Is the Internet ready for VoIP?

Fouad A. Tobagi, Athina P. Markopoulou, Mansour J.Karam

Email: *{tobagi, amarko}@stanford.edu, mans@routescience.com*

ABSTRACT

There is a great interest today in voice communication over the Internet (VoIP). If the Internet were to become the universal network for all communications needs, and thus were to displace the telephone network for voice communication, it must be capable of providing the same level of service quality as the telephone network. Today, this does not seem to be the case. VoIP is plagued with packet loss and variable packet delay in the network. Although measures are taken to overcome these problems (e.g., loss concealment and adaptive playout), their effectiveness is limited to certain ranges of network conditions. Thus it is of great importance to understand the packet loss and delay characteristics of today's Internet, in order to understand the effectiveness of the measures introduced to overcome the impairments. To that end, we examine in detail measurements of packet loss and delay taken over the Internet, and give a characterization thereof. We then comment on the impact that they can have on the quality of VoIP, and on the effectiveness of the measures introduced to minimize their impact. This study is limited to measurements taken only on Internet backbone networks, which represent important components in long distance communication. It reveals that packet loss and delay characteristics are not consistent across all backbone networks. Some backbone networks exhibit fairly good characteristics and may offer good quality communication, leading to a confirmation that packet voice is a sound approach. Other backbone networks exhibit undesirable characteristics that could not be accommodated with any of the measures introduced today. We comment on the possible causes and on the improvements that need to be made in the Internet backbone networks to render them adequate for VoIP.

1 INTRODUCTION

Many among us have used the Internet for voice communications. We all like it because it is free. However, many dislike it because it provides poor speech quality and limited interactivity between the communicating parties. Furthermore, the quality of voice communication is not uniform across all calls. So it is clear that the quality of VoIP service over the Internet is not as good as what we are accustomed to with POTS, and often does not even come close. Then the following questions become particularly interesting. What needs to be done for VoIP to achieve the high quality of POTS? Is the Internet too unreliable to offer a dependable VoIP service? Is the mix of traffic in the Internet so unpredictable and bursty that the delay incurred by voice packets is adversely affected? In that case, would differentiated service solve the problem?

Simply stated, the only possible problems that occur in the Internet and that can affect the quality of voice communication are packet loss and packet delay. Loss may be due to congestion in the network leading to packets getting dropped in switches and routers, or failure of network components (links, switches and routers) leading to a reconfiguration of the network. Here the issue is how extensive is the loss and how bad is its effect. Can its effect be concealed at the destination? What measure can one take, if any, in order to make this possible?

As far as delay is concerned, we distinguish the fixed part of the delay from the variable part. The fixed part comprises packet transmission time over, and propagation time across the links in the path, and any fixed transit delay that may be incurred through network elements encountered in the path. The variable part of the delay comprises queuing delays incurred within network elements and other possible delays introduced by the operation of the network elements (e.g., router) [1].

Even if there were no variations in the end-to-end delay, the magnitude of the latter is important because of its effect on interactivity. To achieve a good level of interactivity, the end-to-end delay should be maintained below a certain maximum, certainly not to exceed 150 ms. For conversations with more stringent interactivity requirements, (that is when the turnaround time is shorter, such as for example when two people take turns reading names or numbers), there is a benefit in having even shorter end-to-end delay. Another effect of large end-to-end delays is the annoyance caused by echoes when no echo cancellation is present in the system. Unfortunately, there is little that one can do about the fixed part of the delay; especially that there is little control that one in general has about the routes packets may take to reach their destinations.

Delay variations (also referred to as delay jitter), on the other hand, may be dealt with at the receiver. A playout buffer is introduced in which packets may be delayed so as to achieve a smooth playback of the speech. The scheduling of packet playback may be of the fixed type, whereby a constant end-to-end delay target is enforced on all packets. Packets that exceed the target delay are dropped. Alternatively, the scheduling of packet playback may be of the adaptive type, whereby the target delay is allowed to vary over time. In one scheme, the target delay is allowed to vary from one talkspurt to another, based on delay measurements made during a talkspurt; within a talkspurt, all packets experience a constant delay, [2, 3]. In another scheme, the scheduling allows the target delay to vary from packet to packet within a talkspurt, thus allowing the rate at which the speech is played back within a talkspurt to also vary, [4].

While the measures proposed to mitigate the effect of packet loss and packet delay jitter are sound, their effectiveness depends on the characteristics of the loss and delay that are experienced in the network; such characteristics include the pattern of packet loss, the magnitude of delay variations and the rate at which these variations take place. Intensive delay measurements are crucial in shedding light to this matter.

