QoS in VoIP [*]

Sarat Chandra
04329009

November 28, 2004

Abstract

During the recent Internet stock bubble, articles in the trade press frequently said that, in the near future, telephone traffic would be just another application running over the Internet. Such statements gloss over many engineering details that preclude voice from being just another Internet application. IP networks are designed to support non-real time applications, such as file transfer or e-mail. These applications are characterized by their bursty traffic and high bandwidth demands at burst times, but they are not very sensitive to delay or delay variation. On the other hand VoIP requires timely packet delivery with low delay, jitter and packet loss values. We need some sort of QoS mechanisms included into IP networks to bring Voice into it. This paper deals with the technical aspects of implementing voice over Internet protocol (VoIP), and various performance improvement issues for it. It also provides a mechanism for Admission Control into Cisco's VoIP system which enables a certain sort of QoS into the current system.

1 Introduction

VoIP (Voice-over-IP) refers to the transmission of voice using IP technologies over packet switched networks. It consists of a set of facilities and protocols for managing the transmission of voice packets using IP. Internet Telephony is one of the typical applications of VoIP. Compared to traditional resource-dedicated PSTN, IP network is resource-shared. Therefore, IP-based VoIP applications are cost-efficient.Service provided by traditional network varies under Congestion or during Convergence, in terms of variable delay, jitter, available bandwidth. Traffic has diverse service needs [5]. For example Voice requires Low Delay, Low jitter, Low loss where as Ordinary Data transfer can be delayed without affecting much to the client's requirement. To withstand to such needs, Certain level of QoS mechanism is needed.However, current IP networks are best-effort services. They lack stringent QoS control. Congestion is inevitable in IP networks and may result in packet loss, delay, and delay jitter, which directly impact the quality of VoIP applications. So, the current IP network architecture must be enhanced by some guaranteeing mechanisms in order to ensure QoS VoIP applications. This paper focusses on management mechanisms for VoIP applications and also includes some performance issues. It also includes a new call admission control mechanism for Cisco VoIP system.

2 Details about Signalling protocol(H.323)

The ITU-T Recommendation H.323 protocol suite has evolved out of a video telephony standard. The Voice-over-IP Activity Group of the International Multimedia Telecommunications Consortium (IMTC) recommended H.323, which had been developed for multimedia communications over

---

[*] This is a Seminar Report done under Prof. Anirudha Sahoo.

packet data networks. These packet networks might include LANs or WANs. VoIP is viewed as a special case of IP Video Telephony.The H.323 protocol suite became the early leading standard for VoIP implementations. H.323 entities may be integrated into personal computers or routers or implemented in stand-alone devices. For VoIP, the important H.323 entities are terminals, gateways, and gatekeepers.[5]

An H.323 gateway provides protocol translation and media transcoding between an H.323 endpoint and a non-H.323 endpoint. For example, a VoIP gateway provides translation of transmission formats and signaling procedures between a telephone switched circuit network (SCN) and a packet network. In addition, the VoIP gateway may perform speech transcoding and compression, and it is usually capable of generating and detecting DTMF signals.

H.323 Gatekeeper Characteristics: H.323 gatekeepers perform admission control and address translation functions. Several gatekeepers may communicate with each other to coordinate their control services. The gatekeeper is logically separate from the other H.323 entities, but physically it may coexist with a terminal, gateway, or an H.323 proxy. When present in a VoIP network, the gatekeeper provides the following functions: Address translation:: the gatekeeper translates alias addresses (e.g., E.164 telephone numbers) to Transport Addresses, using a translation table that is updated using Registration messages and other means. Admission control:: The gatekeeper authorizes network access using H.225 messages. Admissions criteria may include call authorization, bandwidth, other policies. Bandwidth control:: The gatekeeper controls how much bandwidth a terminal may use. The gatekeeper provides the above functions for terminals and gateways that have registered with it.

