# Measurement and Interpretation of Voice Traffic on the Internet

*N. F. Maxemchuk*
AT&T Labs - Research

*S. Lo*
Lucent Tech. - Bell labs

*ABSTRACT*

The Internet is being used to carry voice conversations. Sometimes the quality is acceptable, and at other times the voice is unintelligible. In this paper, we describe the results of a set of measurements on intrastate, cross country, and international Internet connections.

The measurements are used to determine the relationships between the delay that is inserted to compensate for the variation in network delay, the strategy that is used to restore lost packets, and the "quality" of the voice connection. The "quality" of a connection is defined as the fraction of the time that a channel is free of distortion for intervals that are long enough to transmit active speech segments. We also examine the sensitivity of our quality measure to modifying the required length of the distortion free intervals.

We have found that, with the proper combination of packet restoration, receiver delay and patience on the part of users, connections can be established that provide reasonably good voice connections, even during the busy period on the worse channel. Patience is required because a certain fraction of the connections are very bad. In which case the best strategy is to try again later.

It is noted that the quality of international connections is not as good as cross country connections, which is not as good as the quality of intrastate connections. This indicates that Internet voice is better suited for local bypass than its current role as a replacement for long distance connections.

## 1. Introduction

At present, the primary reason for sending voice on the Internet is price. The principle problem is that the service is unreliable. Sometimes the connections are usable and at other times the speech is unintelligible. Our objective is to measure and understand the packet loss and delay characteristics of the Internet as they affect its ability to carry interactive voice traffic.

There are advantages to packet voice beyond price. In a packet voice system, packets are only transmitted when the speaker is active. This makes it reasonable to introduce services in which speakers are inactive most of the time, such as a reverse channel in a lecture or a monitor in a room with a disabled person. In a packet voice system, voice coding is performed at the end user locations, rather than in the network. This makes it possible for coders to evolve more quickly and to use different coders for different connections. Coders can evolve more quickly because only the participants in a conversation must agree to change a coder, rather than modifying the entire network. Since the end users can negotiate the coder that they use, they may elect to use more bandwidth for a speaker phone connection than a telephone connection, and less bandwidth for a monitor. The flexibility offered by packet voice will become important once the quality issues are resolved.

Our measurements are based upon the current best-effort datagram service of the Internet and uses the UDP[1] protocol. Extensive measurements have been made using another Internet protocol, TCP[2,3]. TCP provides a reliable delivery service by retransmitting lost packets. It is inappropriate to use TCP for interactive voice because of the time needed to recover the lost packets. There is an evolving collection of measurements on Internet voice[4,5,6,7]. The work reported here adds to this collection.

The traffic on the Internet is a function of the end-points of the connections and the time of day. Our measurements have been performed on three separate circuits, an intrastate circuit, a cross country circuit, and an international circuit. The measurements were collected for ten minute periods during every hour for fourteen days. The details of the measurements are presented in section 3, along with a number of plots of the delay and loss on the various connections.

In section 4 we define the quality of an interactive packet voice connection in terms of distortion free intervals. A distortion free interval is a period of time, without packet losses, that is long enough to carry useful speech segments. We consider packets that do not arrive before a deadline to be lost. A limited number of lost packets can restored because of the redundancy in speech. We note that restoring single lost packets improves the quality, but when there is more than one packet missing, there are usually too many missing packets to restore. When single losses are restored our quality measure is not sensitive to the exact definition of a distortion free interval. Therefore, we can avoid determining the precise relationship between quality and the length of the interval.

In section 5 we examine quality as a function of the circuit and time of day. As expected, the Internet provides a higher quality on short circuits than on long circuits, and the quality can degrade during busy hours. We also note that the quality of connections can vary significantly on the same circuit, at the same time of day. For instance, the average quality of the international connection during the busy period is relatively poor. However, closer examination reveals that half of the connections are fairly good, and the other half are awful.

In the conclusion we propose voice services based upon our measurements. The services involve both bypassing the telephone network with the Internet, and bypassing the Internet with the telephone network.

## 2. Model

The delay in a packet phone network is the sum of a packet assembly delay at the source, a network delay, and a receiver delay. The delay at the source is selected to balance network efficiency and the response needed in an interactive system. The network delay is beyond the control of the end user. The delay at the receiver is inserted to compensate for the network delay.

The packet assembly delay is the time between the first and last sample in a packet. The source periodically generates samples of the speech signal that are collected for transmission in a packet. The network efficiency is the ratio of samples in a packet to the entire packet length, which includes a header. The network efficiency increases when there are more sample in a packet, but so does the delay imposed on the first sample in the packet.

The network delay consists of a fixed component and a variable component. The fixed component is caused by propagation delays and processing at the network nodes. The variable component is caused by contention for channels or processors. In addition, when channel contention is high, buffers overflow, and packets may be lost.