Measuring loss and delay in the Internet is not a simple task. The Internet is a large system that has a hierarchical structure. At the bottom level in this structure, we identify residential access networks and campus networks that connect residential users and corporate users to regional networks; the latter provide connectivity within "regions" such as large metropolitan areas, and connect the users within the regions to the rest of the Internet; finally, at the top level of the hierarchy are the wide area backbone networks that provide global connectivity. The characteristics of end-to-end packet loss and delay for a given path are a combination of such characteristics of individual networks at the various levels of the hierarchy, characteristics that may vary considerably between levels. For example, wide area backbone networks are generally well provisioned and thus do not exhibit congestion episodes, while regional networks that handle a much higher degree of variability in traffic may exhibit congestion and thus packet loss. Accordingly, to study the characteristics of the Internet and derive any conclusions as to its performance with regards to VoIP, it is more appropriate and fruitful to focus on one level of the hierarchy at a time.

Even when focusing on a single level in the hierarchy, we find that the Internet comprises many separate domains, each administered by a different organization. Each such organization is responsible for the deployment and operations of the networks within that domain. These networks can differ considerably in their provisioning and operations, and as a result, their performance. Therefore to get a realistic assessment of their loss and delay characteristics, it is important to study a good sample of these networks. This is also important given that, in general, Internet users (including VoIP service providers) do not have control over the routes taken by packets. Packets transmitted between two hosts may not take the same route in both directions, and these routes may fall in different domains.

In this paper the focus is on wide area backbone networks. This choice is primarily driven by our access to extensive measurement data collected for such networks by RouteScience Technologies, Inc. There are other reasons why a focus on wide area backbone networks is of interest. These networks are an important part of the end-to-end path for all long distance VoIP calls, including calls that are serviced by a combination of a switched telephone network in the local area and the Internet for the long haul. Performance problems in these networks will be experienced by all such calls; therefore, they need to be well understood and fixed, regardless of what takes place elsewhere in the path.

The outline of the paper is as follows. In Section 2 we describe the measurements available and provide loss and delay characteristics for a number of representative paths of interest. In Section 3, we examine the effects that these characteristics have on VoIP and comment on the effectiveness of the measures usually taken to mitigate these effects. In Section 4 we conclude with some remarks.

2 MEASUREMENTS OF INTERNET BACKBONE NETWORKS

2.1 Measurement Set

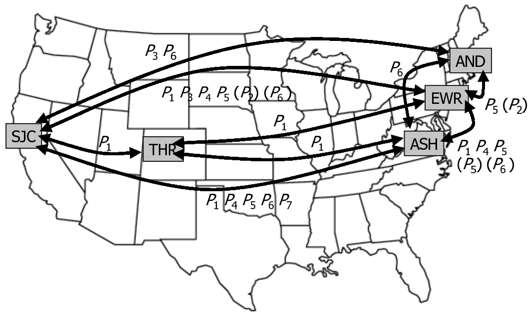


Figure 1: Measurements collection

Since encoded voice is of a constant bit rate, a packetized voice stream consists of a succession of equal size packets equally spaced. Since the packet formation time contributes to the end-to-end delay, the speech data included in a packet corresponds to 10 to 30 ms of speech. Thus a voice source generates packets, one every x ms, where x may take a value in the range of 10 to 30.

RouteScience Technologies, Inc. made extensive loss and delay measurements on Internet backbone networks during the periods of December 1-14, 2000 and June 26-29, 2001. Measurement facilities that are capable of sending and receiving probes and collecting the delays incurred by the probes were developed. These facilities were equipped with GPS receivers enabling them to timestamp the probes and derive end-to-end delays with an accuracy of microseconds. Probes were generated and sent continuously (one probe every 100ms for the first set, and one probe every 10 ms for the second set), 24 hours a day.

The measurement facilities were connected to various Points of Presence (POPs) of several Internet Backbone providers by means of T1 and T3 links. Seven providers and five cities in the US were considered. The 7 providers are referred to in this paper as P1, P2 ... P7 for anonymity purposes. The five cities comprised the following: one on the West Coast - namely, San Jose (SJ) in California; one in Colorado - namely, Thornton (THR); and three cities on the East Coast - namely, Newark (EWR) in New Jersey, Ashburn (ASH) in Virginia, and Andover (AND) in Massachusetts. In Figure 1 we show the paths for which measurements have been collected along with the providers, totaling 43 paths altogether.

Of interest to us in this paper are the measurements collected during the June 26-29, 2001 period because of the smaller 10 ms intervals used, necessary to emulate voice traffic. At this rate, the probes (which are 50 bytes long) constituted a data rate of only 40Kb/s. For the data rates of links used in backbone networks, this rate represents a small fraction. Therefore, the load generated by the probes could not have any effect on the delay and loss characteristics of these networks.

In this section, we study the measurement data collected in June 2001. The measurement period started on Tuesday June 26 at 19:22:00 and ended on Friday June 29 at 00:50:00. (Here time is according to Coordinated Universal Time (UTC) which corresponds to Greenwich Mean Time - GMT). Thus we have measurements for a continuous period covering a little over two full days.

