8

Performance Considerations

This chapter examines several issues pertaining to the performance of VoIP networks. The first part of the chapter examines the trade-offs of packet size, buffer size, packet loss and packet latency. The next part of the chapter provides a summary of three studies conducted on VoIP performance in private internets, and in the public Internet.

PACKET SIZE, BUFFER SIZE, LOSS, AND LATENCY

The VoIP designer must pay attention to buffer sizes, packet sizes, and the packet loss rate. The larger the packet loss, the worse the audio quality will be at the receiver. On the other hand, large packet sizes increase the delay and so do large buffers. To see why, let us examine a simple G.711 64 kbit/s voice signal.

First, consider the loss of user traffic. The size of the packet is quite important for speech because of the concept of packet length (the duration of the packet on the channel). Packet length is a function of the number of user bits in the packet, and the coding rate of the signal (for example, 64 kbit/s). Studies reveal that losing traffic that is around 32–64 ms (for G.711 traffic) in duration is disruptive, because it means the loss of speech phonemes. On the other hand, cell loss of a duration of some 4–16 ms is not noticeable nor disturbing to the listener. Therefore, a payload size of anywhere around 32–64 octets would be acceptable to an audio listener. The actual perception of audio loss is a function of other factors such as the compression algorithms used, etc. But for this general example, the following examples show loss for G.711 traffic. It can be seen that the longer packets suffer more loss. The examples are for packets with 32, 48, and 64 bytes of user voice traffic.[1]

32 octets * 8 bits per octet = 256 bits
256 / 64,000 = .004

48 octets * 8 bits per octet = 384 bits
384/64,000 = .006

64 octets * 8 bits per octet = 512 bits
512 / 64,000 = .008

Next, consider buffer size. A larger buffer will increase delay, and decrease the loss rate, because the larger buffer allows more flexibility in playout, and the machine does not have to discard as many packets. But the continued decrease of the buffer size, while decreasing delay, means more packets will be discarded. In effect, as the buffer size approaches 0, the machine operates at wire speed, but will experience more loss of traffic. So, it is a catch-22 situation. Figure 8–1 shows the relationships between packet loss and unidirectional delay [COX98]

In the next section of this chapter, we examine tests conducted on VoIP products in private intranets. I will use one of these tests here to make some other points about the subject of packet size, latency, and voice quality [MIER99].[2] As just mentioned, the voice packet should be small, in order to reduce latency and improve quality. To amplify this thought, large packet sizes are inversely proportional to interactive voice

---

[1] These examples are from studies performed on ATM cells. After extensive deliberations in the ATM standards working groups, it was agreed that a cell size between 32 and 64 octets would perform satisfactorily in that it (a) worked with ongoing equipment (did not require echo cancellers), (b) provided acceptable transmission efficiency, and (c) was not overly complex to implement. Japan and the United States favored a cell size with 64 octets of user payload; Europe favored a size of 32 octets.

[2] [MIER99] Mier, Edwin, E. "Voice-Over-IP: Better and Better," *Business Communications Review*, January, 1999.

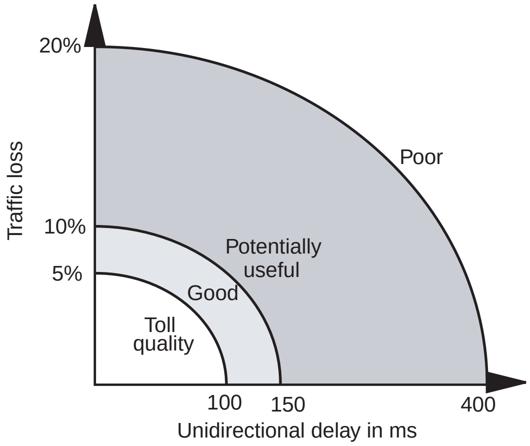


Figure 8–1 Traffic loss versus traffic delay [COX98]

quality. One could take this idea to the extreme and postulate a packet of say 1 byte (or even 1 bit!). Obviously, there is a point of diminishing returns where the overhead of headers to the miniscule user packet is so high that it militates against building an efficient network.

Another consideration is to attempt to build the packet size in consonance with the output of the codec. For example, the G.729 and G.729a codecs produce a 10-ms sample (10 bytes, with 1 byte per ms). In so far as possible, it is a good idea to package the packets in 10-ms bundles.

But this approach may not be possible. For example, if these voice packets are placed inside a fixed-length frame, they may not fit exactly in the fixed payload of the frame. If this is the situation, and the payload size of the frame cannot be altered, the system should be able to break the sample into even-byte boundaries and place it into successive frames in order to best use the payload bandwidth. This approach is common today, and I provide some examples in Chapter 12.

PERFORMANCE OF VoIP IN PRIVATE SYSTEMS

Our focus for this part of the chapter is to examine two studies conducted by Mier Communications Inc. on VoIP products. The general results of the first study are available from Mier Communications

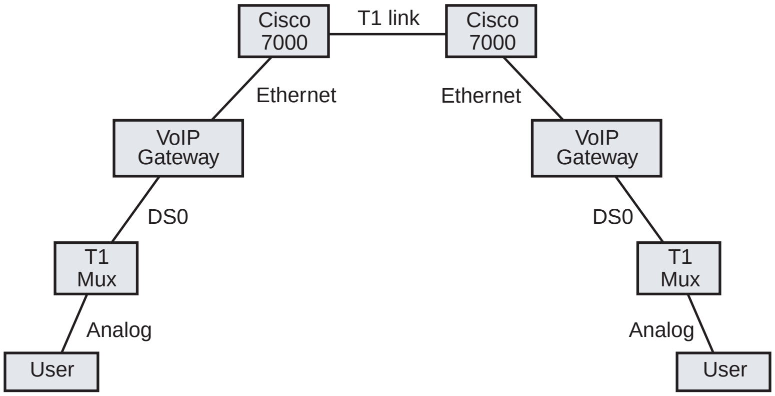


Figure 8–2 Topology for study [MIER98]

[MIER98].[2] The second study is also available from [MIER99] and [MIER99a]. Mier Communications can be reached at ed@mier.com. After we examine the first test, we then look at the new test, and compare the differences.

