MEASURING INTERNET TELEPHONY QUALITY: WHERE ARE WE TODAY ?

Olof Hagsand
SICS
Sweden

Kjell Hanson
Ericsson Business Networks
Sweden

Ian Marsh
SICS
Sweden

Abstract

Users of Internet telephony applications demand good quality audio playback. This quality is largely dependent on the instantaneous network conditions. In this paper we describe a scheme for measuring network connections as well as a motivation for including a new metric when assessing quality. The tests included a wide range of geographically distributed sites and our results can give useful feedback to users, operators and developers of Voice over IP applications. The results indicate that Internet telephony is feasible on todays Internet but we should envisage some problems if the Internet continues to grow at the rate so far.

1 Introduction

The current best effort Internet makes no *guarantee* about the delivery of data in real time to applications such as Internet telephony. Due to the shared nature of many of the resources as well as propagation delays it is a difficult, if not impossible, task. All is not lost however, under most conditions in a well dimensioned network timely delivery of packets is achievable.

Due to the number of simultaneous connections and relative busy periods (9am-5pm) across connected sites, the instantaneous quality of connections can vary dramatically. The goal of this paper is to show the benefit of taking and measuring quality so action can be taken immediately.

Internet telephony and voice over IP applications are already being used on the Internet. Most of them[1] transport data using the Real Time Protocol (RTP) and report statistics using the RTCP, the Real Time Control Protocol [SCFJ96]. A RTP/RTCP sender can obtain information on the network condition, more specifically: total packet in this session, packet loss since the last message, variance in packet arrival (jitter) as well as the round trip time (RTT).

Many studies and measurements have been conducted on the Internet. Two of the most cited (and complete) works were done by Vern Paxson [Pax96] [Pax97] where amongst other items he studied routing behavior in the Internet. We believe this is affect Internet users more and more not least of them Internet telephony users. Work dedicated solely to measurement of audio data include Jean Bolot et. al in 1995 [BCG95] where they performed loss measurements and developed a model indicating that forward error correction would rectify most situations where loss is a problem. More recently, in 1997 work done by Mexemchuck and Lo [ML97] defined "quality" of a connection as the fraction of the time that a channel is free of distortion for intervals that are long enough to transmit active speech segments. In this work they consider delay and show numbers of connections that exceed a given delay bound within a given state, the country and trans-Atlantic. Many more papers describing "telephony type" applications give results from experiments and simulations.These are cited in later sections.

In this work we not only look at loss and delay but take a close look at interarrival variance or *jitter*. This gives an accurate picture of the current state of the network as it uses each packets arrival as "data". Plotting jitter can show differences in connections and a mean value can be calculated also to compare connections. In addition interarrival distributions give useful hints tor application writers when designing adaptive playout schemes. In the next section we give some motivation for designing these tests.

[1] A continually updated list of these applications can be found at http://www.cs.columbia.edu/~hgs/rtp.

2 Measurements

2.1 Motivation

In addition to the simple quality argument we have found a number of motivating reasons for performing these tests:

- Firstly we wanted to present an up-to-date series of results on todays Internet.
- To dismiss that hard limits for delay should be the defining factor for defining todays Internet telephony quality [Int93]. We agree that delay should be minimized as much as possible without sacrificing playout quality, however it should not be the sole *definition* of quality.
- The number of users and their access to the Internet, particularly through wireless connections, will grow considerably in the near future. We wanted to make measurements before this next quantum jump occurs.
- On well provisioned links we can carry packet audio without reservation of network resources. These mechanisms will admittedly enhance quality when queuing occurs at the routers. One goal of this work is to ascertain what quality is available today for the "surviving" best effort service.
- The time of day which one sends and receives audio data clearly has an influence on the quality. During busy periods the calls can be unintelligible whilst at quiet periods remarkably good, quality therefore depends upon the busy/quiet times on *both* sides of the connection. A possible use of these periods is deciding when to call for optimal quality.
- Work done by Paxson [Pax97] cites worrying figures regarding asymmetric routes. Many factors change when routes are not the same in both directions and hence peers engaged in a conversation will have different perceptions on the quality. This is a crucial issue for IP telephony applications.