3 VoIP Architecture and QoS management

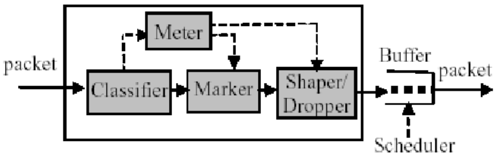
Current implementation of VoIP has two types of architectures, which are based on H.323 and SIP frameworks, respectively. SIP, which is defined in RFC2543 of the MMUSIC working group of IETF, is an application-layer control signaling protocol for creating, modifying, and terminating sessions with one or more participants. Regardless of their differences, the fundamental architectures of these two implementations are the same. They consist of three main logical components: terminal, signaling server and gateway[4]. They differ in specific definitions of voice coding, transport protocols, control signaling, gateway control, and call management. QoS requirements of VoIP include packet loss, delay, and delay jitter. The current H.323 and SIP frameworks support some kind of interfaces to QoS management (e.g., the one between H.323 and RSVP), they do not provide functional QoS management mechanisms. Consequently, products in the market now (e.g., Cisco s and Alcatel s VoIP systems) cannot provide QoS guarantees to VoIP applications.

QoS management architecture of VoIP can be partitioned into two planes: data plane and control plane. Mechanisms in data plane include packet classification, shaping, policing, buffer management, scheduling, loss recovery, and error concealment. They implement the actions the network needs to take on user packets, in order to enforce different class services. Mechanisms in control plane consist of resource provisioning, traffic engineering, admission control, resource reservation and connection management etc.[4]

3.1 DATA PLANE

3.1.1 packet forwarding

It consists of Classifer, Meter , Marker , Shaper/Dropper..
When a packet is received, a packet classifier determines which flow or class it belongs to. All



packets belonging to the same flow/class obey a predefined rule and are processed in a similar manner. For VoIP applications, the basic criteria of classification could be IP address, TCP/UDP port, protocol, input port, IP precedence, DiffServ code points (DSCP), or Ethernet 802.1p class of service (CoS). Cisco supports several additional criteria such as access list and traffic profile. The *meter* is to decide whether the packet is in traffic profile or not. The Shaper/Dropper delays or drops the packets which crossed the limits of traffic profile to bring in compliance to current network load. A marker marks the certain field in the packet, such as DS field, to label the packet type for differential treatment later. After the traffic conditioner, a buffer is used to store packets that wait for transmission.

3.1.2 Buffer management and Scheduling::

Active queue management (RED) which drops packets before a queue becomes full can avoid the problem of unfair resource usage. Predictable queuing delay and bandwidth sharing can be achieved by putting the flows into different queues and treating individually. Schedulers of this type are not scalable as overhead increases as the number of on-going traffic increases. Solution is class-based schedulers such as Static Priority amd Constraint Based WFQ which schedule traffic in a class-basis fashion. But, it is difficult for the individual flow to get the predictable delay and bandwidth sharing. So care should be taken to apply this to voice application which have strict delay requirements.

3.1.3 Loss Recovery::

It can be classified into Active and Passive recovery. In Active recovery we have Retransmission and in passive recovery we have Forward Error Correction(Adding redundancy). Retransmission increases the latency of packets and may not be suitable for VoIP.

3.2 CONTROL PLANE

3.2.1 Resource provisioning and Traffic engineering

Refers to the configuration of resources for applications in the network. In industry, main approach of resource provisioning is over provisioning, abundantly providing resources. Factors that make this attractive are cost of bandwidth in the backbone is decreasing, network planning is becoming simpler and hence provision can be planned.[4]

3.2.2 Traffic Engineering::

It mainly focus on minimizing over-utilization of a particular portion of the network while the capacity is available elsewhere in the network. Multi-Protocol Label Switching(MPLS) and Constraint Based Routing(CBR) provide powerful tools for traffic engineering. With these mechanisms, a certain amount of network resources can be reserved for the potential voice traffic along the paths which are determined by CBR or shortest path routing algorithms.