The receiver delay compensates for the variable component of the network delay. The speech signal is reconstructed at the receiver by periodically playing out the samples. If a packet fails to arrive before the last sample in the previous packet is used, a glitch occurs in the speech waveform. The receiver delays the first packet before it is disassembled, so that most packets will arrive before the last sample in the previous packet is used.

There is a tradeoff between the receiver delay and the fraction of the packets that are lost. If a packet isn't available when it is scheduled to be disassembled, either because it was actually lost in the network or because it experienced a long delay, the receiver replaces it. If the packet arrives later, it is discarded. Effectively, packets that arrive after a deadline are lost. The receiver delay must be long enough that most packets arrive before they are replaced, but short enough that the conversation remains interactive. In the telephone network two satellite delays, about 1/2 second, is considered excessive. Delays of a few seconds or tens of seconds occasionally occur in the Internet. Clearly, the receiver delay in an interactive voice connection can not be long enough to wait for all of the packets that eventually arrive.

## 3. Measurements

In our experiments, traffic was generated by three sources, one at Rutgers University in New Jersey, a second at the University of California at Berkeley, and a third at GMD in Berlin, Germany. The destination for all three sources was at Bell Labs in Murray Hill, New Jersey.

Measurements from each source were taken for a two week period. For ten minutes every hour a source transmitted short, 64 byte, UDP packets every 20 milliseconds. Each source transmitted 30,000 packets per ten minute connection, and more than 10 million packets during the course of the experiment. During a connection a source transmitted 25.6 kbps, which is comparable to the rate of a voice source.

The three sources were on shared machines which we used as guests. We attempted to transmit at regular intervals, but occasionally there were periods as long as 19 seconds when we could not generate packets. Therefore, we had to take measures to distinguish between delays introduced by the end user equipment and the network.

The machine at Murray Hill was a dedicated machine, located outside our firewall. This eliminated contention for resources at the receiver and the additional delay that occurs when passing through a firewall.

The packets we transmitted contained departure timestamps, session numbers and sequence numbers. The session number identifies the ten minute interval in which the packet is transmitted and the sequence number identifies the packet order within a session. These numbers detect packets that are lost or received out of sequence. Timestamps are inserted in each packet just prior to transmission. The timestamps enable the receiver to distinguish between delay introduced by the source and that introduced by the network.

The network delay is the difference between the time when a packet is received and the time when it is transmitted. The $i^{th}$ packet is timestamped with its transmission time $t_i$. When the packet is received, the time, $r_i$, is noted. The difference between the transmit and receive time is the sum of the fixed component for the connection, $T$, and the variable delay for the packet, $x_i$.

When we take a large number of measurements, we are likely to encounter a variable delay component that is close to zero. We estimate the fixed component as

$$\hat{T} = \min_i \left[ r_i - t_i \right]$$

where the minimization is over the 30,000 packets in a 10 minute session. If the clocks at the transmitter and receiver are synchronized, using the Network Time Protocol,[8,9] or a similar procedure, $\hat{T}$ is the fixed delay through the network. If the clocks are not synchronized, $\hat{T}$ is just a constant, and may even be negative, depending upon the setting of the clocks at the transmitter and receiver.

Given $\hat{T}$, we estimate each variable delay as

$$\hat{x}_i = r_i - \left[ t_i + \hat{T} \right]$$

If the source is unable to transmit a packet when planned, jitter occurs between the transmitted packets, but the effect of this jitter is not included in $\hat{x}_i$.

Typical distributions of the variable delay component for intrastate, cross country and international circuits are shown in figure 1. In figure 2 there is a plot of the probability of losing packets on the three circuits as a function of the time of day.

## 4. The Quality of the Channel

The Internet does not provide an acceptable voice communication channel by telephone standards. However, at times the Internet does provide a useful channel. The contradiction is partly due to people setting a lower standard for a service that they feel is free. However, part of the reason is that the noise mechanisms in the phone network and Internet are different, and the measure used in the phone system is not

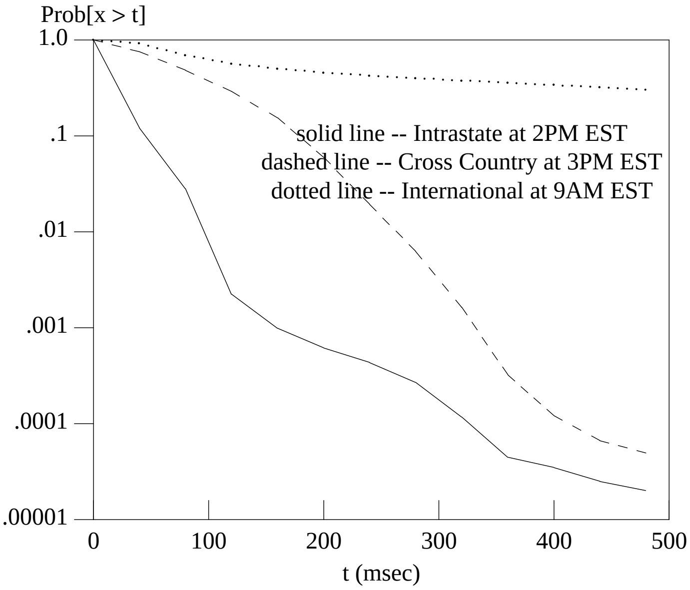


**Figure 1.** Distribution of the Variable Component of the Delay on Intrastate, Cross Country and International Circuits