Figure 8–2 shows the topology layout and configuration for the first study conducted by Mier Communications. The voice traffic was exchanged between two Cisco 7000 routers over an unchannelized T1 link. The voice traffic was created through a simulated analog line and fed into a T1 mux where it was digitized into 64 kbit/s DS0 slots. This signal was fed into the voice-over-IP (VoIP) gateway (the products under test) where they were then input into the Cisco router across an Ethernet 10 Mbit/s channel. The signals were then transported across the unchannelized T1 link to the receiving router where the signal then was sent through an Ethernet interface to the receiving VoIP gateway, then via a DS0 slot to a T1 mux. The mux converted the signal back to analog and transported the signal to the user at either a telephone handset or a speaker.

The study was performed with various voices (female speakers, male speakers, etc.). All speakers recited common voice images such as *do*, *re*, *me*, and so on.

---

[2] [MIER98] Mier, Edwin, E. “Voice-Over-IP: Sounding Better,” *Business Communications Review*, February, 1998.

[MIER99a] Mier, Edwin, E. “VoIP Gateways-Tradeoffs Affecting Voice Quality,” *Business Communications Review Voice 2000*, January, 1999.

Keep in mind that this study did not test voice quality over the Internet. The test was run over the T1 link between the routers. Consequently, you should not infer that these tests state anything about the performance of the Internet in supporting telephony. But the study does provide an interesting and useful assessment of VoIP products. As we shall see, the test reveals that high-quality products are available to run voice over what is essentially data-based protocols, and offer an attractive alternative to public-switched toll services.

Several tests were performed on four products. According to Mier Communications, all four products performed well and provided high-quality telephony images to the end-user. The four products tested were Lucent Technologies, Micom (now part of Nortel), Nuera Communications, and Selsius Systems.

Figure 8–3 reflects the assessment of the quality of these systems when operating under ideal network conditions, The definition of ideal network conditions is a fully available T1 channel which is then unchannelized (not restricted to DS0, TDM slots). In addition, the media operated under relatively error-free conditions without adding delay or latency across the system. That is to say, any latency or delay in the test

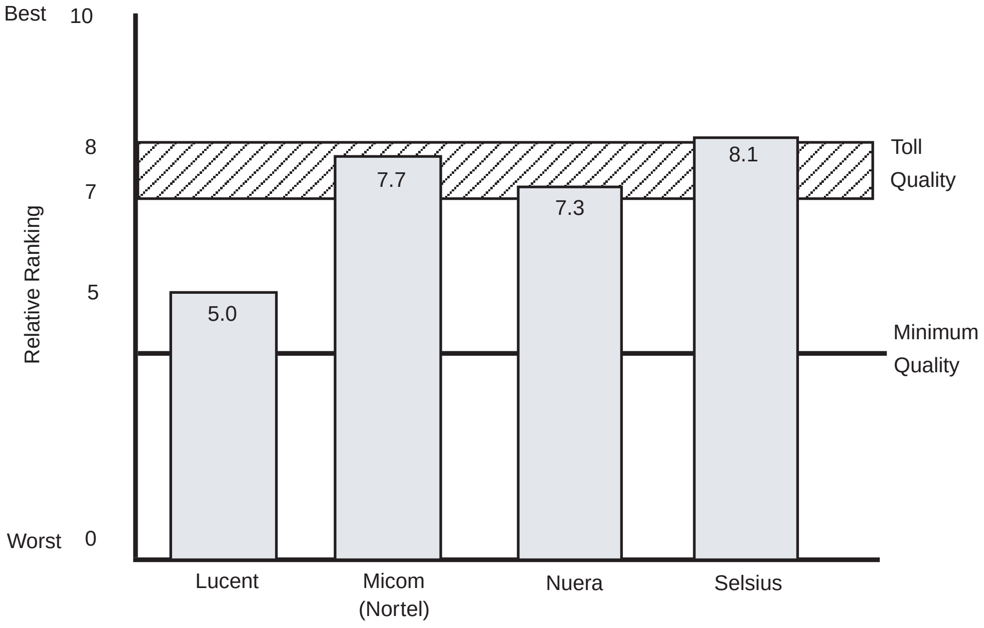


Figure 8–3 Voice quality under ideal conditions [MIER98]

was introduced as a result of the components in the topology and the operations of the VoIPs.

Obviously, the evaluation of the tests is subjective, but the evaluation of voice quality on the telephone network is subjective as well. Notwithstanding, as this figure shows, all products were evaluated as exceeding a minimum quality for the signal. The scale of 0 to 10 was devised with the expectation that a score between 7 and 8 represented the current "toll quality" exhibited in the telephone network. The evaluation of 10 would reflect the best possible telephone voice connection wherein the analog loop is terminated into a digital system one time only, sent a very short distance, and then converted back to an analog signal.