2.2 Measuring Delay

The number of possible parameters one can measure in a system is large. As most traditional telephony users are sensitive to delay we included a delay estimation. We assume the end-to-end delay is equal to the sum of some small packetization delay at the sender, the propagation delay and the receiver induced delay. Artificially introduced delay at the receiver is necessary to achieve smooth playout of packets due to interarrival differences contributed by the statistical multiplexing of packet switched networks. This is a well known artifact and is explained in more detail in [RKTS94] [MKT95] [Sch92].

The packetization delay on most operating systems is relatively small approximately 20ms [SSO83]. We can approximate the propagation delay from the RTT timer in RTCP and dividing by 2. A point to note is if the route is asymmetric then this will be an approximation. The delay added by buffering can be estimated by counting the number of bytes that the packet must "wait"[2].

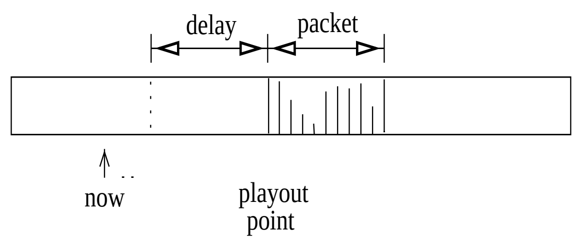

Figure 1: Delaying packets for playout

The time the packet must wait (in milliseconds) is dependent on the sample rate and the number of bits per sample:

$$\text{delay} = \frac{sampling\ rate\ .\ packetsize}{samples\ per\ sec\ .\ bits\ per\ sample}$$

for a typical 160 bytes of payload at 8000 Hz this is:

$$\text{delay} = \frac{8000\ .\ 160}{8000\ .\ 8} = 20\,ms$$

Calculating the amount to delay a packet as is shown in Figure 1 between "now" and the playout point can be determined from the variance of interarrival times of the incoming packets. The variance is simply the difference in interarrival times of two consecutive packets and is re-calculated for every packet. If the variance steps over (or below) a threshold, the buffer space can be adjusted to delay or hasten packets before handing them to the audio device. In this manner a steady number of packets are queued for replaying and hence the listener hears an uninterrupted sequence of audio samples.

[2]This is nicely explained in [Jac94].

2.3 Interarrival times

Interarrival times can be obtained by the difference in the local arrival time and a timestamp in the packet itself with the previous packets, this is continually calculated and a mean variation is obtained by multiplying the mean variation by a smoothing constant. This mean variance or *jitter* gives a good indication of the running condition of a connection This so called "jitter"is reported back to the sender as part of the RTCP protocol. Plots of interrarival times for 3 remote connections are shown in Figure 5.

2.4 Bi-directionality

Asymmetry is a problem in today's Internet. Paxson [Pax96] reports that in 1995 half of the 40,000 measurements taken visited at least one different city in each direction. Observing and verifying this artifact is non-trivial however in many runs we saw significant differences in the two directions. The likelihood this will appear is greater for connections over many hops therefore we concentrated on our longest one, Stockholm to Argentina. The propagation delay in each direction differed from 344 ms from Argentina to Stockholm to 297 ms in the reverse direction versa[3]. In Figure 9 we show the interarrival times for one trace.

