Quality of service for IP voice services - is it necessary?

Marko Luoma [a], Markus Peuhkuri [a], Tomi Yletyinen [b]

[a] Helsinki University of Technology, PO Box 3000, 02015 HUT, Finland
[b] Teleware Oy, Itäkeskuksen Maamerkki, 00930 Helsinki, Finland

ABSTRACT

New services come at cost to the Internet community, therefore determining the needed level of QoS for the services is important. This due to the fact that guaranteeing quality demands extra functionality and extra cost for equipment in the network. We have chosen voice transmission over IP network as a candidate service for the study of QoS issues. We wanted to see what is the difference in perceived quality of service in following cases: a network with conventional forwarding and a similar network using a layer 3 switching (IP-switching). We present measurements with a commercial embedded microprocessor system and a public domain software in a general purpose computer. From experiences we can say that the present general purpose computer architecture is no means optimal for providing controlled quality of service for a real-time communication. We conclude that routers offering real-time services need prioritized traffic handling.

Keywords: ATM, IP, Switching, Measurement, QoS.

1. INTRODUCTION

Internet, the world largest ad-hoc network is running on the basis of solidarity and good will of users and operators. The Internet has come to a turning point. The question is, whether new quality of service (QoS) features should be implemented or should the two decades old way of operation be pursued.

Internet works like a socialistic community. It offers equal misery for everybody, which on the other hand means equal opportunities. It has been argued that new services such as an IP telephony need affirm quality of service scenario in order to work properly, but no real information about experiments with live traffic exists. New services come at cost to the Internet community, therefore determining the needed level of QoS for the services is important. This due to the fact that guaranteeing quality demands extra functionality and extra cost for equipment in the network. One cannot rely on end users and terminals that they provide right information and obey regulations in all cases. In short we can say that “Quality without cost is not guaranteed quality."

Quality of service, in short, means more control over resources. Resources need to be allocated or shared based on the requests of the end systems. Requests have to be mapped to contracts, which can be policed efficiently and charged accordingly. This is a radical change in the way of thinking for the Internet. Currently the bits in the Internet are essentially free for the users. Charging is not in general dependent of the distance traveled or the quantity of the cargo.

The QoS scenario leads to two distinct problems: signaling and charging. Additional QoS signaling in the Internet leads to the necessity of changing our existing software and hardware. Common argument has been that IP is so widely deployed and there is no need to change it in order to scale to the future, this seams to crack with QoS issues.

We have chosen voice transmission over IP network as a candidate service for the study of QoS issues. We wanted to see what is the difference in perceived quality of service in following cases: a network with conventional forwarding and a similar network using a layer 3 switching (IP-switching). IP-switches are capable of offering prioritized service; one way to introduce QoS to network.

Structure of this paper is following. First we will introduce market idea of voice over IP. Then we will explain briefly in chapter 3 some ways to indicate quality of service (QoS) requirements to the network. On chapter 4 we will explain our measurement setups and present preliminary results. Last we will conclude our paper and present some ideas for further studies.

2. VOICE OVER IP

At first the transmission of voice over data networks seem like a bad idea. We already have a well functioning circuit switched phone network that extends through out the seven continents, and forms the largest machine ever built by man. Data networks on the other hand are presently ill-suited for the transmission of voice. Voice is real-time information and requires real-time handling from the network that one can not found in contemporary data networks. Still IP-voice has found a market, the main driver being surpassing the costly long-distance telephony. The popularity of virtually free long-distance calls is proving the fact that even poor quality is satisfactory if the price is right. Old American sales strategic motto: “In a competitive and developed market three things are important: price, price and price" have proven right again. In the future long-distance call charges will sink, not because of VoIP, but because of the increasing competition. The cost advantage of Voice over IP (VoIP) is likely to diminish, but still market experts are predicting a bright future for it. Because of statistical multiplexing and advanced compression methods VoIP should in principle be more cost effective than circuit switched voice transmission. Video conferencing and computer supported collaborative work (CSCW) applications are the new drivers of VoIP; providing added value for voice communication.

Increasing popularity of the Internet has lead operators to enhance their access networks for more suitable data delivery. This has introduced market for different kinds of access line technologies, mainly ISDN and xDSL. Both of these suits for data traffic extremely well and also offer possibility for POTS connections. Question has become more and more relevant: Is IP capable to offer decent QoS for voice even in residential usage? In case access line technologies' scale to the speeds of tens even hundreds kb per second the question moves to the core network. Is IP and its companion protocols able to offer stable service in dynamic environment of Internet?

3. WAYS TO INIDICATE QOS IN IP

There are many ways to introduce QoS capabilities to the network. Nevertheless the question always relates to two aspects:

- Mechanisms that actually implement QoS
- Mechanisms that relay the information of required QoS.

3.1 TOS

The second octet of the IP-packet header defines the type of service (ToS). TOS is divided in two fields: the precedence and the type of service. Precedence shows the priority and the type of service indicates routing. Routing protocols are supposed to compute: a default route when no ToS bit is set, a shortest route (D set), a largest throughput route (T set), a most reliable route (R set) and a cheapest route (C set). The bits should not be normally combined.