Next, the test was conducted in an environment that exhibited poor quality conditions. To simulate these poor conditions, Mier Communications applied burst errors to the T1 line. The approach was to cycle a 5-ms burst of random errors at a rate of 1×10^2, followed by 50 ms of transmissions with a bit error rate of 1×10^6. The 1×10^6 value is a reasonable assumption for BER performance on a typical local loop. The result of this operation is the introduction of errors in about 10 percent of the IP telephony datagrams. As Figure 8–4 shows, Micom and Nuera

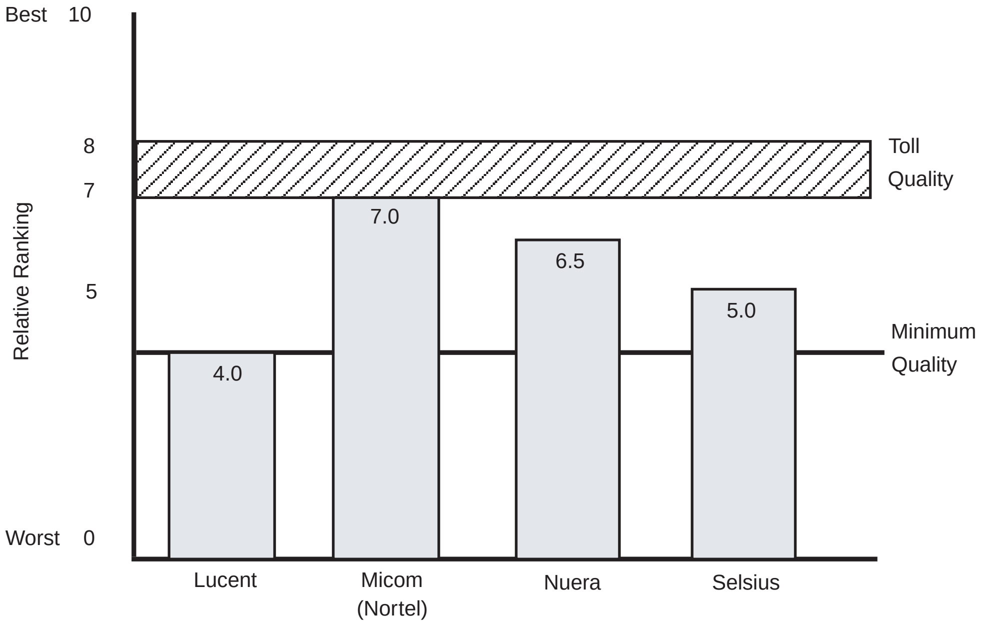


Figure 8–4 Voice quality under poor conditions [MIER98]

continue to exhibit high-quality performance and Lucent and Selsius still exhibited somewhat acceptable quality.

Figure 8–5 shows the bandwidth utilization of the four tested products working under relatively error-free conditions and on a full-unchannelized T1 link. The Nuera system scored the highest with a bandwidth requirement of 14 kbit/s followed by Micom with 18 kbit/s. The Nuera performance approaches that of the sophisticated mobile/wireless vocoders that operate in a range of 13 kbit/s.

Not shown in this figure is the underlying fact that voice quality improved with those systems that used smaller protocol data units (with one exception). Micom's packet size is 91 bytes; Nuera's packet size is 93 bytes; Lucent's packet size is 77 bytes; and Selsius system uses 300 bytes.

In fairness, it should be stated that the performance of Nuera and Micom is partially attributable to the fact that they are using proprietary vocoding techniques whereas Lucent employs the ITU-T G.723.1 and Selsius uses ITU-T G.711. All organizations that belong to the VoIP and the International Media Teleconferencing Consortium (IMTC) have all selected G.723.1 for their basic vocoder.

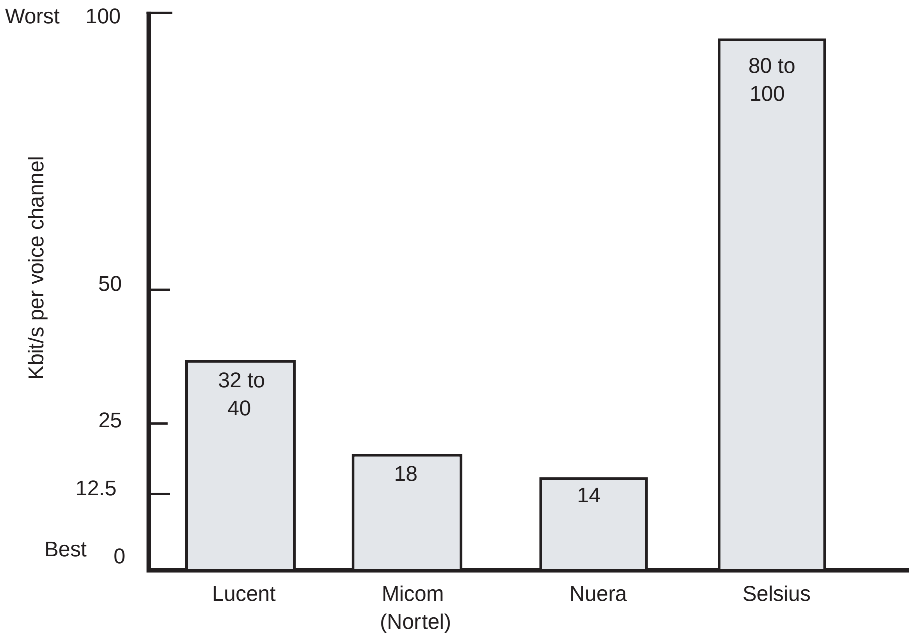


Figure 8–5 Bandwidth used under ideal conditions [MIER98]

It is reasonable to assume that the Lucent and Selsius performance would improve with proprietary schemes.