2.5 Log files

We have implemented a RTP/RTCP pair which is essentially an IP Telephony tool. Under Windows we can record and play PCM encoded files, whilst UNIX we only produce statistics. Some enhancements have been made to the logging and data gathering modules, in particular to the part which measures the mean receiver waiting time as explained in Section 2.2 and is shown in Figure 6 (second column). One further addition is the ability to record and play from log files which have three roles: to record and replay sessions, to test the RTP/RTCP implementations and lastly to utilize machines for tracing on machines without any sound support.

```
I 919339994 729295
T 919339995 732069
E 919339996 38883 172 32869 0 160
E 919339996 68883 172 32869 1 320
--------------------------------
1 2         3     4   5     6 7
(Header: V=2 P=0 X=0 CC=0 M=0 PT=101)
```

Figure 2: Log File Format

One can hand edit the files to for example delete 10% of the arrivals and verify that 10 packets are reported as lost. The log files can be collected on one machine and replayed later on another. We used this technique to collect traces throughout the day and at a later time to analyze and prepare results, often on different machines.

[3] 100 probe packets over 1 hour

3 Results and Discussion

Our selected test sites consisted of one national (Luleå, Northern Sweden), one continental (Cambridge, UK), one trans-Atlantic site in the US, (Amherst, Massachusetts) and finally the "furthest" we could obtain an account for was in Buenos Aires in Argentina. SICS in Stockholm acted as the central site. For all tests a trace file was copied to the remote site so that reproduceable tests could be made. During the bidirectional tests the runs were performed simultaneously for fair comparison. Figure 3 gives details of the trace file, which is an excerpt read aloud from a C book.

Trace File	Value
File Size	584480 bytes
Duration	70 secs
Payload	160 bytes (20 ms)
With Silence Suppression.	3643 packets
W/o Silence Suppression	2064 packets
% Transmitted	56.6 %

Figure 3: Trace file properties

Figure 4 shows the mean queue delay in Stockholm for senders in Argentina and the USA. It is quite evident that the connection to the US fares much better, with an average delay of 50 milliseconds, which when added to the propagation delay gives a delay of about 110ms which is good considering it was one of the busiest times of the day. The end-to-end delay from Argentina at this instant con be estimated at about 420ms (with packetization). Figure 6 shows all 4 sites.

Figure 5 is an attempt to show the difference between the differences in the variance of packet arrivals. Luleå has a sharp peak of jitter variance, with many packets having the same "inter-packet spacing" which therefore results in less frequent changes needed to the playout buffer (see 2.2). The spread gives an indication of how large those changes need to be. The opposite is true for the connection to

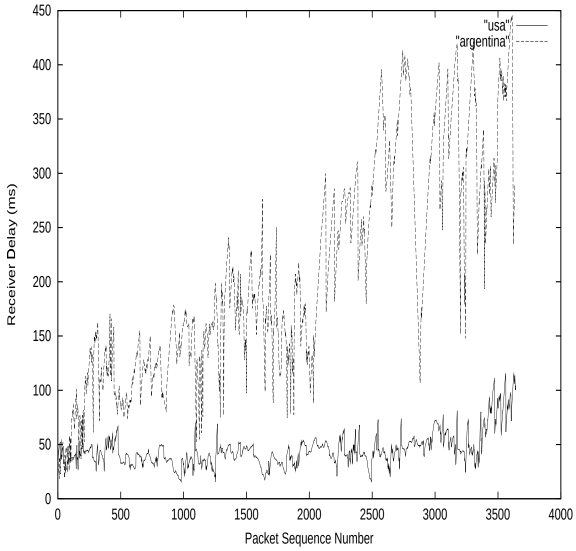

Figure 4: Receiver delay for 2 sites

Argentina where the variance ranges from 20ms to 150ms with no particular dominant packet spacing.