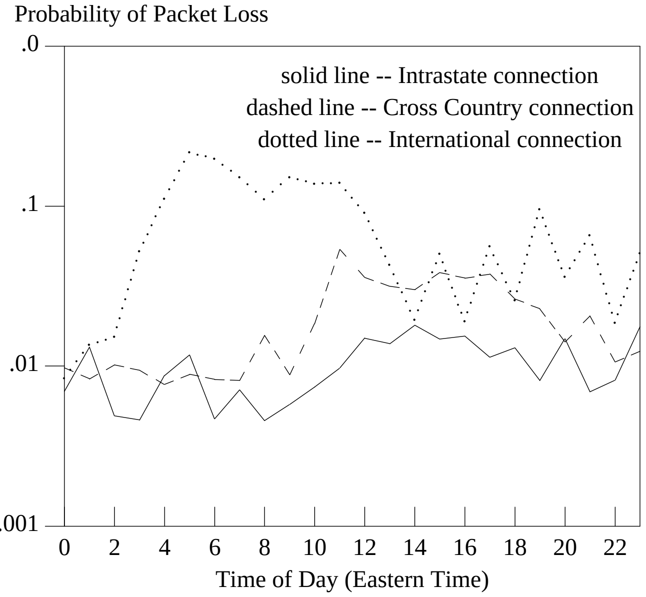


**Figure 2.** Packet loss averaged over a 14 day period as a function of the time of day, on an intrastate, cross country and international circuit

appropriate for the Internet.

Noise in the phone network distorts individual samples while losses in the Internet affect one or more packets of samples. If the line drops out for a brief period in an interactive conversation we can ask the other party to repeat what they were saying. Providing that drop outs don't occur too frequently this type of loss may be more acceptable than constant static. For instance, consider two systems that send packets every twenty milliseconds and lose five percent of the packets. In a system that loses every twentieth packet, two or three packets are lost every second, the received signal has a lot of static and is very difficult understand. If, on the other hand, all of the packet losses in a three minute phone call are clumped together in two or three segments, we lose the channel for nine seconds, but it is perfect for the remaining 171 seconds and provides a useful communications path. On a shorter time scale, if a listener can clearly receive most of the active speech intervals, he can usually understand the conversation.

In order to capture this effect we define the quality of a channel as the fraction of the time that the signal is received without distortion for intervals of time that are "long enough" to convey useful speech segments. Time is divided into distorted intervals and distortion free intervals. An error event is any sequence of lost packets that is not restored. If the time from the tail of one error event to the head of the next is greater than a minimum then that time is a distortion free interval, otherwise the two error events are part of the same distorted interval. The quality of the connection is the sum of the time in distortion free intervals over the total time of the connection.

The quality of a connection depends upon the number of successive lost packets that we restore by interpolation and upon our definition of a minimum loss free interval(MLFI). The accuracy of interpolation decreases, and the amount of processing required increases, as we increase the number of successive lost packets that we try to restore. Therefore, the number of successive losses that we restore should be small. In measurements of speech conversations, Brady[10,11,12] found that the average active interval in speech is about 1.3 seconds. The useful distortion free interval should have a duration that is of the same order of magnitude as an active speech interval. Therefore, we consider defining the minimum error free interval at values between a half second and a few seconds.

In figure 3 there are nine plots, three each for the intrastate, cross country and international circuits. Each of the three plots for a circuit is for a specific MLFI, and shows the quality of the connection as a function of receiver delay for several restoration capabilities.

Figure 3 suggests restoring only single lost packets. In all nine plots, there is a much larger gap between 0 and 1 restorations than between 1 and 2 or 2 and 3 restorations. This indicates that when more than one packet is missing, there are likely to be more than three packets missing.

Figure 3 also shows that when single losses are restored the quality measure is relatively insensitive to changing the MLFI from .5 to 2 seconds. It is worth noting that if single lost packets are not restored then the fraction of the time that each circuit is distorted increases rapidly as the MLFI increases from .5 to 2 seconds. In most of the remainder of this paper we will restore single lost packets and set the MLFI to 2 seconds.