The trade-off of proprietary versus standardized schemes pertain to the fact that private systems might opt for proprietary schemes because of their superior performance. But if different vendor systems are to interoperate with each other, the standards must be used.

Figure 8–6 shows the latency introduced by the tested systems. All operated within the 250-ms threshold (which is not considered very good performance), and none operated in accordance with the telco standard for one-way latency of 100 ms. The 250-ms delay is considered by some to be the threshold at which it becomes noticeable and disturbing to the users of the system.

In addition, the general study did not focus on where the delay was encountered. Recent tests with voice-over-telephony, in some products (not those cited in the study), indicate that significant delay is occurring in the line card located at the customer premises. Therefore, a clear analysis must focus on the specific components that create the delay because the VoIP may not be the actual culprit for these delay performances.

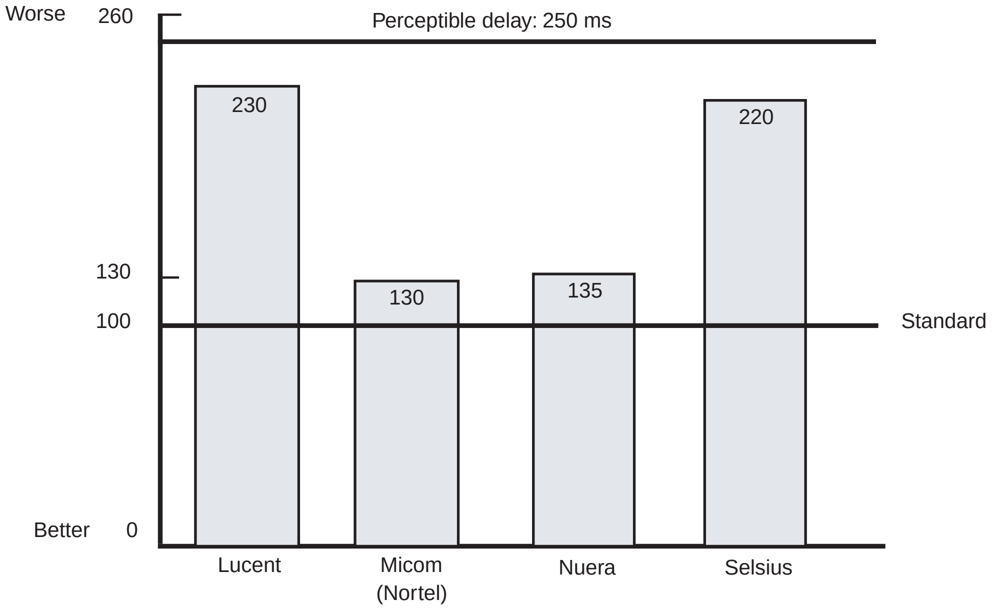


Figure 8–6 Latency: One-way delay [MIER98]

As a result of the interest shown in Mier's first test, the company conducted another test [MIER99]. You can reach Mier at (609) 275–7311 if you need more information about these studies.

Figure 8–7 shows the layout for the study, and notations of the vendors' testing equipment (the tested systems are not shown here). Several points should be emphasized here. First, the test layout is more elaborate and sophisticated than the earlier test. Second, G.723.1 and G.711 were not used by the vendors in this test. Instead, G.729 and G.729a were used. Third, H.323 was part of a test, and it is considered an important component of VoIP by all the tested vendors. Fourth, the results of the test were considerably higher than the earlier test, reflecting better coders and the overall maturation of the technology.

The analog voice files were prerecorded and sent through an Adtran TSU/100 T1 mux, where they were converted to DS0 signals, then to a