The ToS is a way for a user or an application to inform network if packet contains real-time information that requires low delay and high priority. In general the ToS is not used in the Internet because of fear of misuse. Type of service routing is defined in some routing protocols (BGP and OSPF) and some routers are capable of priority queuing necessary to provide different service categories. Recently there has been available solutions, as IP PBX studied in later chapter, which uses TOS bits.

<table>
<tr><td>Ver</td><td>IHL</td><td colspan="2">ToS</td><td colspan="2">Total length</td></tr>
<tr><td colspan="4">Identification</td><td>Flags</td><td>Fragment offset</td></tr>
<tr><td colspan="2">Time to live</td><td colspan="2">Protocol</td><td colspan="2">Header Checksum</td></tr>
<tr><td colspan="6">Source Address</td></tr>
<tr><td colspan="6">Destination Address</td></tr>
<tr><td colspan="4">Options</td><td colspan="2">Padding</td></tr>
</table>

Figure 1 IP header with ToS bits highlighted

3.2 DIFFERENTIATED SERVICES

Differentiated services (DS) are intended to provide scaleable service discrimination in the Internet without the need for per-flow state and signaling at every hop. The differentiated services approach to providing quality of service in networks employs a small, well-defined set of building blocks from which a variety of services may be built. The services may be either end-to-end or intra-domain.[1] A wide range of services can be provided by a combination of:

- Setting bits in the DS byte (eq. TOS octet) at network edges and administrative boundaries
- Using DS byte to determine how packets are treated by the routers inside the network
- Conditioning marked packets at network boundaries in accordance with the requirements of each service

A differentiated services capable network node includes a classifier that selects packets based on the DS byte and is capable of delivering the treatment corresponding to that marking of the DS byte. Setting of the DS byte and other conditioning of the dynamic behavior of marked packets need only be performed at network boundaries and may vary in complexity.

IN	PHB	0 0

Figure 2 Structure of DS byte

Per hop behaviors (PHB) vary based on the mechanism used to create them and wishes of single ISP. Routers may have number of parameters that can be adjusted in order to achieve certain kind of classification for traffic. DS byte is used to select appropriate forwarding function in single router. Selecting what per hop behavior should be implemented is matter of ISP therefore PHBs can vary between operators. It is noted in many occasions that experiences will eventually lead to harmony with different ISPs and their implementation of PHBs.[2]

3.3 SIMPLE INTEGRATED MEDIA ACCESS

Simple Integrated Media Access (SIMA) can be seen as one way to implement differentiated services into Internet. With SIMA one have all the necessary elements for a charging scheme, linked with a corresponding traffic control system. The primary idea of the SIMA service is to maximize the exploitation of network resources with a simple control scheme while keeping the ratios of QoS levels offered to different flows unchanged under changeable traffic conditions. The maximization is based on three key features: all flows with different QoS requirements share the total capacity of every link, the network attempts to avoid any unnecessary packet discarding, and flow (or call) level blocking can be totally avoided. The approximate constancy of QoS ratios and simplicity are achieved by using 8 priority levels that make possible a fair packet discarding scheme inside the network without keeping track on the traffic of every flow.[3]

Heart of SIMA is the scheduling and buffering unit. It deals the extra complexities introduced by SIMA. Incoming packet train is compared with negotiated nominal rate and network load. Based on actual and nominal rate a priority is assigned to each packet. In addition to that there is a separate queue for real-time packets that is always served in highest priority.

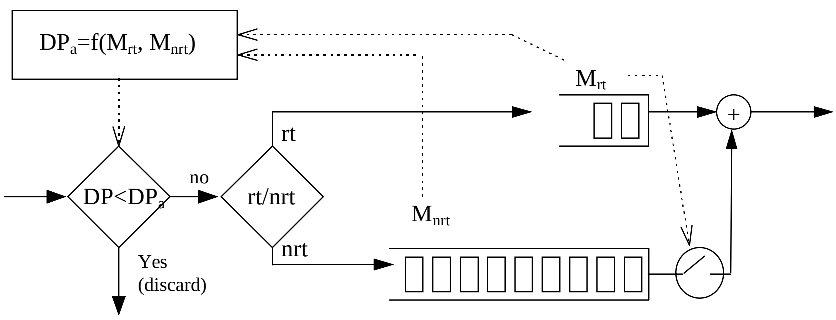

Figure 3 Scheduling and buffering unit of SIMA networking concept

3.4 INTERNET INTEGRATED SERVICES

The Internet integrated services framework (IIS) provides the ability for applications to choose among multiple, controlled levels of delivery service for their data packets. To support this capability individual network element along the end-to-end path of the application's data packets must support the mechanisms to control the quality of service delivered to those packets. This function has been specified by the guaranteed service and the controlled-load service. Furthermore, the applications requirements must be conveyed to the network elements along the path together with QoS management information. This can be implemented in a number of ways, of which resource reservation protocol (RSVP) is one alternative.