Before proceeding, it is worth commenting on the loss mechanisms on the Internet. Increasing the MLFI has little effect when single lost packets are restored, but has a large effect when no lost packets are restored. This indicates that the error process that generates multiple losses has an inter-occurrence interval of more than two seconds, while the process that generates single losses does not. The fact that most bursts are more than 3 packets long indicates that when a component in the network starts losing packets, it is likely to stay in a bad state for more than 80 msecs. This observation is

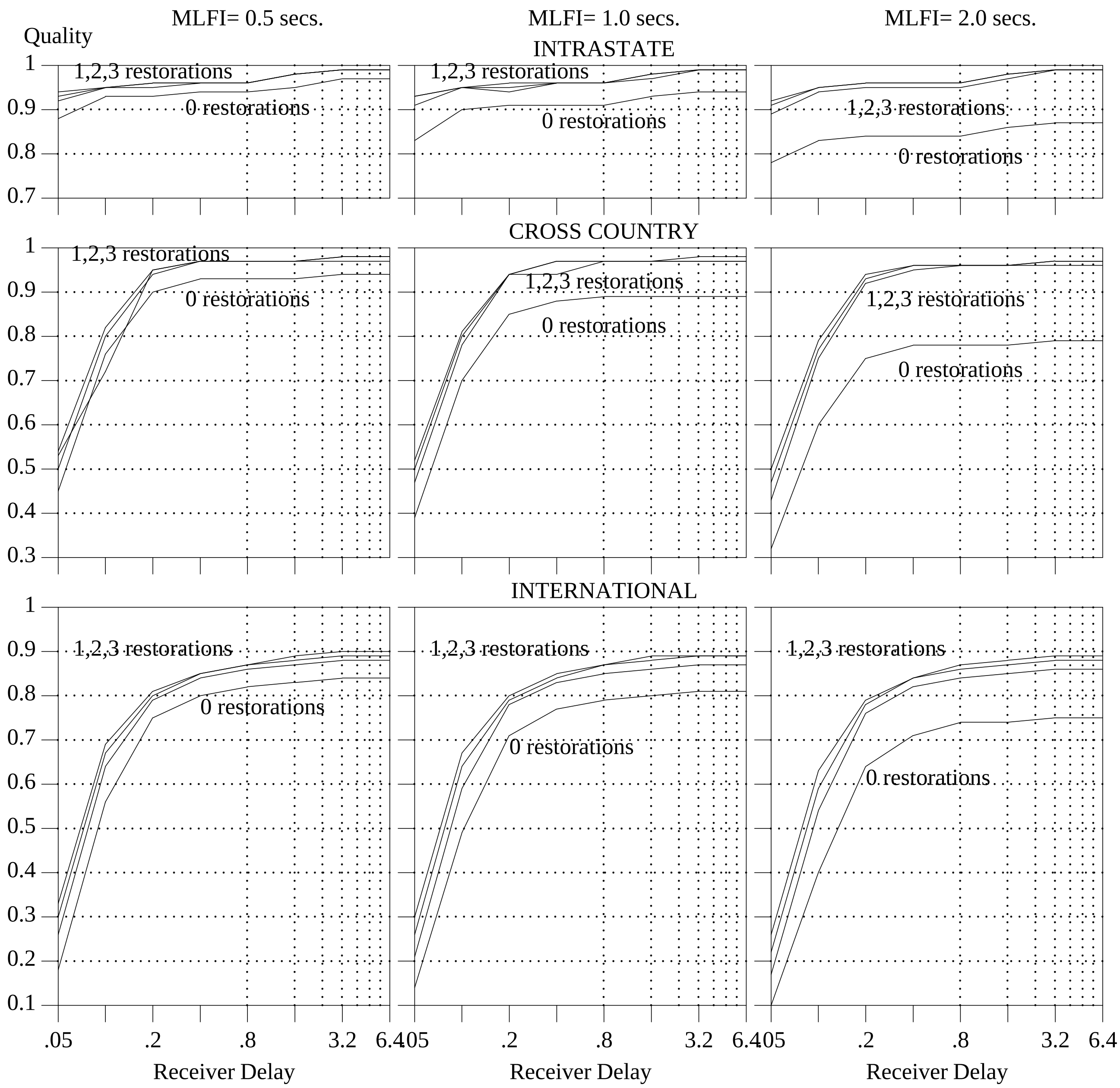


**Figure 3.** The quality of a connection as a function of the delay at the receiver, when the minimum interval without loss (MLFI) is .5, 1 or 2 seconds, and the number of successive lost packet that are restored is 0, 1, 2, or 3.

consistent with other measurements that have been made on Internet circuits[13,14,7].

## 5. Factors Affecting the Quality of a Connection

The quality of a connection is determined by the contention for resources along the path. Contention varies with the length of the circuit, the time of day, and the specific set of active users sharing the facilities.

Sources and destinations that are further apart are more likely to have a poor quality connection. A packet is more likely to encounter a facility that is temporarily busy as it passes through more intermediate routers. A packet is also more likely to traverse an expensive, and therefore more heavily utilized, path the further it travels. This is particularly true when the source and destination are on different continents.