3.2.3 Admission Control::

Limits resource usage of voice traffic within the amount of the provisioned resources. IP networks have no admission control and can offer only best effort service. *Parameter based Admission Control* provides delay guaranteed service to applications which can be accurately described, such as VoIP. When traffic is bursty, it is difficult to describe traffic characterstics which makes this type to overbook network resources and hence lowers network utilization. It uses explicit traffic descriptors to limit the amount of traffic over any period (typical example is token bucket).
Algorithms used in parameter based admission control are::
Cisco's resource reservation based (RSVP), Utilization based (compares with a threshold, based on utilization value at runtime it decides to admit or reject), Per-flow end-to-end guaranteed delay service (Computes bandwidth requirements and compares with available resource to make decision.), Class-based admission control.

4 Performance Improvement on VoIP applications

4.1 End-To-End Delay

When it exceeds a certain value,the interactiveness becomes more like a half-duplex communication. Delay can be of 2 types:: Delays due to processing and transmission of speech and Network delay(delay that is the result of processing in end systems,packet processing in network devices and propagation delay between network nodes on the transmission path)[2]

Network delay = Fixed part + variable part
Fixed part depends on performance of the network nodes on the transmission path, the capacity of links between the nodes ,transmission delay and propagation delay. *Variable part* is the time spent in the queues which depend on current network load. Queuing delay can be reduced by introduction of advanced scheduling mechanisms (Expidited Forwarding and priority queuing). IP packet delay can be reduced by sending shorter packets. Useful technique for voice delay reduction on WAN is link fragmentation and interleaving. Here a longer packet is fragmented into smaller packets and transmitted. In between those small packets, *VOICE* packets are sent.

4.2 DELAY JITTER

Delay variation, also called jitter, obstructs the proper reconstruction of voice packets in their original sequential and periodical pattern. It is defined as difference in total end-to-end delay of 2 consequetive packets in the flow. Removing the jitter requires collecting packets and storing them long enough to allow the slowest packets to arrive in order to be played in the correct sequence. Solution is to employ a playout buffer at the receiver to absorb the jitter before outputting the audiostream. packets are buffered until their scheduled playout time arrives. Scheduling a later deadline increases the possibility of playing out more packets and results in lower loss rate, but at the cost of higher buffering delay.

Techniques for Jitter Absorption

1. Setting the same playout time for all the packets for entire session or for the duration of each session. 2. Adaptive adjusting of playout time during silence periods regarding to current network

delays. 3. Constantly adapting the playout time for each packet, which requires the scaling of voice packets to maintain continued playout.

4.3 FRAME ERASURE(F.E)

It basically happens when the IP packet carrying speech frame does not arrive to the receiver in time. Loss may be single frame or a block of frames.

Techniques used to fight the frame erasure:: 1.Forward Error Correction(requires additional processing) depends on the rate and distribution of the losses. 2.Loss concealment(replaces lost frames by playing the last successfully received frame) effective only at low loss rate of a single frame.

High F.E and delays can lead to a longer period of corrupt voice. The speech quality perception by the listener is based on F.E levels that occur on the exit from the jitter buffer after the Farward Error Correction has been employed. To reduce levels of frame loss, Assured forwarding service helps to reduce network packet loss that occur because of full queues in network nodes.

4.4 OUT OF ORDER PACKET DELIVERY

Occurs in the complex topology where more than one path exists between the sender and the reciever. The receiving system, must rearrange received pkts in the correct order to reconstruct the original speech signal.

Techniques for OUT-OF-ORDER PACKET DELIVERY

It is also done by Jitter buffer whose fuctionality now became 1. Re-ordering of out of order pkts (based on seqence number) 2. Elimination of Jitter.