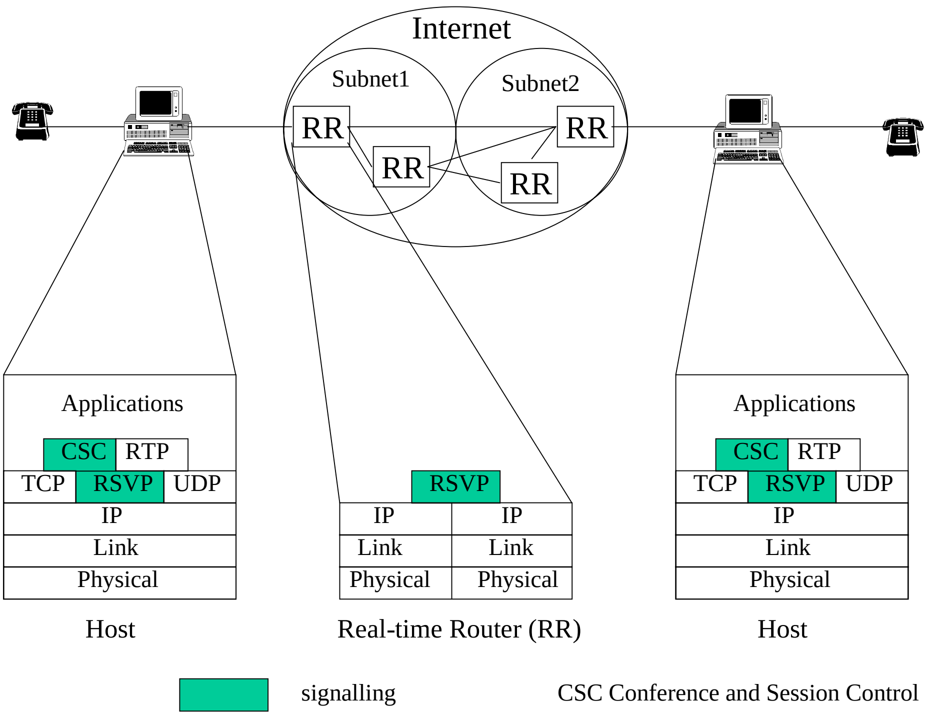

Figure 4 Real-Time Internet [4]

The original reference implementation framework for real-time router included four components: the packet scheduler, the admission control routine, the classifier, and the reservation setup protocol. The implementation framework for a host is generally similar to that of a router, with the addition of applications. Rather than being forwarded, host data originates and terminates in an application. An application needing real-time QoS for a flow must somehow invoke a local reservation setup agent. The way might be an explicit API, such as Winsock2 for network resource setup, or the setup might be invoked implicitly as a part of the operating system scheduling function.

3.5 IP SWITCHING

The QoS of IP switching is policy based as opposed to RSVP and ATM contract-based QoS, where the user requests a certain degree of QoS. The policy basis means that the network manager sets policies according to which the traffic is given QoS differing from that of the default connection. The policies can be based on IP addresses, IP ToS, TCP/UDP port, or a combination of the three.

Different QoS is achieved with:

- Flow actions: The flow can be cut through the switch or it can go through the default path through the processor.
- QoS actions: A traffic flow can receive four priorities: high, medium, normal and low.

IP switching uses Class-Based Queuing (CBQ), a combination of priority scheduling and FIFO queuing. When a source (a flow) exceeds its target rate the excess traffic is discarded or put to a buffer called waiting room, i.e. incoming flows go first through a class policer. The packets conforming to the rate and the packets in the waiting room are then run through a Weighted Round Robin (WRR) scheduler to the priority scheduler, from where they are forwarded to the link.

The IP switching way of improving QoS has been argued to have draw-backs, even to the extent that no improvement in QoS might be seen.

At least four types of problems could occur:

- The switch controller might select a route that can not meet QoS requirements. This could happen because RIP does not support QoS routing and OSPF has only limited capabilities.
- Traditional packet forwarding is used until all the switch controllers along the route have established a virtual circuit path. The performance before this can be poor.
- The current implementation of IP switching only supports the limited IP header ToS for defining QoS.
- No mechanisms for ensuring that all cut-through stubs in each switch along the path are offering the same level of QoS, i.e. there are no guarantees of a consistent level of end-to-end quality.

4. MEASUREMENT STUDIES

4. 1 CLIENT STUDY

We did measure two different workstation environments: Pentium 100MHz PC with Windows95 operating system and SUN UltraSparc 170E with Solaris 2.5 operating system . The goal was to resolve possible bottlenecks inside client system. These bottlenecks are either on hardware, i.e. audiocard or software, i.e. coding and packetizing.