2.2 Measured Packet Loss Characteristics

There were three paths (namely, SJC-AND for provider P3 and SJC -ASH for providers P5 and P6) where no loss was experienced during the entire measurement period. For all other paths, there are probe loss events that occur. These loss events can have different characteristics pertaining to the pattern of packet loss during the events. For the purpose of accurate description of packet loss characteristics, we identify two types of events: elementary packet loss events which consist of consecutive probes getting lost (comprising one or more packets) separated by relatively long periods of time, and complex loss events which correspond to the occurrence of several elementary probe loss events concentrated over a short period of time.

For several providers (namely, P1, P2, P4, P6 and P7), we note that the number of loss events during the two full days is rather small, on the order of 10's for the entire measurement period. These are mostly of the elementary type, but they do comprise some complex events. As a concrete example, we consider the path ASH-SJC of provider P6. This path incurred 12 elementary events, of which 6 consist of a single packet lost, and 6 consist of 19-22 consecutive packets lost. This path also incurred a single complex event that lasted 15 s during which packet loss occurred in the form of single packets leading to a loss 9.4%. We show the occurrence of this complex event in Figure 2 in which we plot the delay incurred by probes as a function of the probe's send time; for probes that are lost, we show a delay of zero. In Figure 3, we show a blow up of a portion of the graph shown in Figure 2.

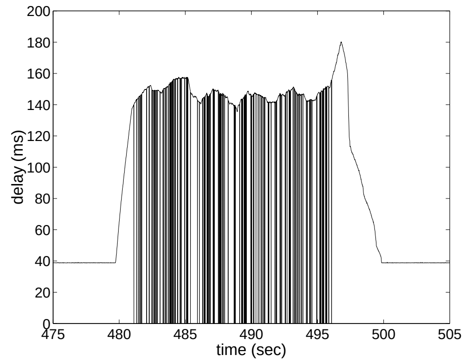


Figure 2: A complex loss event on path ASH-P6-P1, on Wed 06/27/01 at 3:30 (UTC)

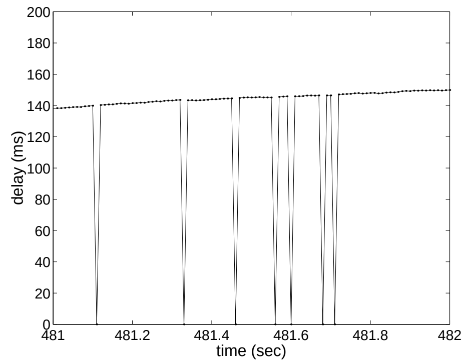


Figure 3: Zooming in on the complex loss event on path ASH-P6-SJC, on Wed 06/27/01 at 3:30 (UTC)

As another concrete example, we consider EWR-SJC of provider P2. This path incurred 27 elementary events, among which 20 consisted of single packets lost, and 7 consisted of 17-24 packets lost. This path also incurred 5 complex events lasting from around 20 s to around 60 s with loss rates ranging from 19% to 42%. We choose two complex events in this path to illustrate the possible loss patterns incurred, which we plot in Figures 4 and 5. The event shown in Figure 4 comprises a mixture of single packet loss events spanning a period of 30 s sandwiched between two elementary multi-packet loss events lasting 5 s each. The entire duration of the complex loss event is 50 s, and the packet loss rate during that period is 24.6%. The event shown in Figure 5 comprises a number of elementary multi-packet loss events each lasting up to 1.4 s, spanning a total duration of 30 s and leading to a packet loss rate of 41.4%. It is interesting to note that this loss event is synchronized with another loss event incurred on another path of the same provider, (EWR-ASH) with exactly the same characteristics.

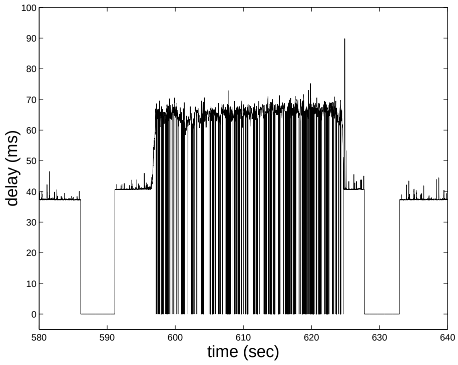


Figure 4: A complex loss event on path EWR-P2-SJC, on Wed 06/27/01 at 3:30-3:50 (UTC)

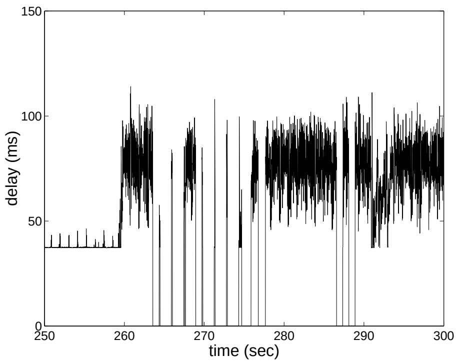


Figure 5: A complex loss event on path EWR-P2-SJC, on Thu 06/28/01 at 20:10 (UTC)

As yet another concrete example, we consider the paths ASH-SJC and SJC-ASH of provider P7, (the only paths for this provider,) each of which has incurred a single elementary loss event once every 24 hours, synchronized with each other. Such events lasted approximately 2 minutes and were accompanied by a change in the fixed part of the end-to-end delay. This change in delay is indicative of a change in the route taken by the paths. An instance of these events is displayed in Figure 6. There are no other loss events on these paths except for three single packet loss events on the path SJC-ASH.