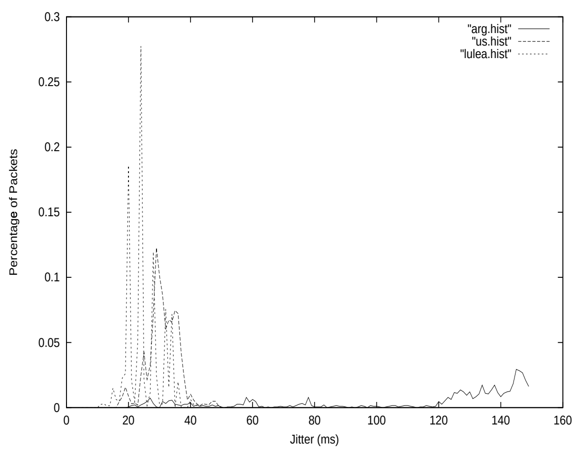

Figure 5: Interarrival histograms for 3 sites

Figure 6 shows a summary of the delays (to Stockholm) from the remote sites. Worthy of note is the tests were conducted at 17:00 Central European time which is 11:00 in Massachusetts and 12:50 In Buenos Aires, all sites and the transatlantic link is at it's busiest.

Figure 7 shows the number of hops and the jitter for each source. The jitter (to Stockholm) increases as expected with the "Internet distance" to the remote hosts.

The final two plots possibly indicate the asymmetry that was reported in [Pax96]. The tests were conducted simultaneously in both directions and one

Src	Prop. delay (ms)	Rec. (ms)	Tot (ms)
Luleå	12.13	42.14	54.27
UK	33.3	54.39	87.68
USA	57.3	50.82	108.12
Arg	273.0	119.12	392.12

Figure 6: Delay to Stockholm in ms (Mon:17:00)

Source	Hops	Jitter
Luleå	9	25.12 ms
UK	15	27.38 ms
US	15	31.25 ms
Arg	18	137.69 ms

Figure 7: Mean Jitter Values 17:00

can see there is a clear difference in both packet loss (from RTCP) and variance in jitter. In Figure 8 the receiver in Stockholm is experiencing much worse packet loss (90 compared to 10). A similar situation can be seen with the jitter (Figure 9) which shows much worse variation for the packets en route to Stockholm. Unfortunately, we were not able to find the mapping between a traceroute run and the location of the routers to verify Paxsons' numbers, but suspect this is a case.

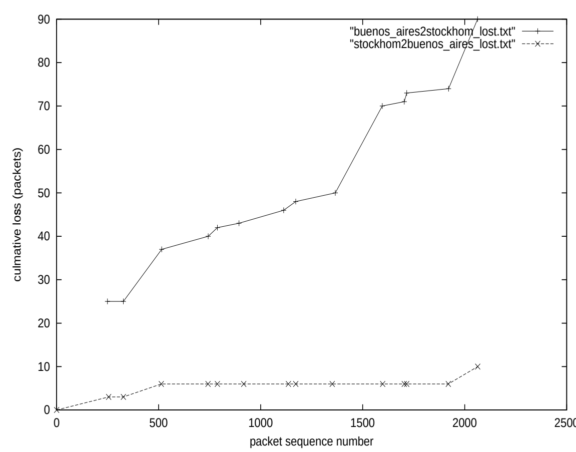

Figure 8: Bidirectional Loss (Cumulative)

4 Conclusions/Future Work

In this short paper we have presented some criteria for investigating the quality of Internet Telephony connections. We have shown through estimates of end-to-end delay and interarrival histograms that the

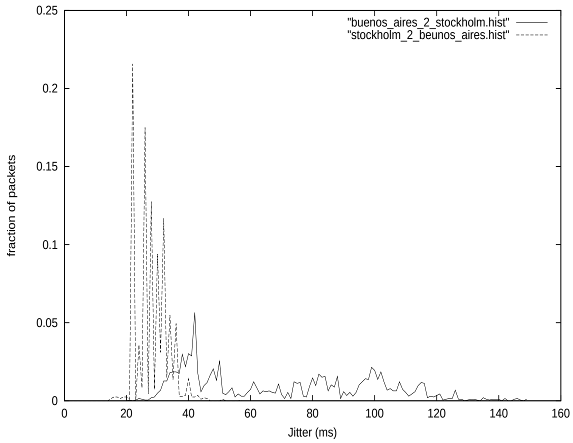

Figure 9: Bidirectional Interarrival Histograms

Internet is capable of carrying voice with acceptable delay and quality. We tried to select remote sites that would not always give good results and the busier parts of the day to conduct the tests. However there is some concern about the possible difficulties in future regarding asymmetric paths.