First we wanted to measure the hardware caused delays in the workstation. For this the setup was that the trigger 1 of the HP 5300A delay measuring device connected to the line input of audiocard with pulse generator and trigger 2 to the line output of the audiocard. Measurement was done by directing the line input to the line output. This way input signal was A/D-converted and then D/A converted in the audio hardware. The samples did not go to the VoIP program for coding, and thus no buffering and coding delays were introduced. The measured delay was between 7,77 and 12,85 ms, with an average of 8,9 ms.

In the next measurement we wanted to find out the total delay in the workstation including the buffering and coding delays of the VoIP client software (Nevot). Nevot uses RTP over UDP/IP to deliver voice packages. In the software the packet length was adjusted to 20 ms. This means that the input was sampled at 8000 Hz (one sample = 125 us), and that a 20 ms block of these samples was coded and packetized at a time. The program must therefore buffer the samples in memory before coding. In Nevot we used the monitor option so that the coded samples were decoded and then D/A-converted and fed into the output. The delay varies between 100 and 120 ms. Subtracting the coding and decoding sample lengths of 2 x 20 ms and the hardware delay of around 9 ms gives a buffering and processing delay of 54 ms. It should be noted that no playout delay was inserted.

Currently the most popular Voice over IP platform is a Pentium PC running Windows 95 and one of the VoIP clients such as Microsoft Netmeeting, Intel Internet Video Phone and Vocaltec Internet Phone. We used two Pentium PCs running Windows95 and Microsoft Netmeeting. PCs were connected together using emulated ATM LAN at 25 Mbps NICs.

For PC lack of debugging options in software forced us to measure end to end delays in order to resolve software delays. Measured delays for different components were:

- Terminal hardware delay 5,4 ms
- Network delay <1 ms

- End-to-end delay over LANE 190 ms

Which leads to the conclusion that buffering, coding and processing delay was about 89ms.

4.2 LAN STUDY

4.2.1 LAN MEASUREMENT

In LAN measurement we had two Sun Ultras connected to a shared but separated 10BaseT Ethernet. On both workstations we had VoIP clients, Nevot (Network Voice Terminal) version 3.34 and Sun standard audio hardware. To the workstations we attached an HP delay meter. In addition to this using the debugging option of Nevot the timestamps of incoming audio packets at the receiver were recorded. The timestamps can be used for calculating the packet spacing difference of a pair of packets at the receiver compared to that at the sender. From the packet spacing differences we can, for example, calculate network caused packet interarrival jitter of the audio stream (most stressing QoS parameter for voice).

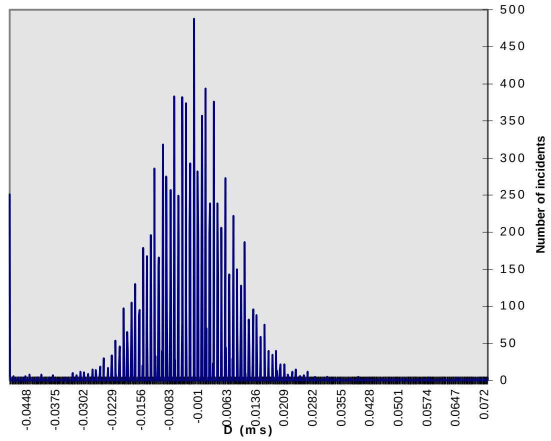

Figure 5 IAT for LAN with no load

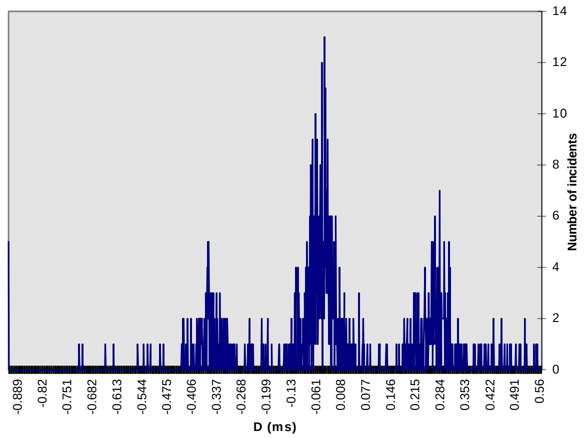

Figure 6 IAT for LAN with small packet load

Figures 2-4 shows clearly effect of contention to packet inter arrival times (IAT).Inter arrival distribution expands as large as 200ms which in some cases can mean intolerable variation for playout delay optimization.

4.2.2 SUBJECTIVE STUDY

We made subjective study for IP voice in laboratory environment where two persons were communicating with other using LAN connected computers in separated segment. This LAN segment was loaded with third computer running floodping with different packet sizes. With floodping we were able to introduce collisions to LAN and extra load to another machine. These collisions introduce distortion to the audio signal which by assumption is annoying for human ears.

Subjective study was carried out using same equipment that was used in LAN measurement as a part of University course in Laboratory of Telecommunications Technology. Students were introduced to the concept of voice over IP and after that they made subjective and objective evaluations about quality of voice connection. Subjective studies showed following points:

- Quality of speech is intelligible when load of the network is small or moderate.
- Quality of speech is better than in GSM but worse than in PSTN.
- With small or moderate load one can recognize other speaker by voice but when load is high there is so much distortion that recognition is impossible.

