure 5 illustrates a typical scenario of a Cisco IP telephony system.

In this IP telephony system, voice and data can be integrated in the wide area network (WAN) by permitting long distance calls to traverse the existing data infrastructure between remote locations. By using routers and gateways to connect the PBX, voice traffic can be carried over data IP networks. Call management software and IP telephones are deployed in the existing IP networks at each remote site. This will reduce the cost of WAN consolidation while at the same time eliminating the cost of installing a second network at each remote location. Using the analog access gateway (at the remote site), local calls can be enabled for remote users. Long distance calls can be routed over the WAN link and consolidated from the central site. With this approach, the transport for IP telephony becomes transparent to users, who will be unable to distinguish whether a call is placed over a packet network, a circuit-switched network, or a combination of both. The networks can support multiple classes of services (CoSs) and provide guaranteed QoS to real-time communications. QoS functions and mechanisms are distributed between cooperating edge/aggregation devices and core/backbone switches. Packet classification and user policies are applied at the edge of the network. Packet classification identifies and categorizes network traffic into multiple classes. The Cisco IP phone can set the IPv4 ToS at the ingress to the network.

The QoS guarantees are primarily provided by two mechanisms: the call manager equipped with a resource reservation protocol (e.g., RSVP) and a priority queue mechanism. The priority queue mechanism is maintained in the core routers, and is responsible for high-speed switching and transport as well as congestion avoidance. Congestion avoidance uses packet

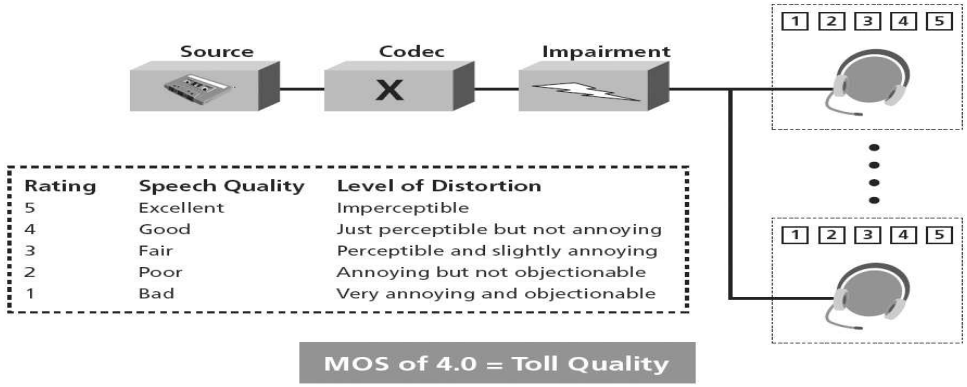

Figure 3: MOS Score Ratings.

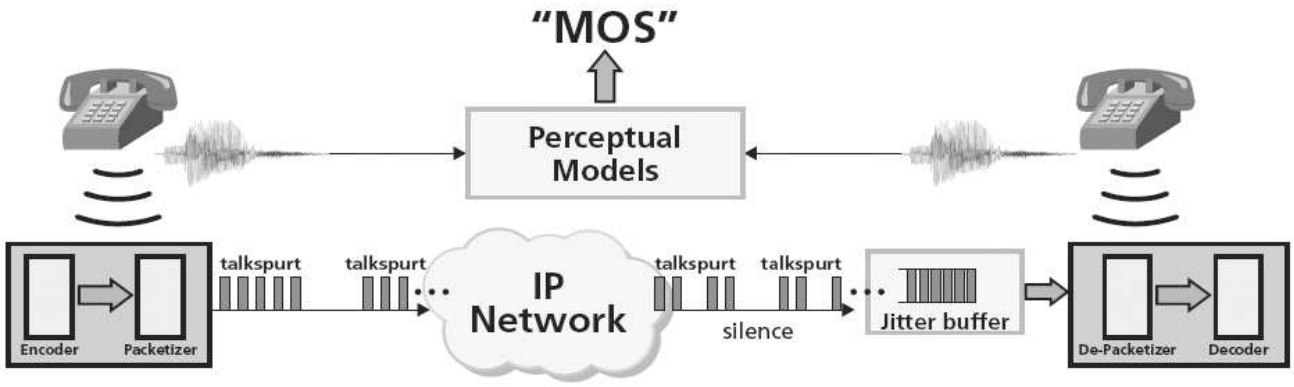

Figure 4: Perceptual Model Diagram.