5 General Call Admission Control Techniques in Cisco VoIP system

A call admission control (CAC) mechanism has to be introduced in IP networks in order to ensure that sufficient resources are available to satisfy the requirements of both the new and the existing calls after the new call is admitted. Current VoIP systems have realized the importance of CAC to provide QoS guarantees. Several CAC mechanisms, such as Site-Utilization-Based CAC (SU-CAC) and Link-Utilization-Based CAC (LU-CAC), have been used in current VoIP systems. However, none of current VoIP systems can really provide QoS guarantees to VoIP. The basic reason behind this is that none of them are able to well apply and support CAC mechanisms. For example, the SUCAC mechanism performs admission control based on the pre-allocated resource to the sites, which can represent a host or a network with different sizes. It demands an approach to do resource pre-allocation to sites at the configuration time. Unfortunately, current VoIP systems, such as the Cisco s VoIP system, have not been able to define such an approach. Resource pre-allocation in these systems is performed in an ad hoc fashion. Hence, no QoS can be guaranteed although the SU-CAC mechanism is applied. Another example is the case of the LU-CAC mechanism. With the LU-CAC mechanism, admission control is based on the utilization of the individual link bandwidth. This mechanism needs resource reservation on individual links in the network[9]. Current VoIP systems rely on resource reservation protocols, such as RSVP, to do explicit resource reservation on all routers along the path of the traffic in the network. Such a resource reservation approach will introduce the significant overhead to the core-routers, and hence greatly compromise the overall network performance.

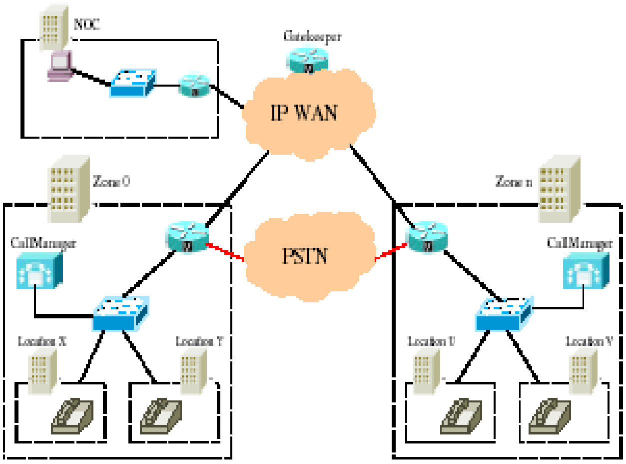


Figure 1: The Basic QoS architecture

6 Typical Control Plane architecture in Cisco VoIP system

A typical multi-site control-plane architecture of the system is shown below. CallManager (CM) is the main component in the architecture, which is a software-based call-processing component. It provides the overall framework for communication within a corporate enterprise environment. As an optional component, *Gatekeeper* can provide services such as address translation and call admission control to the calls and can be configured to work with CMs to do admission control. *CM* and *GK* communicate with each other by using the H.323 signaling protocol[9].

CM as well as *GK* perform admission control for calls inside or outside a corporate enterprise environment, aiming to provide a certain degree of QoS to voice in IP networks. To the call within a corporate enterprise environment, only the CM located in the enterprise environment is invoked to perform admission control. However, to the call across multiple corporate enterprise environments, not only CM (both in the environment where the call is originated and in the one where the call is terminated), but also the related GK(s) may be involved to do admission control.

7 Architecture that can improve Cisco's VoIP System

The architecture consists of 2 important componants as described below

7.1 QoS Manager(QoSM)

QoSM implements three basic functions:: 1.To provide user interface to control and monitor the components in the same QoS domain . In each QoS domain, we can deploy one QoSManager and multiple distributed agents, which are registered to the QoS Manager. 2.To provide registration to the distributed agents and coordination among distributed agents in the same QoS domain. 3.To co-operate with peer QoSMs that belong to other QoS domains.

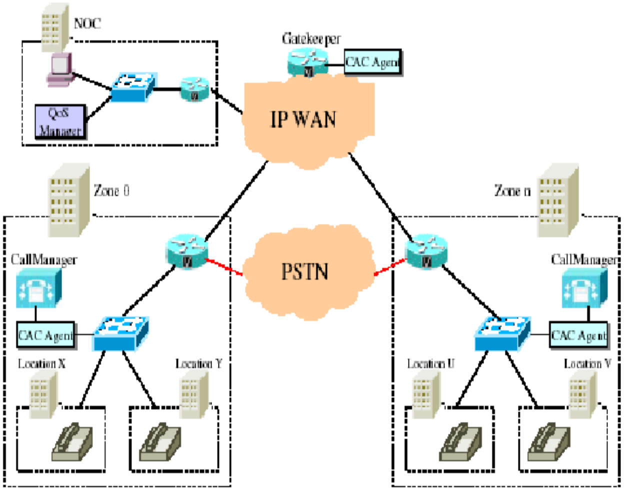


Figure 2: The Proposed System Architecture

7.2 Call Admission Control Agent(CACA)

CACA implements three main functions: 1.Doing deterministic or statistic delay analysis and obtaining the bandwidth utilization. 2.Performing admission control with specific CAC mechanism. 3.Processing call signaling.