- Delay of 200ms was found out to be maximum tolerable delay for conversation, after that conversation turns to 'half duplex'.
- Adjustment of playout delay is crucial. More delay less cracks in voice but intelligity is jeopardized if delay is more than 20ms.
- If playout delay is less than delay variance in network, use of service is inconvenient.
- Speech color is metallic if delay variance increases near to playout delay value.
- Crackles and gaps (in small amounts) were seen better option than metallic color of voice and possible excessive delay causing conversation to turn 'half duplex'.
- Echo was received after a second or two which made conversation annoying. Echo cancellation decreases echo but does not remove it.
- Quality of microphone/headset is more stringent than in normal telephony cause computer generates background noise which disturbs conversation (distortion) if grabbed into connection.
- Normal telephone handset was seen better option than headset.
- Silence detection was not good enough because computers generated high background noise.
- Usage of offered level of QoS depends highly on price.

4.3 IP FORWARDING STUDY

In this measurement we studied an IP voice connection over an IP switching network that was set to packet store and forward mode. In other words the network did not make any cut through connections over the ATM switch fabric. All packets including our long lived IP voice flow were processed in the conventional IP router way. The sending voice terminal was connected with a 100BaseT Ethernet to an IP Switch Gateway router. From there the packets were routed and forwarded through an ATM link running on STM-1 to an ATM switch through the default VP/VC. From there the packets go through the default VP/VC to the IP switch processor. The IP switch processor routed and forward the packets to the neighboring IP switch processor ATM switch pair. There the switch processor routes and forwards the packets to the next IP switch, which is an IP switch gateway. Our receiving voice terminal was connected to this gateway using 100BaseT. On the receiving terminal we were again recording packet timestamps.

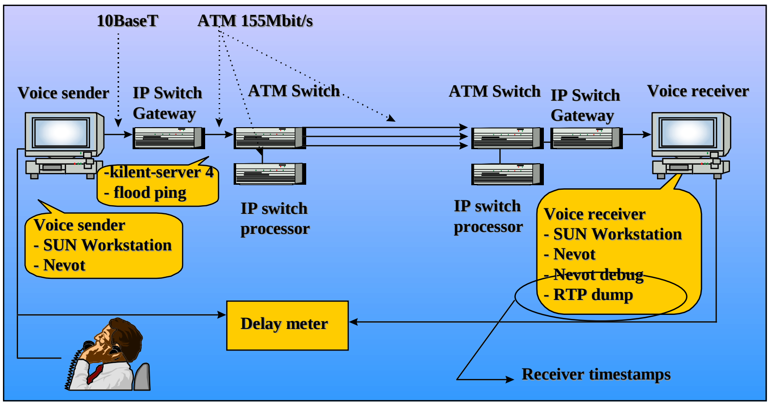

Figure 7 Measurement environment

We wanted some understanding of the network behavior under load. We had the assumption that the router functions in the IP switch processors and the border gateways would be the bottlenecks, causing delay variation to the voice stream. We wanted to push the forwarding functions in the switch processors. Caching routes relieves the route table look-up bottleneck in a small campus network like ours, so there was no point in trying to push the loading of that further. We decided to use a more direct approach. Flooding ping is a very effective way to load a network, it sends packet as soon one gets reply to previous or 100 packets per second, whichever is more. We ran the ping processes from the routers themselves in order not to load the LANs in the edges. This way we get only the performance difference between IP switching and IP forwarding in the two core switches. The workload on the edge routers and the traffic in the connected LANs is equivalent in the IP switching and IP forwarding case. We made several measurements with raw TCP-traffic loading the network and causing voice stream delay variance. Using a client-server network traffic generator and throughput measurement application (Kilent-Server version 4 (KS)) we loaded the network with large, 1500 byte TCP-packets. The client application running on the IP switch gateway sent the packets to the server application running on the other end of the network in an IP switch gateway. When running the client application it starts sending TCP packets between the server and client using a user defined port. The client starts a timer and starts sending packets with MTU size 1500 bytes and packet frame size of 1500 bytes. We experimented with one, two, four and eight sending processes [6]. The average load in the network did not increase notably with the addition of sending processes which implicates that the sending gateways forwarding function was saturated already with two processes. The CPU loads were from 20% to 40% on the gateways, which also supports this.