For provider P3, elementary loss events consisting mostly of individual packet loss occurred regularly, spaced

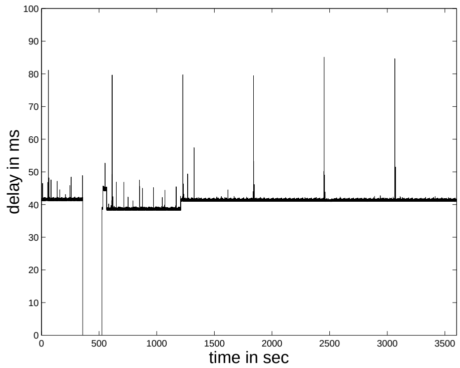


Figure 6: A complex loss event on path SJC -P7-ASH, on Wed 06/27/01 at 4:00 (UTC)

by an interval averaging 4 s. leading to a packet loss rate of 0.25%.

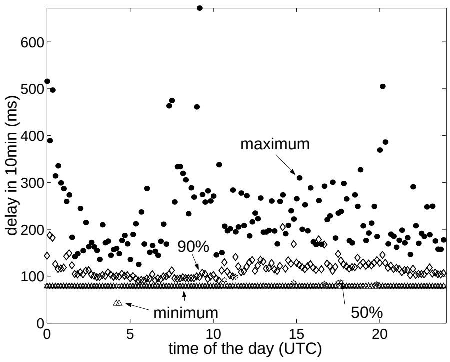
For provider P5, there are some paths (e.g., ASH-SJC) for which the number of loss events was in the 100's during the 48-hour period, a large fraction of which is concentrated over a period of 8 hours of the day in the morning.

2.3 Measured Delay and Delay Jitter Characteristics

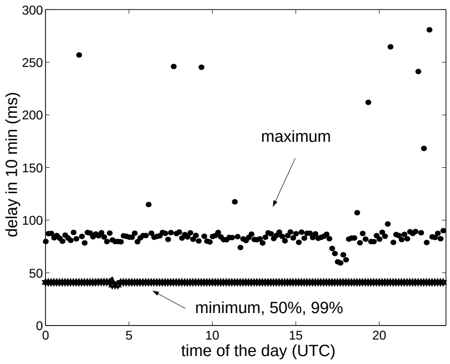
To aid in the analysis of delay for such a large set of measurement data, we begin by examining the statistics of delays incurred by probes over 10 minute intervals. We record for each such interval the minimum and maximum delays, and various delay percentiles (primarily the 50th and 90th percentiles). We then plot these for all 10 minute intervals for a 24 hour period. We show in Figure 7 such a plot for a number of paths; namely, the path THR-ASH of provider P1and the paths SJC-ASH of providers P7 and P2.

The minimum delay observed corresponds to the fixed portion of the end-to-end delay; it usually remains constant across time. There are cases observed when the fixed delay has changed as a result of a route change. The minimum delay is typically below 10 ms for paths joining cities on the same coast (EWR, ASH and AND on the East Coast), and in the range of 30 to 45 ms for paths joining cities across the US (SJC on the West Coast and the three cities on the East Coast). One exception to the above is the path THR-P1-ASH that has a minimum delay of 78 ms.

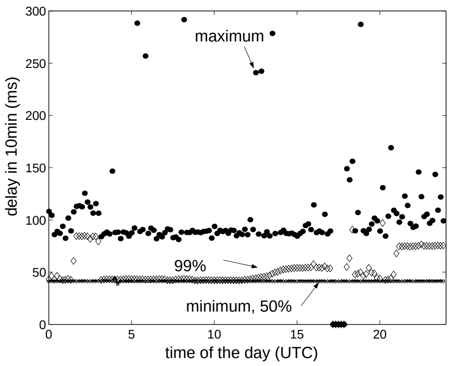
The maximum delay and delay percentiles are important to identify intervals during which probes have experienced delays that are large compared to the minimum. If in a one 10-minute interval we observe a high maximum accompanied by increased values of the percentiles, then it means that the interval is of interest for further study. The delay statistics exhibited in Figure 7 are also useful to give an indication of the effect of time of day on measured delay. It also aids us in comparing paths; for example, from Figure 7 we see that the path THR-ASH of P1 is a path



(a) Path THR-P1-ASH on Wed 06/27/01 (UTC)



(b) Path SJC-P7-ASH on Wed 06/27/01 (UTC)



(c) Path SJC-P2-ASH on Thu 06/28/1 (UTC)

Figure 7: Delay percentiles per 10 minute intervals for a 24h period

that exhibits high peaks as well as high percentiles most of the day, while at the other extreme the path SJC-ASH of P7 is a path that exhibits rather low delays. The path SJC-ASH of P2 is a path that is usually good (similar to P7) for most of the day, but does incur higher delays over a certain period of the day.