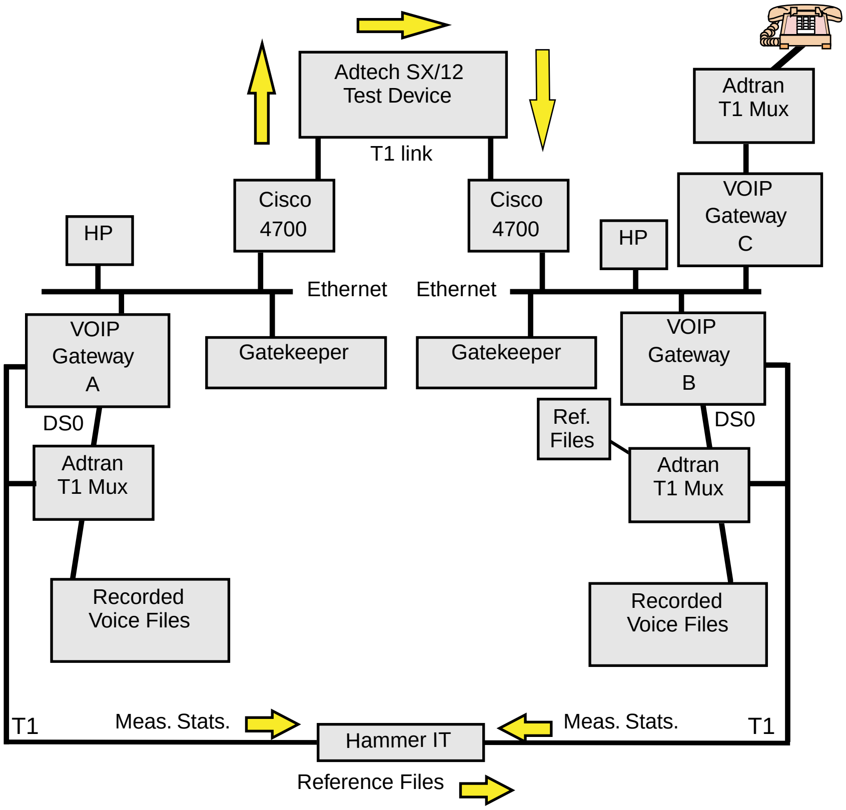


Figure 8–7 Revised test topology [MIER99]

VoIP gateway (Gateway A). Here they were subjected to the vendor's codec and VoIP operations. The resultant VoIP packets were sent through a Cisco 4700 router across an unchannelized T1 (and through an Adtech SX/12 test device) to the receiving site. At this site, the process was reversed (Gateway B), with the voice message rerecorded onto the voice files. A back-to-back T1 line was used to provide reference signals at the receiver. These toll quality samples were compared (by a panel of people) against the tested samples. Gateway C was used to test real-time quality and latency between two people.

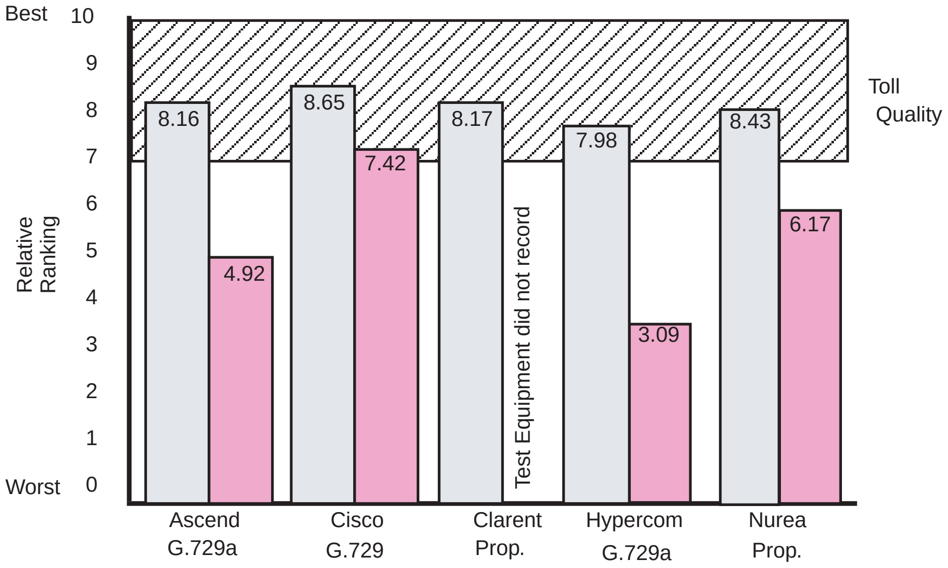
The original voice files were sent to the receiver to act as reference (ref.) files for the comparison tests. A scope from Fluke Corp. was used to measure the latency that each vendor's product added (not shown in the figure). A Hammer IT VoIP Test System also measured latency and generated loads on the gateways, as well as measured voice quality and a variety of other statistical support.

The testing for voice quality was performed by Mier's staff by listening to the rerecorded voice samples, as well as using interactive tests between two people. As Figure 8–8(a) shows, all vendors' products performed very well, under ideal conditions (the higher bars for each vendor). All scored well above minimum toll quality. Mier's conclusions are that the reception of packet voice with the tested VoIP gateways cannot be distinguished from POTS services. The lower bars in Figure 8–8(a) show the quality during poor conditions, in which periodic error bursts were placed on the link.

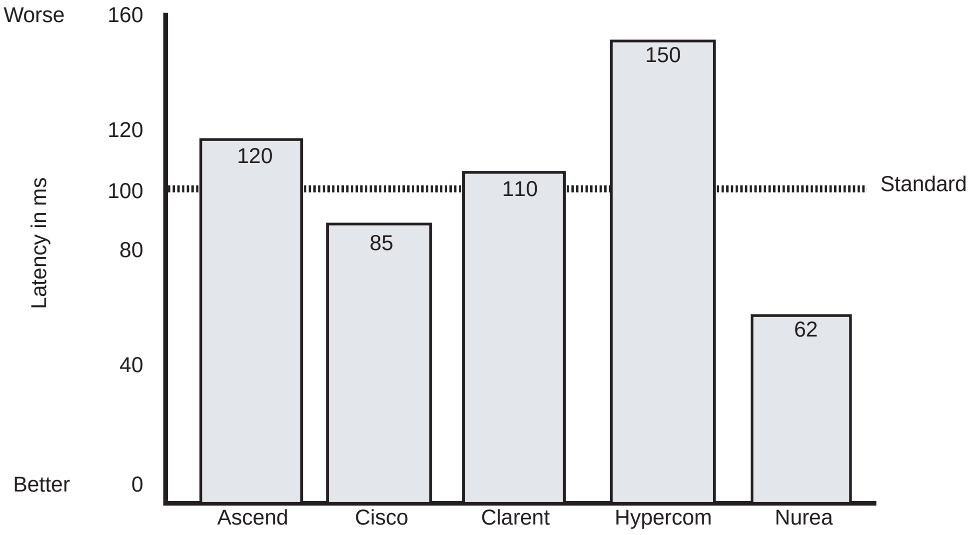
The figure also shows the codecs used for the tests. The vendors used G.729, G.729a, or proprietary coders. All vendors have products with other codecs, such as G.711, and most of them have or are planning to support G.723.1.