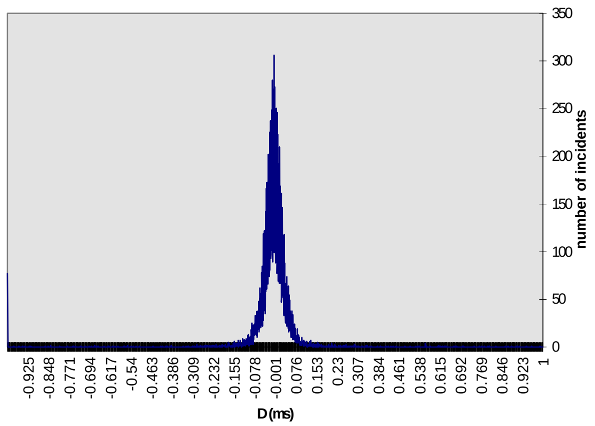

Figure 8 IAT distribution with no load

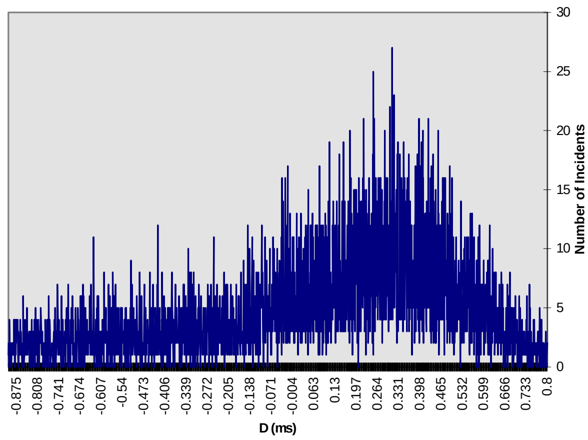

Figure 9 IAT distribution with KS-2

4.4 IP SWITCHING MEASUREMENT

4.4.1THE SETUP OF THE MEASUREMENTS

In this measurement we studied an IP voice connection over an IP switching network that was set to establish switched flows for long lived connections ports and IP addresses. All long lived IP packet flows were identified (more than five packets), and cut through connections through the ATM switch fabric were established. The IP switch gateways on the edges were still using conventional store and forward, because cut through connections can only be established between IP switches. The sending voice terminal was connected by a 100BaseT Ethernet to an IP Switch Gateway router. From there the packets were routed and forwarded through an ATM link running on STM-1 to an ATM switch through the connection specific VP/VC. There using ATM cell switching the packets in cells are switched to the correct port on the correct VP/VC. From there the packets in cells go to the neighboring ATM switch input. Cell switching is used and the flows switched to the correct port from where they finally go to the border IP switch gateway. The gateway routes and forwards the packets to the correct LAN. Our receiving voice terminal was connected to this gateway using 100BaseT. On the receiving terminal we were recording packet timestamps.

4.4.2 RESULTS

The results for IP switching show that IP switching is really capable to offer more effectively bounded delay variation. Pictures 12 and 13 show similar incidents that pictures 9 and 10 for IP forwarding. One can easily notice that distribution is narrower for IP switching than for forwarding. Picture 16 shows mixed load situation where IP switches were loaded not just ping or KS but mixture of those two.

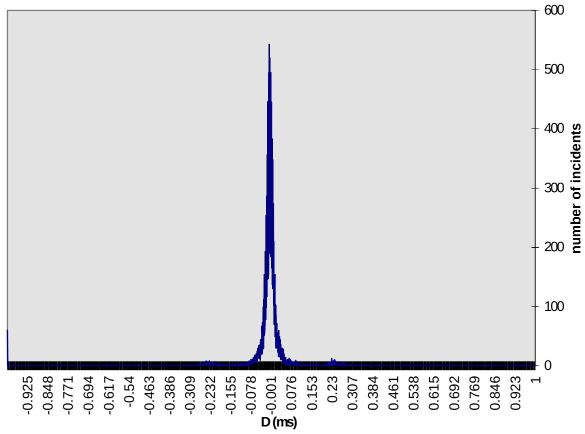

Figure 10 IAT distribution with no load

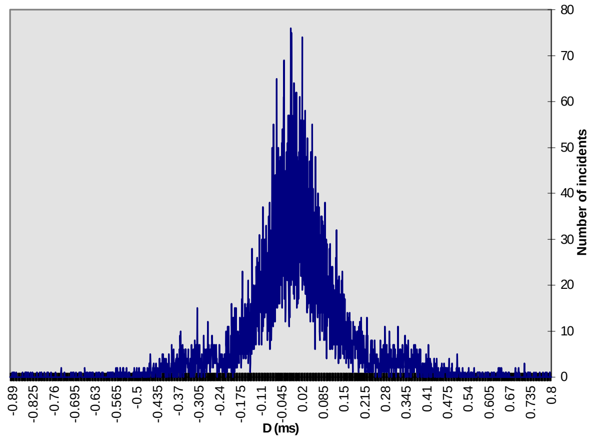

Figure 11 IAT distribution with KS-2

The worst case we were able to achieve, a measurement with the network loaded with both flood ping and Kilent-server is in figure 15. The 85th percentile of D is around 0 ms and the 99th percentile is around 10 ms.