In this paper, we are primarily interested in analyzing the delay variations experienced by probes in order to identify the various delay jitter patterns that may be found and characterize them. This requires that we plot the delay of individual probes versus their respective send times. A typical example is shown in Figure 8. The delay variations that we see show that the delay is constantly varying within a certain relatively small range above the minimum. There are frequent visits to the minimum, indicative of the fact that during the periods displayed the corresponding paths are lightly loaded. This type of delay variation prevails and corresponds to what one might call the normal or regular patterns.

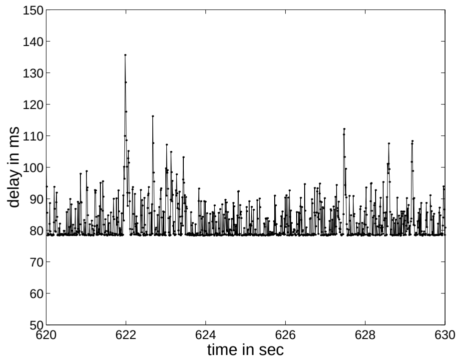
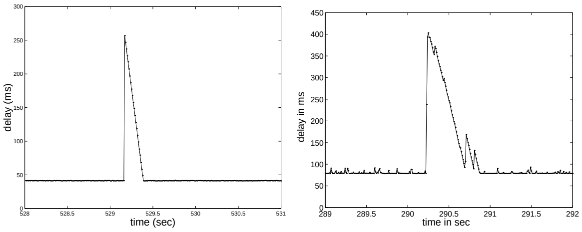


Figure 8: Delay of individual probes on path THR-P1-ASH, on Wed 06/27/01 at 2:10 (UTC)

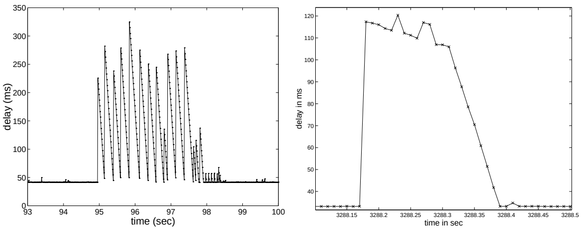
There are however larger delay variations that occur. Examples are shown in Figure 9. These delay variations occur in the form of spikes, whereby a spike consists of a sudden sizeable jump in delay for a probe, followed by a succession of probes delays decreasing by 10 ms each. We note that since probes are sent deterministically one every 10 ms, the delays of probes succeeding the peak follow a line with a slope of -1. (See Figure 9-a). The only parameter characterizing such a spike is the magnitude of the jump, or equivalently the peak delay. The rest simply follows. The spike shown in Figure 9-b is not as simple as that of Figure 9-a; there are several smaller peaks that follow the first and tallest peak. In this case, the entire event may be characterized by the magnitude of the first (highest) peak and the width of the spike; i.e., the number of probes involved in the spike.

There are yet other situations that differ from the above description. An example is shown in Figure 9-c. It consists of a rapid succession of spikes lasting over seven seconds. Another example is shown in Figure 9-d. In this case, following the sudden jump in delay, a number of probes incur roughly the same delay as the peak, before the linear decrease in delay is observed. This is an exception to the triangular spike shape that happens on provider P5 for a 5



(a) "Model" spike, on path SJC-P7-ASH, on Wed 06/27/01 at 2:00 (UTC)

(b) High spike, on path THR-P1-ASH on Wed 06/27/01 at 0:00 (UTC)



(c) Clustered spikes, on path EWR-P4-SJC, on wed 06/27/01 at 21:00 (UTC)

(d) "Spike", on path SJC-P5-EWR, on Thu 06/28/01, at 17:00 (UTC)

Figure 9: Example spikes

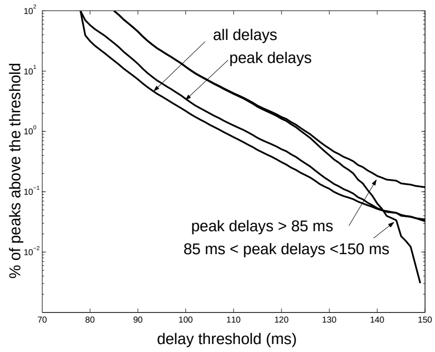
hours period. However, the large majority of spikes in the traces follow the triangular shape of 9-a.

The characteristics of spikes and the specific pattern of occurrence vary from path to path and over time. We illustrate this fact by examining paths from the three providers discussed above: P1, P7 and P2. We are guided by the delay statistics generated for 10 minutes intervals and shown in Figure 7 above to select periods of time on which to study in greater detail.