8 Architecture of CACA

8.1 Communication Module

It is used to communicate with QoS Manager. For example, the voice traffic model, network topology, and voice traffic deadline requirement can be downloaded from QoS Manager through the Communication Module.

8.2 Utilization CalculationModule

It has two main functions. One of the functions is to compute the maximum link utilization. The other function is to optimize the utilization of resource to prevent wasting too much resource in applying SU-CAC mechanism.

8.3 Admission Decision Making Module

It is used to make admission decision based on maximum link utilization and the incoming call information (i.e., bandwidth required and address of the caller/callee, etc.).

8.4 Call Signaling Processing Module

It monitors and intercepts call setup signaling from Gatekeeper or CallManager, withdraws the useful message and passes it to Admission Decision Making Module, and executes call admission

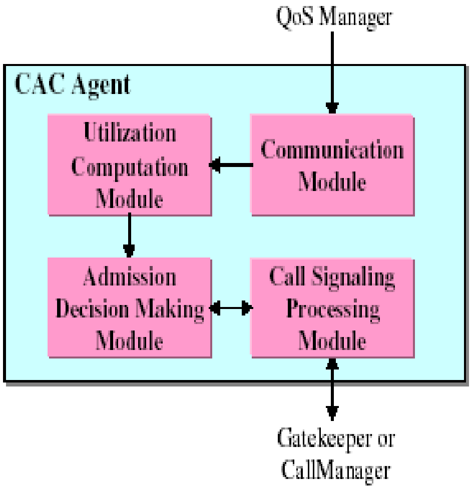


Figure 3: The architecture of CACA

decision made by Admission Decision Making Module. SU-CAC does not require the last two modules as it works on preallocated bandwidths.

9 Utilization CalculationModule

9.1 Utilization Module

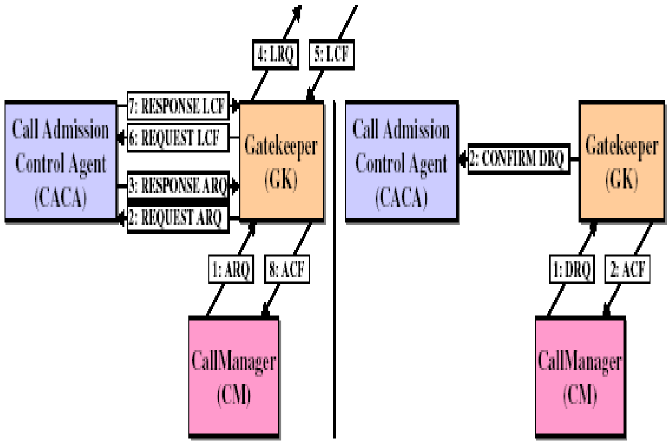
Utilization Computation Module has a utilization verification function. Given the voice traffic model, the network topology, and the voice traffic deadline requirement, for any input of link utilization u, we calculate the worst-case delay (deterministic case) or delay distribution (statistical case) with our delay analysis methods.Then we can verify whether the utilization is safe or not to make end-to-end delay meet the deadline requirement. Using binary searching method for utilization, we can obtain the maximum link utilization. We need information about flow population to have such analysis. In the absence of such information, the worst-case delays must be determined assuming a worst-case combination of flows.

9.2 Optimization of Bandwidth Utilization

As mentioned, SU-CAC mechanism tends to underutilize the network resource while providing end-to-end delay guarantee. One of our objectives is to optimize the overall bandwidth utilization. Given the network topology and the limitation of link bandwidth allocated to voice traffic, to optimize the overall bandwidth utilization, an optimization problem is formed.