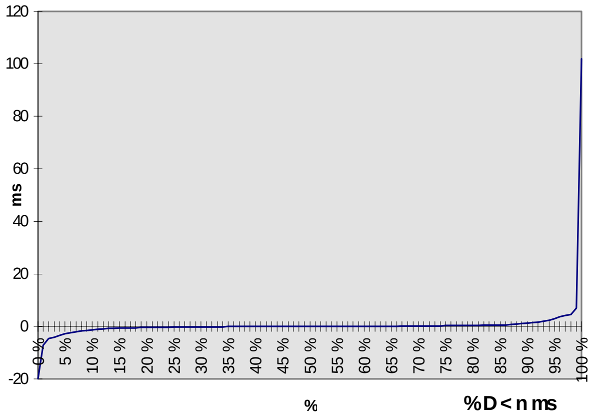

Figure 12 IAT percentiles with mixed load

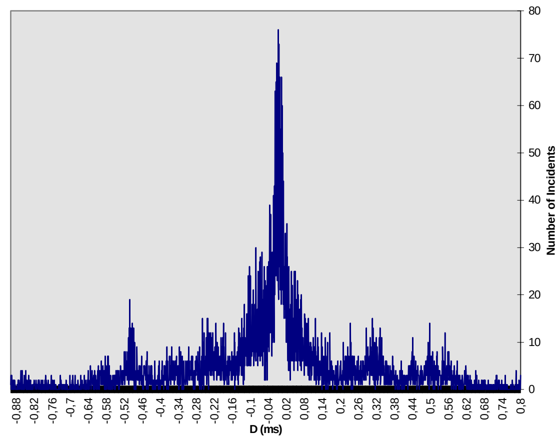

Figure 13 IAT distribution with mixed load

4.5 IP PBX STUDY

4.5.1 LAN STUDY

We made also short subjective study about QoS of commercial IP PBX which uses ‘ordinary’ telephones connected to LAN. Ordinary in sense that they look like an ordinary telephone but they are actually highly sophisticated computer units. These phones were connected to an Ethernet segment which was first separated from rest of world and then interconnected to the university network. This IP PBX is a Windows NT -based call manager (CM) software which supervises phones and deals with call setups and supplementary services. Implementation is based on control channels on top of TCP/IP and information channel on top of UDP/IP. Measured system implemented TOS implication for routing. Control packets were marked for normal TOS whereas conversation packets were marked for low delay. This indicated that system is designed to interoperate routers capable of offer prioritized services.

System has two option for each subscriber class: local and remote. Local subscriber has usually better connection and more bandwidth, so it is possible to utilize less efficient coding; normally allowing better quality for voice. In this case coding for local subscribers is G.711 also known as PCM coding. For remote subscribers resources are more scarce this forces to use more efficient coding methods, often leading worse quality of voice, like GSM coding. In this case coding is G.723 also known as QCELP.

G.711 produced UDP packets of size 260bytes. These packets were sent to network so that during talkspurt the average sending rate was 63,6kbit/s. Because of silence detection longer time average is about 51,3kbit/s.

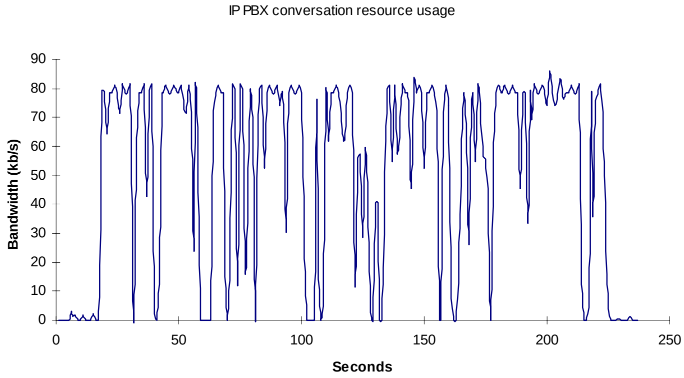

Figure 14 Usage of resources in IP PBX conversation with G.711 coding

G.723 produced UDP packets of size 44bytes. These packets were sent to network so that during talkspurt the average sending rate was 13,1kbit/s. Longer time average is about 5,3kbit/s.

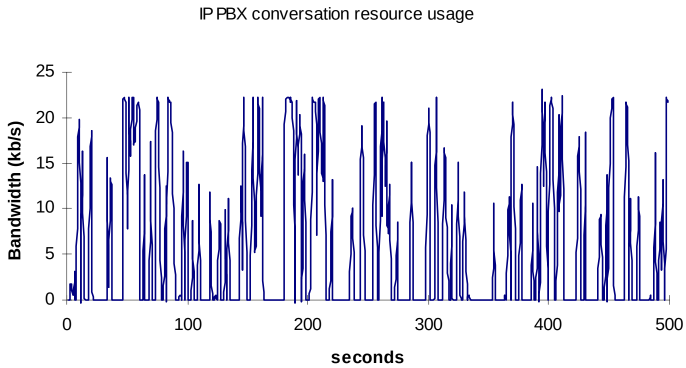

Figure 15 Usage of resources in IP PBX conversation with G.723 coding

Difference in subjective quality with these coding methods was negligible. G.723 showed even better when there was background noise. This seems to be due to the compression algorithm used. Overall quality is excellent, loading in 10 Mbit/s Ethernet could be as high as 5Mbps and still one could not hear any disturbing effects in voice. One can presume that system implements dynamic algorithm for playout delay since there was an increase in transfer delay through network when the load was increased. This algorithm smoothes arrival time differences of voice packets fairly well after 6Mbps load one could hear gaps in voice and also communication started to turn half duplex. Worst degradation was the fact phones occasionally rebooted due to the excessive collisions in Ethernet. Actual reason was that the phones need to change messages with PBX in regular intervals and collisions caused timers to expire before communication was successful.