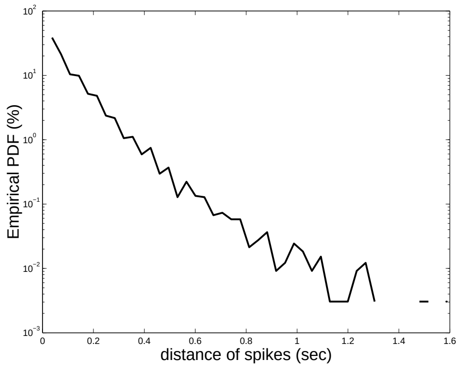
2.3.1 *Discussion of Path from P1*

The path from THR to ASH belonging to provider P1, see Figure 7-a, is a highly loaded path that exhibits high delay variations: high peaks and periods of time during which the spikes occur at high frequency.

Most of the time, delay is low (roughly below 150 ms) and follows a random pattern, consisting of spikes with random peaks that happen at random intervals. For example, the pattern shown in Figure 8 corresponds to such a regular period (from 2:00 - 3:00 UTC on Wed 06/27/01). We consider peak delays of a considerable size to be those above 85 ms and we observe that their distribution follows an exponential shape (with a mean of 92 ms). Figure 10-a shows the complementary cumulative distribution function (CCDF) for all probe delays and for the peak delays in particular. It is interesting to note that the distribution of all probe delays is very close to the distribution of the peak delays, which can be justified by the triangle shape of the spikes. The period of time separating these spikes also follows approximately an exponential distribution, as shown in Figure 10-b. The same observations hold for most of the day, when delays are small, i.e. roughly below 150 ms. The truncated distribution for peak delays below 150 ms is also shown in Figure10. However, we observed that larger delays typically follow regular patterns that we now discuss in detail.



(a) Complementary Cumulative Distribution Function for delay

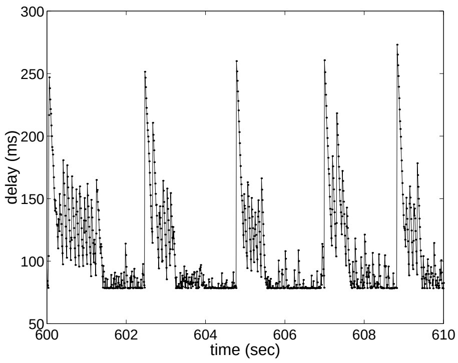


(b) Complementary Cumulative Distribution Function for the distance between consecutive spikes

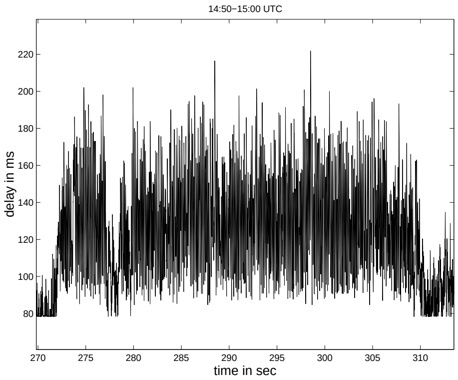
Figure 10: Magnitude and frequency of spikes on the path THR-P1-ASH, during the period: 2:00-3:00 (UTC) on Wed 06/27/01.

There are three distinct patterns observed on provider P1 that occur when delays are high, i.e. above 150 ms. The first one is shown in Figure 9-b: one high spike followed by a few smaller ones. These are the highest peaks observed (as high as 400 ms - 700 ms) and they happen every 10-20 ms. From Figure 7 we see that they happen during the periods 0:00-1:00, 6:00-10:00 and 20:00-21:00, 23:00-00:00.

The second regular pattern is shown in Figure 11-a. It consists of a spike with a peak at 250 ms followed by smaller oscillations; this shape is repeated every 1.5-2 seconds. This pattern occurs 9 times in the entire measurement period and it leads to an increase to the median delay.



(a) A regular pattern



(b) Sustained increase in delay

Figure 11: Additional delay patterns of provider P1

The third pattern is the one shown in Figure 11-b. The spikes are more frequent and there is a sustained increase in the delay range lasting for tens of seconds. This pattern also leads to an increase in the median delay, see Figure 7 and it happens 18 times during the entire measurement period. It also happens often on providers P2 and P5.

2.3.2 Discussion of Path from P7

The path from SJC to ASH on P7 is a path in a very well provisioned network that exhibits very low delay variations. Delay lies in a narrow range between 40.5 and 42 ms. However, we observe spikes as shown in Figure 9-a that occur periodically every 10 minutes with peaks at 80-90 ms and occasionally 250-300 ms. Around the loss event shown in Figure 6, there are some additional small spikes. This behavior leads to the delay percentiles shown in Figure 7-b. The same observations hold for the only other path of this provider.

2.3.3 Discussion of Path from P2

The path from EWR to SJC on provider P2 exhibits mixed delay characteristics. For most of the day, there is low delay variability similar to P7. Delay is between 37 and 45-50 ms, due to clusters of two spikes, as high as 45-50 ms spaced 1 sec apart. Every 10 minutes there are some higher spikes of magnitude 90-100 ms. This results in the low 99th percentile and the higher maximum observed in Figure 7-c. However, between 0:00 and 2:00 as well as between 14:00 and 20:00, the pattern of delay variation changes; in addition to the regular pattern, there are spikes at least 100 ms high occurring every 1 sec. This results to an increase of the 99th percentile in Figure 7-c during those periods. Figure 12 shows delays for the period from 1:00 and 2:00, during which there is change from the lower delay to the higher delay pattern. Most of the loss events for provider P2, discussed in Section 2.2, coincided with such changes.