Figure 8–8(b) shows the one-way latency for the tested products. Mier's ongoing tests have made them attuned to quality vs. latency, and they conclude that once latency drops below 90 to 80 ms, a person cannot tell the difference between the vendors, or for that matter, between the vendors and the public switched network. After all the tests, the testers could actually tell the differences between latencies under 100 ms, slightly more than 100 ms, and considerably more than 100 ms.

This study was amplified further in [MIER99a]. As you might expect if you read the first part of this chapter, the vendors whose traffic experienced the lowest latency also experienced the best quality. What you might also be thinking is how large a packet do the vendors' products produce? After all, we learned that larger packets lead to more loss on noisy links, and they increase latency.



(a) Voice quality tests



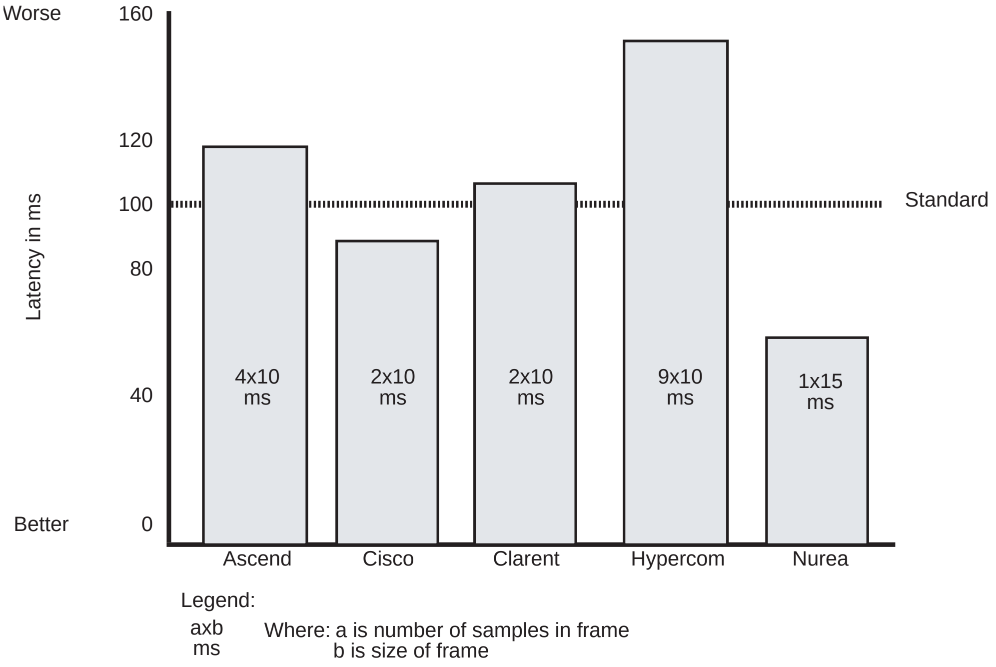
(b) One-way latency tests

Figure 8–8 Voice quality and latency [MIER99]

Then it should come as no surprise that the vendors (with some minor exceptions) whose product gave the best voice quality also exhibited the lowest latency. The number and size of the voice samples in each packet are shown in Figure 8–9.

From these tests, it seems prudent to construct a small packet if one is interested in voice quality. Indeed it is, but there is a price to pay. As stated in the introductory part of the chapter, the packet that contains a small number of samples suffers from more overhead than a packet that contains more samples. This is so, because number of bytes in the associated headers (protocol control information) remain the same for a small or larger user payload—the IP header is always 20 bytes, etc.

Mier's study corroborates these claims, as shown in Figure 8–10. The bandwidth consumption was tested on the LAN and a simulated WAN. Statistics were gathered on the LAN by an HP Internet Advisor, and on the WAN by an Adtech SX-12. In addition, on the WAN, the Adtech gradually reduced the amount of bandwidth available (to the point where the quality was unacceptable). As expected, the systems that



Note: Ascend also had a 2 x 10 packing arrangement

Figure 8–9 Latency and frame size [MIER99]

Figure 8–10 Bandwidth consumption [MIER99]

had larger packet sizes and more latency also had a more efficient use of bandwidth.

As stated in the first part of this chapter, there is a tradeoff of latency, packet size, and perceived voice quality. If bandwidth is plentiful, one would opt for the system that provides higher quality, and so on.

In addition, these vendors' gateways are configurable. Frame packing (number of samples per frame) can be set on most of them, as can the jitter buffer. Many other factors come into play in evaluating VoIP gateways, such as costs and ease of use (configuration). So, it is best to delve into more detail about a vendor's product, but the Mier studies are excellent places to start.

PERFORMANCE OF VoIP IN PUBLIC SYSTEMS

This part of the chapter examines a test conducted by 3Com on the performance of VoIP in the public Internet. The source for this study is [COX98].