On each of the circuits there are different busy periods. There are times of the day when more network users are active, most often during the normal work day. There are also times when sources and destinations in different time zones are more likely to communicate.

Finally, at a given time of day, on a given circuit, different levels of contention can be encountered. For any connection, there are non-zero probabilities that a source destination pair will encounter poor quality or good quality.

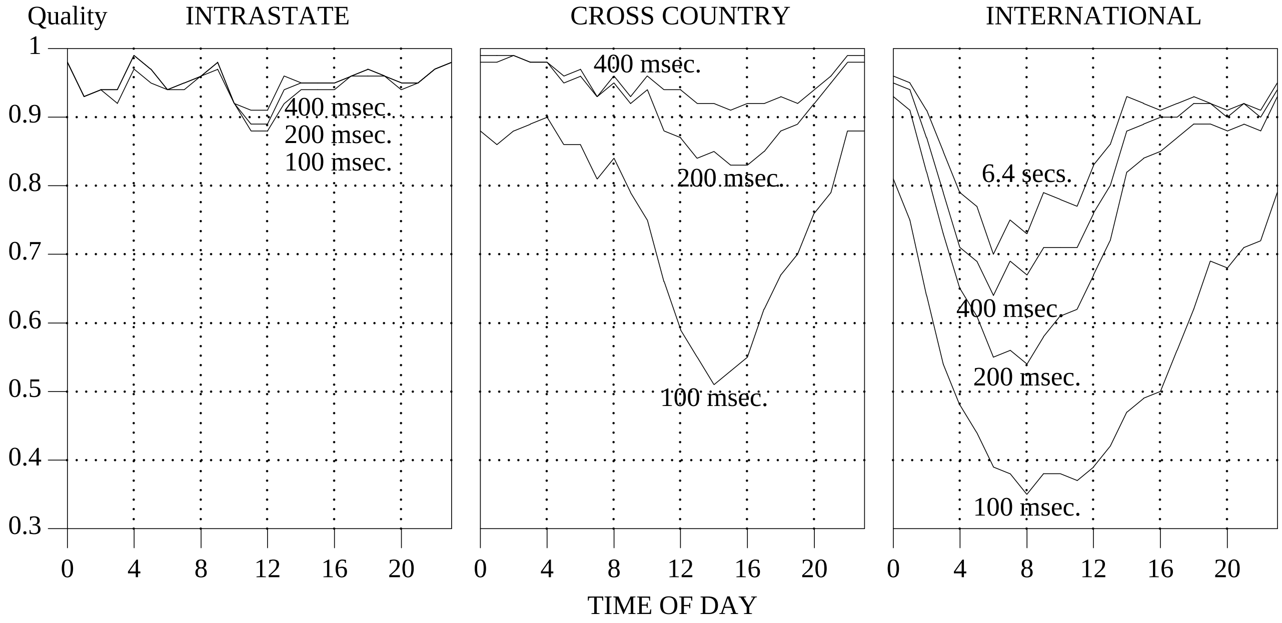


**Figure 4.** The quality of a connection as a function of the time of day, with single lost packets restored, a 2 second minimum loss free interval, and receiver delays of 100, 200, and 400 milliseconds. ( Also 6.4 seconds for the International circuit)

### 5.1 The Length of the Connection

In figure 3 the x-axis is the receiver delay. The delay is plotted on a log scale in order to both examine values between 50 to 500 milliseconds, the range in which we want to operate, and to determine the fraction of the packets that are lost by the network. We assume that most packets with delays exceeding 6.4 seconds are lost by the network.

When single losses are restored, and the MLFI is 2 seconds, the knee of the delay curve is our recommended receiver delay. Decreasing the delay below the knee causes a sharp decrease in the quality of the connection, and setting the value far above the knee does not significantly increase the quality. The recommended receiver delay for the intrastate circuit is 100 msec., for the cross country circuit is 200 msec., and for the international circuit is 400 msec. With this set of receiver delays, the quality of the intrastate and cross country connections exceeds 95%, and the quality of the international link is between 85% and 90%. [The quality is the fraction of the time spent in distortion free intervals.]

Note that on the cross country and international links, the fraction of the time that the connection is distorted is not reduced significantly if the receiver delay is increased to 6.4 seconds. Therefore, the residual distortion is most likely caused by packets that have been lost by the network, rather than packets that we did not wait to receive.

### 5.2 The Time of Day

A commonly reported problem with the Internet is that the delay is much greater during the busy hours of the day. In order to study this phenomenon, we examine the fraction of the time that each circuit is undistorted as a function of the time of day. We use the operating points suggested in the previous section: The MLFI is 2 seconds and single lost packets are restored. In order to test our recommendation to set the receiver delay at the knee of the curve, we consider several values of receiver delay that are around the knee.