10 Signaling Processing and Admission Decision Making in CACA

Basically there are 2 methods for doing this. One is Front-End approach and the other is Back-End approach.



(a) Call Request

(b) Call Tearing-down

Figure 4: An illustration of Successful call request procedure in CACA for GK

10.1 Front-End approach

In this approach, call setup requests and response between endpoint and admission decision unit (e.g., CM) must pass through the agent. The agent can directly enforce its call admission decision to the call setup request by manipulating signaling between the endpoint and the existing admission decision unit.

10.2 Back-End approach

In this approach, call setup requests and responses will be forwarded to the agent by the existing call admission decision unit (e.g., GK). The agent will indirectly execute its call admission decision to the call setup request by negotiating with the existing call admission decision unit.

10.3 Signaling Processing Module for Cisco GK

Cisco GK provides interface for external application servers to offload and supplement its features. The interaction between the Cisco GK and the external application is completely transparent to the H.323 endpoint. The communication between the Cisco IOS GK and Signaling Processing Modules is based on Cisco s propriety protocol, GK Transaction Message Protocol (GKTMP). GKTMP provides a set of ASCII RAS request/response messages between Cisco GK and the external application over a TCP connection. If an external application is interested in receiving certain RAS messages, it must register its interest with the Cisco GK. The following above figure shows how the Back-End approach is used for signaling processing for Cisco GK.

10.4 Signaling Processing Module for Cisco CM

Unlike GK, CM does not provide an external application to supplement Call admission mechanism. So we use Front-End approach for signaling handling as shown in Figure.

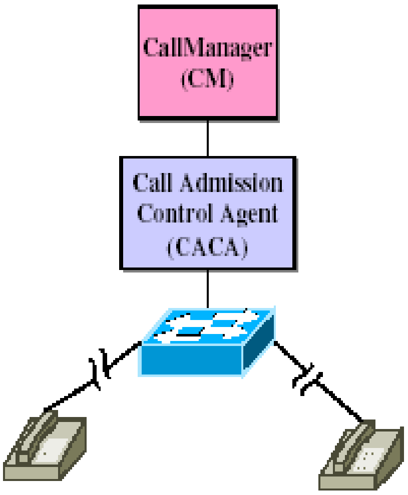
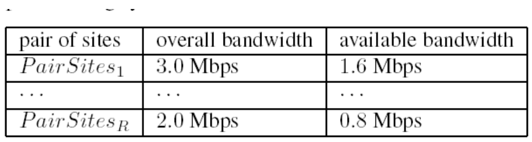


Figure 5: An Illustration of Communicaton protocol for CACA for CM



pair of sites	overall bandwidth	available bandwidth
$PairSites_1$	3.0 Mbps	1.6 Mbps
...	...	...
$PairSites_R$	2.0 Mbps	0.8 Mbps

Figure 6: The bandwidth table in SU-CAC

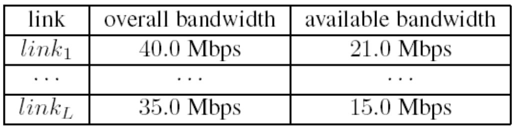
11 Admission Decision by CACA

11.1 Admission Decision Making in SU-CAC

To support the SU-CAC mechanism, the Admission Decision Making module keeps neither the information about the overall bandwidth nor the available bandwidth for each individual link of the network. It takes a fixed amount of bandwidth for each pair of sites or a fixed total amount of bandwidth from/to a site, which is statically configured in Bandwidth Table. Note that the fixed bandwith is allocated by Utilization Computation Module in our QoSprovisioning system. Everytime before admitting, it looks whether it crosses the limit between the sites if it services the present request and takes the appropriate decision.