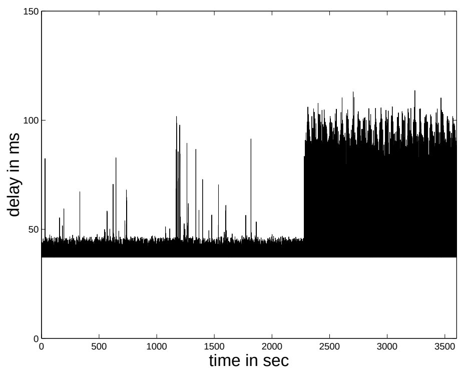


Figure 12: Delay on the path EWR-P2-SJC, from 1:00 until 2:00 (UTC) on Thu 06/28/01.

3 Effects of Measured Loss and Delay Characteristics on the Quality of Voice Communication over the Internet

In this section, we concern ourselves with the effects that measured loss and delay have on the quality of voice communication over the Internet (backbone networks), and whether the measures we have at hand can remedy these effects. We begin the section with a short summary of the characteristics of voice communication and the kinds of impairments that are incurred as a result of loss and delay. We then provide a quick review of the measures that have been proposed to mitigate these impairments and comment on their effectiveness in light of the measurements discussed in Section 2.

3.1 Voice Communications Characteristics and Possible Impairments

The quality of voice communication in the presence of impairments is assessed by a measure referred to as Mean Opinion Score (MOS) that reflects the subjective rating given by listeners. It is a quantitative measure given on a scale of 1 to 5. The meaning given to ranges of values of MOS is provided in Figure 13.

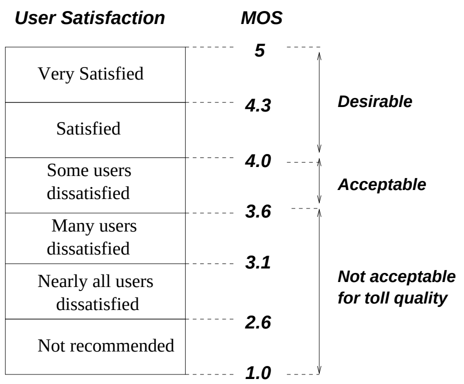


Figure 13: Mean Opinion Score and its relation to voice quality levels and user satisfaction.

There are several factors that affect the quality of speech: the encoding process, loss of speech, echo and the total ("mouth-to-ear") delay. The encoding process at the source introduces degradation. The MOS after encoding and without any other impairment is given for various encoding schemes in Table 1.

Table 1: Standard encoders and their characteristics

Standard	Codec type	Rate (Kbps)	Frame (ms)	MOS intr.
G.711	PCM	64		4.43
G.729	CS-ACELP	8	10	4.18
G.723.1	ACELP	5.3	30	3.83
G.723.1	MP-MLQ	6.3	30	4.00

The effect of loss of speech has been studied extensively; for a survey of such studies can be found in [5]. We show, for example, in Figure a summary of the degradation due to speech loss for G.711 by plotting the MOS attained as a function of the fraction of packet loss. We note that the degradation due to loss depends on the duration of clipped speech: the longer the clipped speech durations, the worse the degradation. Furthermore, the quality degradation is very high if loss concealment is not used. However, loss concealment has its limit. For example, a study of loss concealment in G.723.1 has indicated that its effectiveness decreases rapidly with the duration of clipped speech, [6]. It is shown that, for a given packet loss rate, loss concealment is quite effective when clipped speech is equal to 30 ms (a single frame in G.723), less effective when clipped speech is equal to 60 ms (2 frames), and hardly effective when clipped speech is equal to 120 ms (4 frames). This is explained by the fact that when clipped speech starts exceeding 60 ms, it starts affecting intelligibility, since durations of speech of 60 ms and higher cover phonems.

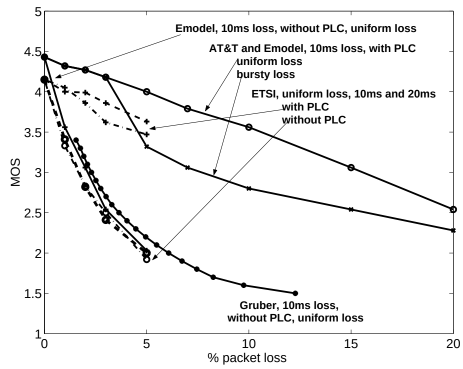


Figure 14: G.711 quality under various packet loss conditions.

The presence of echo in speech represents a major source of quality degradation. Indeed, echo is not perceptible only if the end-to-end delay is very short (below 10 ms), and the longer the delay is the more annoying its effect becomes.