The study entailed sending and receiving traffic between three nodes: University of California, Davis; University of Illinois, Chicago;

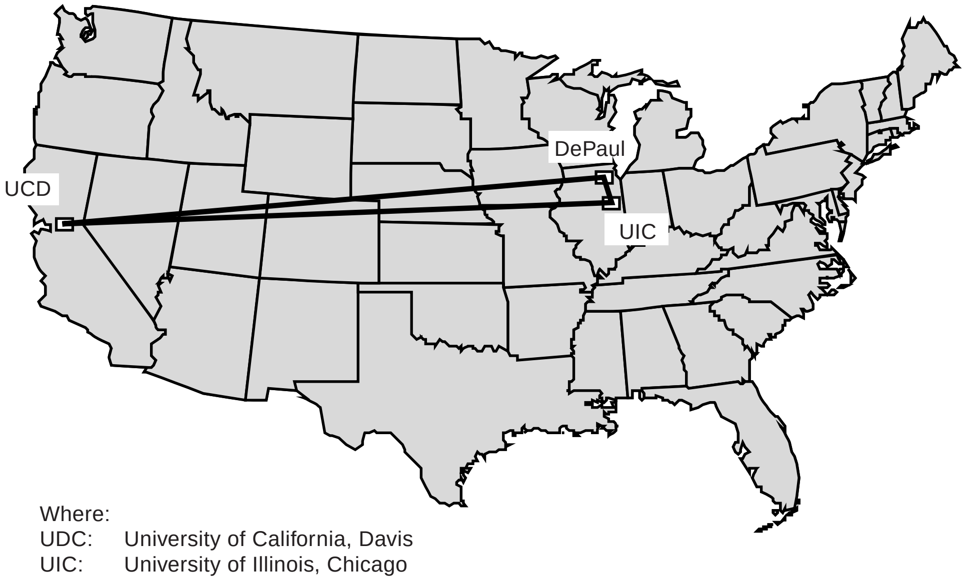


Figure 8–11 Topology for the study [KOST98][3]

and DePaul University, see Figure 8–11. The tests were run for a six-month period. During these tests, a client would transmit once per hour to a server for three minutes. The transmissions involved a trace which allowed the analysts to judge RTT as well as packet loss. The engineers designed their own "ping" program and did not rely on internet pings in order to control how routers handle conventional ping packets. Observations were made in the evenings as well as various times during the business day, and on weekends.

In addition, different codecs were employed in the tests with the emphasis on G.723.1 and G.729A. Tests were done using PC-to-PC communications and VoIP gateways.

We can summarize several key aspects of the study with the examination of the three figures depicted in Figure 8–12. Figure 8–12 (a) compares the average RTT (in ms) in relation to the hop count. The hop count represents the number of nodes traversed between the client and the server. This figure reveals some interesting facts. The first fact is that RTT exceeds 200 ms. This conclusion is borne out by other studies.

---

[3] [KOST98] Kostas, T. J., Borella, M. S., Sidhu, I., Shuster, G. M., Grabiec, J., and Mahler, J. of a 3Com, paper "Real-Time Voice Over Packet Switched Networks, *IEEE Network,* January, February 1998.

The second fact is that the delay is highly variable. On occasion a delay going through the same number of nodes is, say, 100 ms and on another occasion it may be 200 ms.

Keep in mind that these data represent RTT measured with a ping only and does not include analog-to-digital conversion, codec operations, or other factors that would increase RTT.

Figure 8–12(b) is the same figure as that of 8–12(a). The oval is placed on the figure to emphasize that despite the variability in the RTT in relation to hop count, increased hop counts do indeed contribute to delay.

Figure 8–12(c) has some numbers placed around several of the points in the graph. These numbers represent the geographical distances in miles between the sender and receiver of the trace. It is clear that geographical distance cannot be correlated to RTT. Indeed, a short distance of only 477 miles with a hop count of 21 resulted in a 24-ms RTT. Therefore to emphasize, hop distance is a key factor in delay, and geographical distance is less a factor.

Figure 8–13 provides a summary of the 3COM study. Based on the two alternatives of (a) use of the telephone and VoIP gateways, and (b) use of PCs and routers, the study reveals that the telephone/gateway approach provides significantly better performance than the PC/router approach.

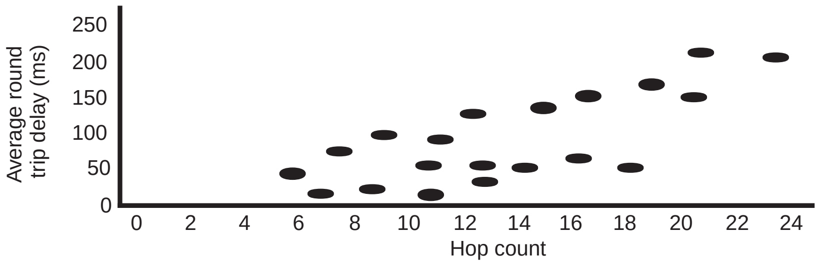
Under ideal conditions, option (a) meets the RTT established by the ITU-T specifications. Option (b) does not meet these requirements, but comes close. During less-than-ideal conditions, neither approach meets the requirements, but a large segment of the population would likely find the performance acceptable for option (a). Option (b) pushes the envelope of acceptable quality.

So, is Internet telephony feasible? Yes, but under the present Internet environment, it may not be acceptable to some people, and may be acceptable to others.