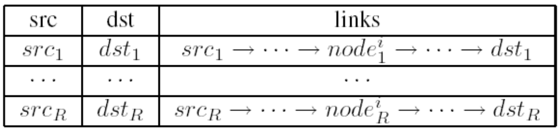
11.2 Admission Decision Making in LU-CAC

To support the LU-CAC mechanism, the Admission Decision Making module has to keep the network topology information and the routing information. There are two tables in supporting this mechanism: Bandwidth Table and Routing Table. Bandwidth Table is used to keep the information of how much of the configured bandwidth on the links is currently consumed by voice traffic and how much link bandwidth is available for calls as shown. Routing table consists of routes between various sources and destination. Everytime before admitting, it looks whether enough resources are available along the path from source and destination and takes appropriate decisions.



link	overall bandwidth	available bandwidth
$link_1$	40.0 Mbps	21.0 Mbps
$\cdots$	$\cdots$	$\cdots$
$link_L$	35.0 Mbps	15.0 Mbps

Figure 7: The bandwidth table in LU-CAC



src	dst	links
src_1	dst_1	$src_1 \rightarrow \cdots \rightarrow node_1^i \rightarrow \cdots \rightarrow dst_1$
$\cdots$	$\cdots$	$\cdots$
src_R	dst_R	$src_R \rightarrow \cdots \rightarrow node_R^i \rightarrow \cdots \rightarrow dst_R$

Figure 8: The Routing table for LU-CAC

12 Interesting reads

This article[1] mainly deals with the Quality of a VoIP service. This gives an overview of state-of-art quality assessment technologies for VoIP, including recent work on improving their accuracy.

This article[10] deals with VoIP mobility in the context of IP and Cellular World. Contains some approaches to VoIP mobility.

This article[8] explores different mobility management schemes for VoIP services, with a focus on Mobile IP and SIP.

This paper[6] presents an overview of convergence of voice and data in satellite communication links with respect to the VIP-TEN and ICEBERGS projects. This paper[11] evaluates QoS of the number of VoIP endpoints, in terms of mouth-to-ear delay, clock skew, silence suppression behaviour and robustness in packet loss.

A guide for QoS for IP Telephony is available at www.cisco.com [7].

QoS technologies can allow service providers to offer more services, such as real-time traffic support, or specific bandwidth allocations at the time of Service Level Agreement(SLA). This paper[3] are many ways of Providing Differentiated Services for the traffic, out of which we implemented it using one of the metrics “Delay”.

13 Conclusion

The introduction of QoS to IP networks does have effect on all the four performance measures such as delay, jitter, frameloss, out-of-order packets. It is therefore compulsory that the network equipment needs QoS mechanisms into IP networks. QoS today comes in many different flavours. It can be offered in two basic ways. Absolute QoS levels that are offered by technologies such as ATM and RSVP and relative QoS levels are offered by TOS field in IP networks. In this paper an overview of a new implementation of QoS system has been discussed which can improve the effiency of the existing Cisco VoIP system along with discussion of many other performance improving parameters.

References

[1] N. K. Akira Takahashi, Hideaki Yoshino. Perceptual qos assessment technologies for voip. IEEE Communications Magazine, 2004.

[2] S. T. Anton Kos, Borut Klepec. Techniques for performance improvement of voip applications. IEEE, 2002.

[3] A. K. Borut Klepec. Performance of voip applications in a simple differentiated services network architecture. IEEE, 2001.

[4] X. L. M. W. Z. Chen, Wang. Survey on qos management of voip. Proc. International Conference on Computer Networks and Mobile Computing(ICCNMC'03), 2003.

[5] B. Goode. Voice over interner protocol. Proc.of IEEE, 2002.

[6] Z. S. B. C. H.Cruickshank, Antonio Sanchez. Voice over ip over satellite links. IEEE, 2001.

[7] C. Systems. Cisco ip telephony qos design guide. http://www.cisco.com.

[8] S. D. S. D. Ted Taekyoung, Mario Gerla. Mobility management for voip service. IEEE Wireless Communications, 2002.

[9] M. X. Z. Wang, Mai. Implementation of qos provisioning system. Proc.of IEEE Real-time and Embedded Technology and Applications Symposium, 2002.

[10] J.-C. L. Wanjiun Liao. Voip mobility in ip/cellular network internetworking. IEEE Communications Magazine, 2000.

[11] S. Wenyu Jiang, Kazummi Koguchi. Qos evaluation of voip end-points. IEEE, 2003.