Long mouth-to-ear (m2e) delays affect the interactivity between communicating parties. The m2e delay that can be tolerated (and thus is not considered to introduce a degradation in MOS) depends heavily on the type of task undertaken; the latter is characterized by the frequency at which the communicating parties alternate. On one extreme of the spectrum, the task consists of two people taking turns reading random numbers as quickly as possible; on the other extreme is relaxed free conversation. The E-model states that on average interactivity is not adversely affected if the m2e delay is 150 ms or lower, [7]. For the most interactive task, m2e delay of 150 ms introduces a decrease in MOS of about 0.5.

3.2 Comments on the Effects of Measured Loss and Delay in Backbone Networks

We address first the effect of packet loss. We note from the above summary that elementary packet loss events that encompass a few packets, say up to 2 or 3 consecutive packets lead to a mere increase in background noise as long as the percentage of speech loss remains relatively low. For example, we find in this category paths belonging to provider P3 in which packet loss involve individual packets and the total loss rate is 0.25%. Furthermore, for such events where the number of consecutive packets lost is 5 or less, loss concealment techniques perform adequately in mitigating the effect of packet loss. Elementary events that span larger number of packets, such as those we identified that span about 20 packets cannot be concealed, and may cause loss of intelligibility. Also, nothing can be done about longer periods of consecutive packet loss (such as those lasting seconds or minutes). The only remedy in these cases is to improve the reliability of the network and decrease the network reconfiguration time when failures occur.

Given the magnitude and frequency of spikes that have been observed, it is clear that delay jitter in Internet backbone networks poses a serious challenge. As stated in the introduction, a principal way to overcome delay variations is by means of buffering and play-out scheduling. The question then is: can the proposed techniques cope with the delay variations observed? Consider for example fixed playout scheduling. Either the target end-to-end delay has to be fairly large to accommodate the high spikes, or significant speech clipping is to occur, leading to poor speech quality, especially when such clipping is to occur every few seconds. The problem then becomes: how to determine the magnitude and frequency of spikes in order to make the appropriate choice? There is so much variation in delay that an a priori characterization of paths in terms of delay jitter seems impossible (except for the well behaving paths). One has to consider schemes whereby learning about the path's characteristics takes place as the call progresses, aiding in adjusting the fixed scheduling. The learning in question should consist of identifying the delay spikes when they occur and identifying their magnitude and frequency. A conservative response to the learning would consist of adjusting upward the playout delay to always guarantee good speech quality in the remainder of the call, regardless of the effect that this may have on interactivity. A less conservative response would consist of adjusting upward the target delay only if the frequency of occurrence of spikes is above a certain threshold. Yet a third approach is to provide the user the ability to express his or her preference.

It is not clear that the adaptive scheduling schemes proposed in [2, 3, 4] would perform well with the delay variations observed in the measurements. Indeed, schemes that follow closely the delay profile, ignore the fact that spikes repeat in time. These schemes work well only if the variations in delay take place at a much slower rate than is the case in reality. This was shown to be indeed the case in [5].

Perhaps the best approach to finding a solution to the problem is simply to prevent large magnitude spikes from occurring. We have no clear explanation for their occurrence. So the challenge is to find out the causes for such large variations and address them seriously.

4 Conclusion

In this paper, we have studied loss and delay measurements collected over the backbone networks of major ISPs in the US. We also discussed how these characteristics affect voice quality and to what extent existing techniques are able to cope with them. We found that, although packet voice is in general feasible, many of the measured backbones are not ready to support VoIP today, due to delay variability (in the form of spikes) and loss events. The causes of this behavior seem related to network reconfiguration, router internal operations and protocol exchanges, and not due to congestion. The causes need further investigation and the problems need to be fixed before the Internet becomes ready to replace the telephone network.

References

[1] K. Papagiannaki, S. Moon, C. Fraleigh, P. Thiran, F. Tobagi, C. Diot, " Analysis of Measured Single-Hop Delay from an Operational Backbone Network", *in the Proc. of Infocom 2002.*

[2] R. Ramjee, J. Kurose, D. Towsley, "Adaptive playout mechanisms for packetized audio applications in wide-area networks", *in the Proc. of Infocom 1994.*

[3] D. Moon, J. Kurose, D. Towsley, "Packet audio playout delay adjustment: performance bounds and algorithms", *ACM/Springer Multimedia Systems*, vol.6, pp.17-28, January 1998.

[4] Y. Liang, N.Farber, B. Girod, "Adaptive Playout Scheduling and Loss Concealment for Voice Communication over IP Networks " , *IEEE Transactions on Multimedia*, April 2001.

[5] A. Markopoulou, F. Tobagi, M. Karam, "Assessment of VoIP Quality over Internet Backbones", *in the Proc. of Infocom 2002.*

[6] S. Voran, "Speech Quality of G.723.1 coding with added temporal discontinuity impairments", *in the Proc. of ICASSP* 2001.

[7] *ITU-T Recommendation G.107*, "The Emodel, a computational model for use in transmission planning", December 1998.