The results are displayed in figure 4. The time of day on all of the plots is the time on the east coast of the U.S.. Since the experiment ran for 2 weeks, there are 14 samples for each hour. We found that simply averaging the 14 samples resulted in curves that were very difficult to interpret. A deeper look at the data pointed out a big difference in the quality of the connections. On some connections there are almost no distortion free intervals. These "bad" connections have a big effect on the average, and there are large changes in the average depending upon whether we happen to encounter one or two of these outlying samples. The sample size must be made larger to make the average meaningful. As an alternative to increasing the duration of the experiment, we have plotted the samples for three adjacent hours at each point. For instance, the point at 5 AM is the average of the 42 samples from 4 AM, 5 AM, and 6 AM.

As expected, figure 4 shows that there are periods of time when the Internet is busier than others. The busiest periods for connections that cross the country, or the ocean, occur during the overlap in the workday between the two end points. It is curious to note that on the intrastate connection the busy period isn't as pronounced as in the other two connections. One reason is that the Internet is no longer only used for work. Another reason is that the quality of the intrastate connection is good even during the busy hour.

The plots in figure 4 partially confirm our use of the knee of the curve to select the receiver delay, but also point out the danger in looking at averages. The figure shows that the quality of the cross country connection can be significantly improved by increasing the receiver delay from 200 msec.

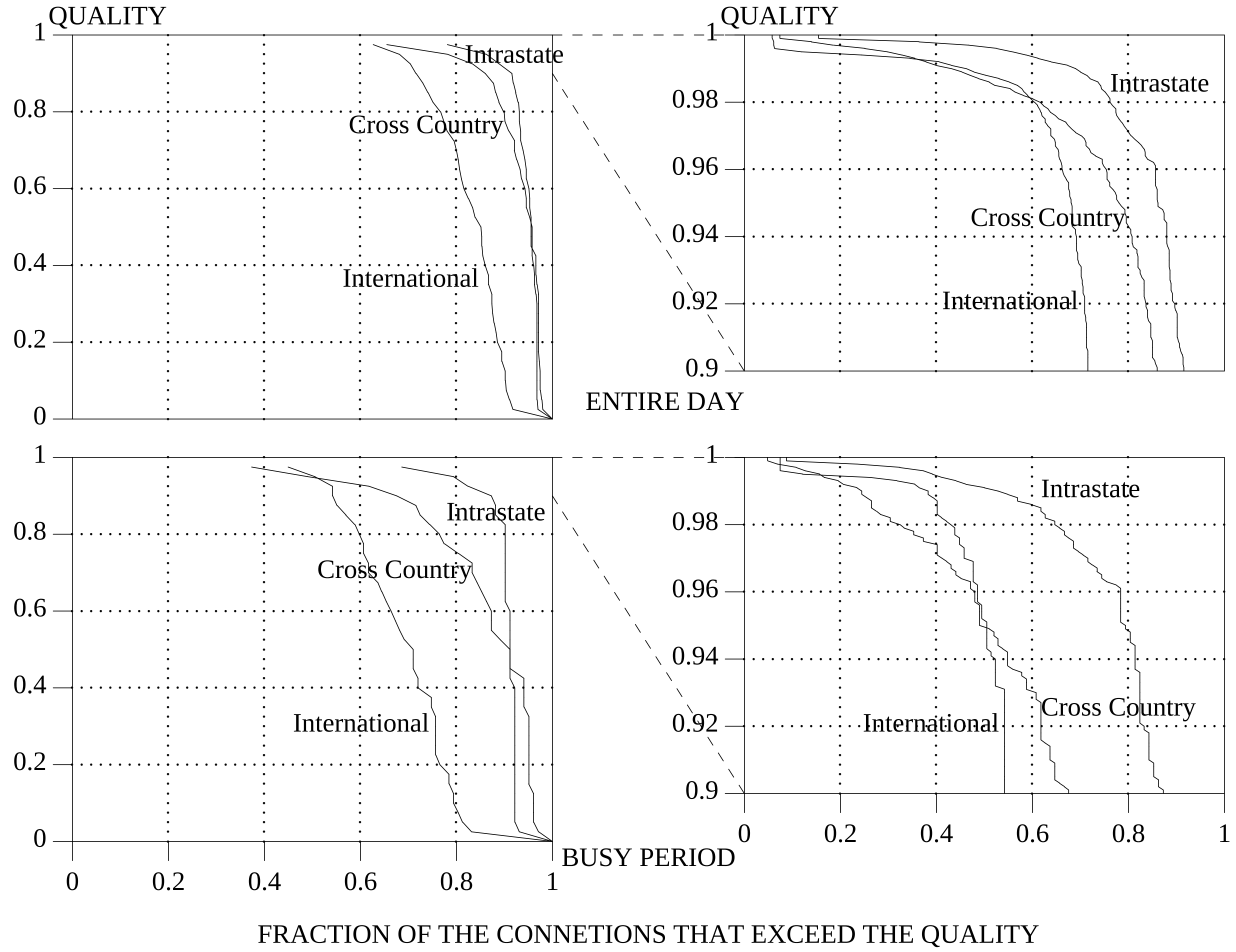


**Figure 5.** The fraction of the connections that exceed a quality level during the busy period and during the entire day. Single missing packets are restored The receiver delay on intrastate connections is 100 msec., on cross country connections is 200 msec, and on international connections is 400 msec.

to 400 msec. during the busy periods.