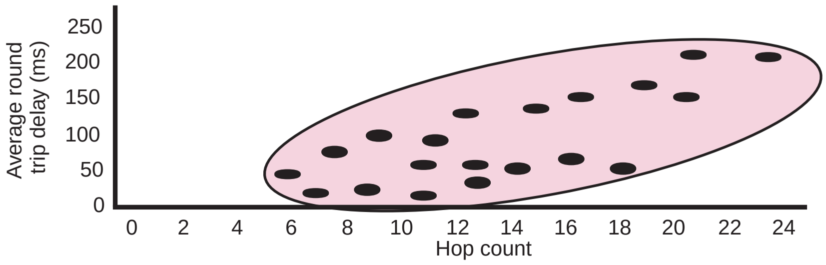
Nonetheless, given the attractive features of Internet telephony (one link to the home, integrating voice and data, and low costs), it will surely succeed.

One other aspect of this subject bears examination: the deployment of high-speed (a) ADSL modems, (b) cable modems, or (c) fixed wireless access technologies on the local loop. Once the customer has these technologies available, the equation changes.

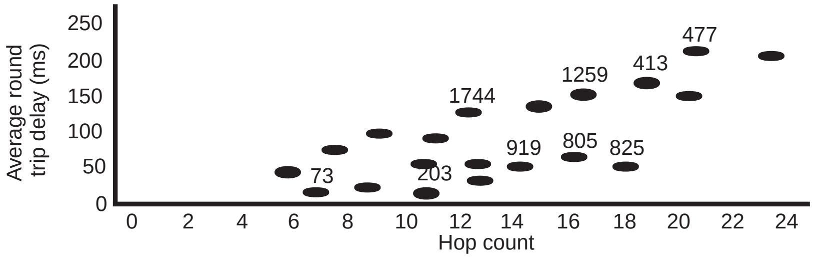
First, overhead (headers and trailers) is not as significant a factor, since the increased bandwidth can support this overhead. Second, new PCs will be upgraded to support faster voice coders to take advantage of



(a) Average delay and hop count



(b) Correlation of delay to distance



(c) Relationship of delay and geographical distance

(note: numbers represent distance in miles)

Figure 8–12 Round trip delay versus hop count [KOST98]

the higher-speed local loop. Third, the increased pipes into and out of the Internet will force an upgrading of the Internet's capacity. Fourth, the increase of voice (and video) traffic will also force the Internet to look more and more like the telephone network, but with significantly enhanced multiapplication capabilities.

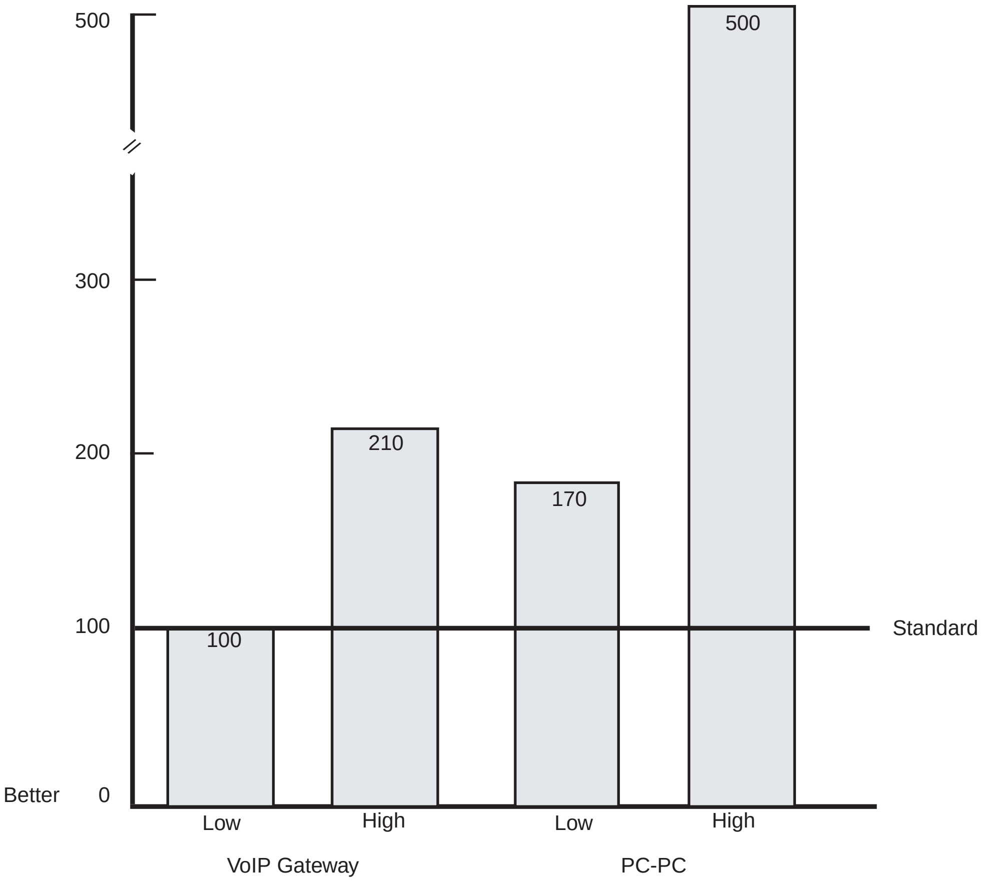


Figure 8–13 Unidirectional delays [KOST98]

SUMMARY

The Mier and Cox studies demonstrate that VoIP is quite feasible in private networks with leased lines, but marginal in the public Internet. It is a good idea to remember that packet voice is in its infancy, and the test results from Mier's labs are impressive. It is reasonable to expect that the Internet will also deliver toll-quality speech—in a few years.