We have not dwelled on packet loss in too great detail as this is covered in other work, rather we have shown that Internet telephony (with good to reasonable quality) is feasible for most users connected to the fixed Internet. Obviously more connections need to be tested but this is now very simple with the framework we have established.

We have developed a tool that probes, gathers and produces statistics based on the RTP and RTCP protocols. Additionally we have included time driven tracing as well as looked at bi-directional quality. The tool is compatible with RTP/RTCP implementations therefore it can be used in conjunction with other packages.

We plan to extend the tool to operate as a daemon in which the tests can run without intervention. This will allow any site to connect to another allowing a full mesh topology rather than the centralized system. We plan to make the central site an RTP/RTCP mixer as defined in the RTP draft, thus allowing aggregation of flows and more testing possibilities. Additional encoding will also be introduced and currently we have only drivers for the Windows environment, UNIX (VoxWare based) ones will be forthcoming. Finally we plan to look at asymmetric Internet access means such as cable modems.

Due to the brevity of this paper it is not possible to show all plots and results that we have obtained in designing and testing this tool. Also discussions have been kept deliberately terse for space reasons. However more details will appear at http://www.sics.se/~ianm/Voip/voip.html in the near future. We would like to thank Thiemo Voigt for his careful reading and Bengt Ahlgren for his timely comments on this paper, Ericsson and Telia for their financial support. Also thanks to Steve Pink, Jim Kurose and Pablo Giambiagi for the accounts that we have used.

References

[BCG95] J. Bolot, H. Crepin, and A. Garcia. Analysis of audio packet loss in the internet. Lecture Notes in Computer Science, pages 163–174, Durham, New Hampshire, April 1995. Springer.

[Int93] International Telecommunication Union (ITU). Transmission systems and media, general recommendation on the transmission quality for an entire international telephone connection; one-way transmission time. Recommendation G.114, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, March 1993.

[Jac94] Van Jacobson. Multimedia conferencing on the Internet. London, England, August 1994. Tutorial slides.

[MKT95] Sue B. Moon, Jim Kurose, and Don Towsley. Packet audio playout delay adjustment algorithms: performance bounds and algorithms. Research report, Department of Computer Science, University of Massachusetts at Amherst, Amherst, Massachusetts, August 1995.

[ML97] N. F. Maxemchuk and S. Lo. Measurement and interpretation of voice traffic on the internet. Montreal, Canada, June 1997.

[Pax96] Vern Paxson. End-to-end routing behavior in the internet. Stanford, California, August 1996.

[Pax97] Vern Paxson. End-to-end internet packet dynamics. Cannes, France, September 1997.

[RKTS94] Ramachandran Ramjee, Jim Kurose, Don Towsley, and Henning Schulzrinne. Adaptive playout mechanisms for packetized audio applications in wide-area networks. pages 680–688, Toronto, Canada, June 1994. IEEE Computer Society Press, Los Alamitos, California.

[SCFJ96] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. RTP: a transport protocol for real-time applications. Request for Comments (Proposed Standard) 1889, Internet Engineering Task Force, January 1996.

[Sch92] Henning Schulzrinne. Voice communication across the Internet: A network voice terminal.

Technical Report TR 92-50, Dept. of Computer Science, University of Massachusetts, Amherst, Massachusetts, July 1992.

[SSO83] Daniel C. Swinehart, L. C. Stewart, and S. M. Ornstein. Adding voice to an office computer network. pages 392–398 (11.4), San Diego, California, November 1983. IEEE. also Xerox Report CSL-83-8.