### 5.3 The Specific Connections

In the previous section we mentioned that there were some connections that had high quality and others that had a very low quality. In this section we quantify that statement by plotting the fraction of the connections on each circuit that exceed a specific quality. The fractions are taken both over the entire day and over the busy period of the day.

In order to obtain a reasonable number of samples, the busy period is defined as the six hour interval with the heaviest use. This results in 14*6 = 84 samples per circuit. The busy period is selected as:

— 10 AM to 3 PM on the intrastate connection,
— 12 noon to 5 PM on the interstate connection, and
— 4 AM to 9 AM on the international connection.

The plots in figure 5 use single restorations, an MLFI of 2 seconds, and our recommended receiver delays.

In general, the intrastate connections have very good quality. Each of the connections lasted for ten minutes. During the busy period, 80% of the connections are free of distortion for 9 minutes and 30 seconds of the ten minute call, and nearly 70% of the calls have less than 10 seconds of distortion during a ten minute connection.

In contrast, the international connections have a bimodal distribution, they are either good or very bad. More than half of the connections have reasonable quality, with less than 30 seconds of distortion in the 10 minute connection. On the other hand, 20 % of the connections have extremely bad quality with less than 30 seconds that are free of distortion out of the 10 minutes.

Figure 5 shows the need to reexamine the selection of parameters on the cross country circuit. With the selected parameters, the cross country circuit generally has a higher quality than the international circuit, however, a larger fraction of the international connections are distortion free more than 95% of the time. We have already observed that during the busy period it may be worthwhile increasing the receiver delay on the cross country link to 400 msec.

Figure 6 shows that the quality of the cross country connection can be significantly improved by changing our selection of parameters. By increasing the receiver delay from

200 msecs. to 400 msecs. and the number of packets that are restored from 1 to 2,

— the fraction of connections that have less than 30 seconds of distortion increases from slightly more than 50% to almost 90%, and

— the fraction of connections with less than 10 seconds of distortion increases from less than 1/3 to more than 3/4.

The results for the cross country connection indicate that while restoring double missing packets does not increase the average performance very much, it can make good connections very good. The results also indicate that the receiver delay should be allowed to float up to a maximum value, rather than being fixed. One way to do this is to set the delay, but also count the packets that arrive late and increase the delay when it is warranted.

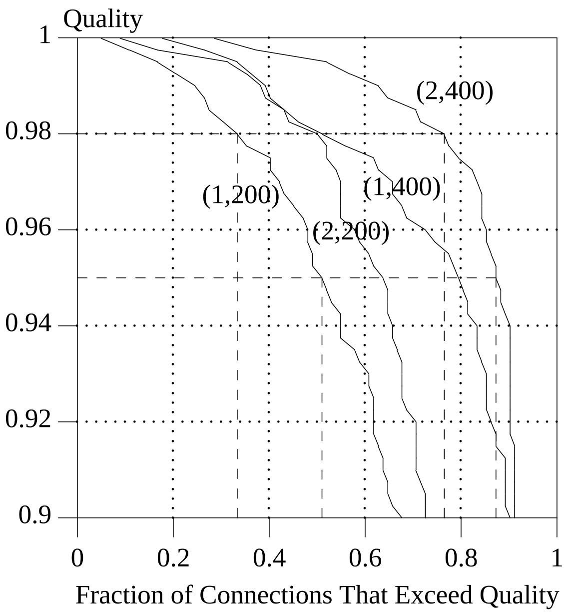


**Figure 6.** Effect on the cross country circuit during the busy period of increasing the receiver delay from 200 to 400 msecs. and increasing the successive packets that can be restored from 1 to 2

## 6. Conclusions

Single lost packets should be restored when sending 20 msec. long voice packets over the Internet. Restoring two consecutive lost packets may also be useful to make good connections very good. However, when more than one packet is lost there are likely to be a large number of lost packets and restoring more than two consecutive losses is not worthwhile.

We have defined the quality of a connection as the fraction of active intervals that are received without distortion. Once we have restored single lost packets, the quality measure is not affected by the exact size of the interval.

In our experiment, we found that the quality of the intrastate circuit is better than that of the cross country circuit, which is better than that of the international circuit. A body of measurements is developing that indicates that this is generally true[4,5,6,7]. Based upon this generalization, most people who are currently using the Internet for voice are using it backwards. The Internet is best suited for carrying voice for short distances, while most applications use it as a long distance alternative. Two reasonable applications of the current Internet include bypassing the local telephone company and corporate networks.