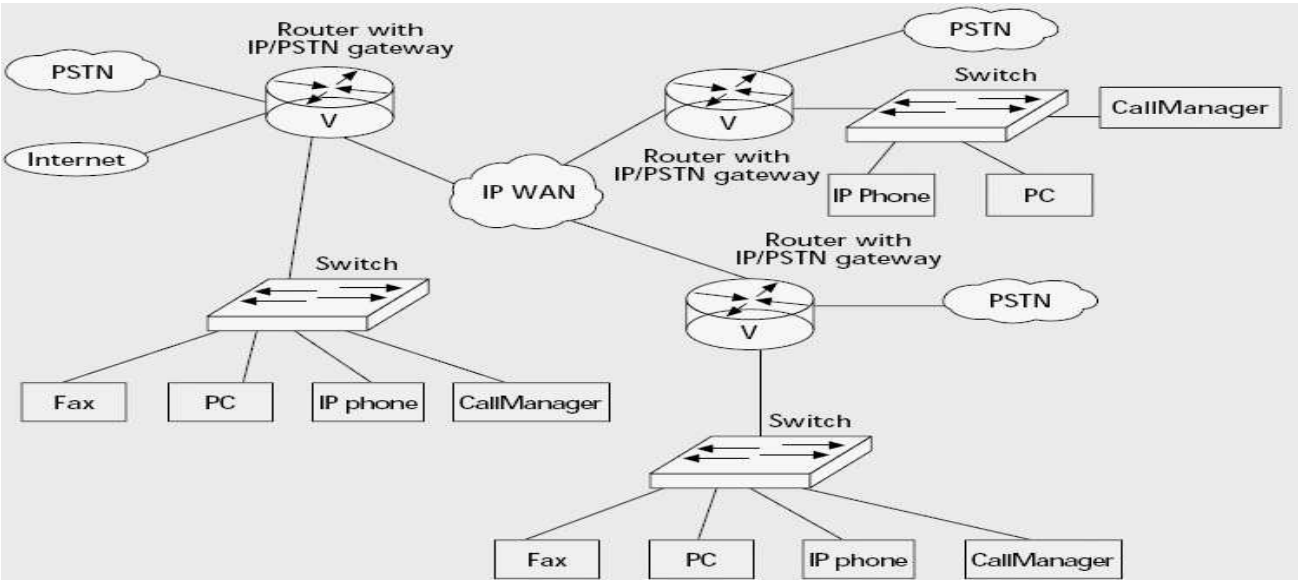

Figure 5: The Cisco data and IP telephony network configuration.

discard mechanisms such as weighted random early detection (WRED) to randomly drop packets on a congested link. WRED ensures that the voice packets will get higher-priority services while no one user monopolizes network resources.

8 Conclusion

The introduction of QoS to IP networks does have effect on all four performance measures (delay, jitter, frame loss and the out-of-order packets). It is therefor understandable that network equipment manufactur-

ers are putting high bets and hopes for the introduction of QoS mechanisms into IP networks. With the introduction of service differentiation, different types of traffic can experience different network conditions and therefore can have different average values for the delay, jitter, packet loss and the number of out-of-order packets. If, for instance, real-time traffic such as voice gets high priority over data, then its performance rises to the levels at which its real-time transmission over IP networks is no longer a question.

With the use of QoS techniques many new services will be possible, among them will certainly be a high-quality IP telephony. Equipment manufacturers see the benefit mainly in converging all the different devices into one that is connected to an IP network and offers the functionality of several separate devices that are a part of several different transmission networks. QoS today comes in many different flavours. It can be offered in two basic ways. Absolute QoS levels (absolute values of bandwidth, delay and other parameters are agreed) that are offered by technologies such as ATM and RSVP and relative QoS levels (performance relative to priority class) that are offered by technologies such as TOS in IP networks or Precedence in Frame Relay networks.

Additional to the introduction of QoS to the network there are other means to improve the performance of applications that require real-time transmission. However they only have limited effect. These techniques are: use of advanced jitter buffers that can adapt its length to the changing network conditions, use of FEC and loss concealment, use of long packed fragmentation. In private IP networks (Intranets) it is fairly easy to prioritise real-time traffic in all network nodes, but the traffic patterns must be known, specifically the share of real-time traffic. We can divide techniques for improving the transmission of VoIP traffic into two groups. In the first group we find mechanisms for QoS and in the second all other techniques. The introduction of QoS to the Internet is a very complex task, specifically from the viewpoint of defining new services and their support on the entire transmission path that traverses many ISPs (Internet Service Providers). In local area networks and Intranet environments short term prospects are a bit better. There is a lot of network equipment that can be configured with QoS and can offer it to their users. There has been significant work done to establish the foundation to support VoIP. However, much remains to be done in order to ensure the QoS for VoIP and for multimedia traffic in general in the next-generation Internet.

References

[BBC+98] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. An Architecture for Differentiated Service. RFC 2475 (Informational), December 1998. Updated by RFC 3260.

[G.100] ITU-T Recommendation G.107. The E-model, a computational model for use in transmission planning, March 2000.

[HJ98] M. Handley and V. Jacobson. SDP: Session Description Protocol. RFC 2327 (Proposed Standard), April 1998. Updated by RFC 3266.

[HKW04] Christian Hoene, Holger Karl, and Adam Wolisz. A perceptual quality model for adaptive voip applications. *International Symposium on Performance Evaluation of Computer and Telecommunication Systems*, July 2004.

[HSSR99] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg. SIP: Session Initiation Protocol. RFC 2543 (Proposed Standard), March 1999. Obsoleted by RFCs 3261, 3262, 3263, 3264, 3265.

[LHJC00] Bo Li, Mounir Hamdi, Dongyi Jiang, and Xi-Ren Cao. Qos enabled voice support in the next generation internet: issues, existing approaches and challenges. *IEEE Communications Magazine*, 38(4):54–61, April 2000. ISSN:0163-6804.

[SR99] Henning Schulzrinne and Jonathan Rosenberg. IETF Internet telephony architecture and protocols, March 1999.