4.5.2 EMULATED INTERNET STUDY

In order to clarify whether these commercial IP phones will survive in real Internet, where heavy and usually random delay can range from tens to hundreds of milliseconds and even losses of 30% can occur, we emulated Internet environment with NIST NET [5]. NIST NET is an free network emulator software on Linux router which allows extensive parameter adjustment even for single IP flow. Parameters that had meaning for our study were delay, delay variance, packet losses, allowed bandwidth and packet duplication.

Values for these parameters we resolved from our former and on going measurements of delay to various Internet sites. We have been running ICMP echo request with various sizes to numerous Internet hosts in order to resolve delay and connectivity of Internet routes. Based on these measurements we have categorized values for different parameters and regions. Values of Asia, Africa, Russia and Japan are totally based on present status but other values are more accurate, longer term results. Bandwidth values were not probed so values were chosen so that for each case the boundary limit for decent quality were resolved.

This study was made only for G.723 coding, because it was seen only viable option for large scale Internet usage. Required bandwidth for G.711 is too much for many cases and subjective studies in LAN case showed that quality of voice was comparable with G.723.

Subjective studies showed that amount bandwidth needed is the bandwidth of talkspurt. Lower values do not provide decent quality for voice. Voice turns to sound mumbling which is very stressing if used for a long time. Packet losses can be toleraetd up to 15% which comparatively high value if considered values in current Internet. Regions that do not satisfy this are mainly areas of Asia and Africa. Delay variation was found to have great influence to intelligence of speech. More than 20ms variation started to made listening of speech bit by bit inconvenient. This is due to playout delay compensation which was probably not able to adapt variances higher than that. Delay was not so important as we expected. Man on the moon effect started to be noticeable after 150ms but became problem after 300ms.

Region	Delay (ms)	Delay variance (ms)	Packet Loss (%)
Finland	18	10	0.5
Western Europe	58	12	2
Russia	275	30	17
Asia	260	25	22
China	2930	170	20
Japan	250	16	20
Australia	356	16	15
USA	115	10	6
Africa	585	60	18

Table 1 Average delay and loss parameters for various regions in Internet

Based on previous aspects we can say that our IP PBX system is operable in Western Europe and in USA. Some values were a bit high in some cases but overall quality of speech was satisfactory. For cases were losses and variances were too high one can not use implementation like one which we have. For large delay variations more adaptive playout delay is required but it comes on the expense of conversation convenience.

CONCLUSIONS

We have presented some preliminary results from quality of service measurements in IP voice system. First we introduced IP voice platform and ways to indicate QoS requirements to the network. Then we presented measurement results from public domain software (Nevot and Netmeeting) in SUN and PC environment using LAN, IP forwarding and IP switching. Last we presented some results from commercial IP PBX system in separated, loaded network and emulated Internet.

Conclusions from measurements are that in order to have a decent quality for the voice, IP routers must offer prioritized packet handling in forwarding. Prioritized packet handling makes soft guarantee for delay variance and packet losses due to congesting file transfers. Introducing of new concepts like differentiated services and resource reservation protocol is seen necessary if real-time service are to be used in wide extent in Internet.

We have presented subjective studies about quality of service in IP PBX environment. Next question is to verify our results through measurements. We see that more and more questions in traffic prioritization need to answered before real-time services can be deployed widely. Some of them relating to actual parameters of communication and some to the adaptivity of control functions to dynamic fluctuation of Internet traffic profiles.

ACKNOWLEDGEMENTS

This work has been done in the Laboratory of Telecommunications Laboratory of Helsinki University of Technology as joint research study in MITTA- and IPANA-projects.

REFERENCES

1. http://diffserv.lcs.mit.edu/
2. K. Nichols, S. Blake, “Differentiated Services Operational Model and Definitions”, INTERNET-DRAFT, 15 p., February 1998
3. J. Ruutu, M. Kalervo, K. Kilkki, C. Lanting "Multimedia Services Over an Internet: Horizontal and Vertical Aspects of Quality of Conveyance" in Interworking'98 -conference, Ottawa, Canada, 6-10 July 1998.
4. Eriksson A. Real-Time services over the Internet, ISS97.
5. http://www.antd.nist.gov/itg/nistnet/
6. Yletyinen, Tomi, “The Quality of Voice over IP”, Master`s Thesis, HUT / Department of Electrical and Communication engineering, 1998