In a bypass network intrastate Internet connections are used to collect voice packets in a region, the telephone network is used to carry the packets over the long distance, then the Internet is used to deliver the packets near the destination. The Internet is used to bypass the local telephone network for intrastate communications, and the telephone network is used to bypass the Internet for long distance communications.

There are two advantages to using packet voice for local bypass. First, local access charges are avoided, and second, fewer bits need to be transmitted. The first advantage is regulatory and will likely disappear. The second advantage results from performing voice coding at the end-points, where it is easier to take advantage of recent advances in coders than in the network. To note the extent of the second advantage, the phone network transmits 64 kbps voice channels while most Internet voice coders use 16 kbps or less, and only transmit during active intervals. As a result 8 to 10 packet voice connections can be transmitted on each telephone connection. Advantages due to performing coding at the end-points rather than in the network are likely to remain.

In a corporate network, PC's in offices place voice packets on a LAN, a multiplexer collects the traffic that is designated for a remote LAN and uses the circuit switched network to transport that traffic. At the distant location the voice packets are placed on another LAN and delivered to the recipient's PC. In a corporate network the utilization, and hence the quality of the connections on the LAN's can be controlled. The same reduction in capacity is obtained between the distant locations in a corporate network as in bypass strategy. The office environment has the additional advantage that most users already have PC's so there is no need for additional equipment.

Although the Internet is not suitable for long distance voice communications today, it does not mean that it will never be suitable for this type of communications. There is considerable work on reservation systems for the Internet[15] that will improve the quality of connections.

There is an interesting analogy between sending voice on the Internet today and using data modems on the telephone network in the late 60's. In that period, 1200 bps modems were being deployed, but not all telephone circuits were suitable for carrying modem signals. Telephone technology was capable of carrying data, however, not all parts of the network were able. As the need for data communications increased, the telephone network was upgraded to meet the need.

Our experiments indicate that the Internet is capable of carrying voice, although it is not always able. Just as the telephone network evolved to meet the need for data, it is likely that the Internet will evolve to meet the need for voice.

### *REFERENCES*


[1] J. Postel, "User Datagram Protocol," RFC-768.

[2] R. Caceres, P. Danzig, S. Jamin, D. Mitzel, "Characteristics of Wide-Area TCP/IP Conversations," Proc of SIGCOMM 1991, pp. 101-112.

[3] J. Zinky, F. White, "Visualizing Packet Traces," Proc. of SIGCOMM 1992, pp. 293-304.

[4] D. Sanghi, A. K. Arawala, O. Gudmundsson, B. N. Jain, "Experimental Assessment of End-to-End Behavior on Internet," IEEE Infocom ’93, San Francisco, Ca., Mar 30- Apr 1, 1993, pp 867-874.

[5] R. Ramjee, J. Kurose, D. Towsley, H. Schulzrinne, "Adaptive Playout Mechanisms for packetized Audio Applications in Wide-Area Networks," IEEE Infocom ’94, June 14-16,1994, Toronto, Canada, pp. 680-688.

[6] J. Bolot, "End-to-End Packet Delay and Loss Behavior in the Internet," SigComm ’93, pgs 289-298.

[7] J. Bolot, H. Crepin, A. Garcia, "Analysis of Audio Packet Loss in the Internet," Proceedings of NOSSDAV ’95, 1995, pp. 163-174.

[8] D. L. Mills, "Internet Time Synchronization: The Network Time Protocol," IEEE Trans. on Communications, Vol. 39, pp. 1482-1493, Oct. 1991.

[9] D. L. Mills, "Improved Algorithms for Synchronizing Computer Network Clocks," IEEE/ACM Trans. on Networking, Vol. 3, No. 3, pp. 245-254, June 1995.

[10] P. T. Brady, "A Technique for Investigating On-Off Patterns of Speech," BSTJ, Vol. XLIV, No.1, Jan. 1965, pp 1-22.

[11] P. T. Brady, "A Statistical Analysis of On-Off Patterns in 16 Conversations," BSTJ, Jan. 1968, pp. 73-91.

[12] P. T. Brady, "A Model for Generating On-Off Speech Patterns in Two-Way Conversation," BSTJ, Sept. 1969, pp. 2445-2472.

[13] H. Schulzrinne, J.F. Kurose, D.F. Towsley, "Loss Correlation for Queues with Bursty Input Streams," ICC ’92, pp. 219-224.

[14] J.M. Ferrandiz, A. Lazar, "Consecutive Loss in Real-Time Packet Traffic," Proceedings of the 4th International Conference on Data Communication Systems, June, 1990, pp. 353-365.

[15] D. J. Mitzel, D. Estrin, S. Shenker, L. Zhang, "An architectural comparison of ST-II and RSVP," Proc. of IEEE Infocom ’94, June 1994, Toronto, Canada.