Performance Evaluation of Internet Telephony Systems Through Quantitative Assessment

*C.H. Ng, S. Foo, S. C. Hui
School of Applied Science, Nanyang Technological University, Singapore

ABSTRACT

An Internet telephony system functions like a conventional telephone to support real-time voice communication over the Internet. The performance of such a system depends upon a number of factors which include network performance, computing resource, voice processor, signal acquisition, output transducer, and method used to establish a connection between users. A performance evaluation using quantitative assessment techniques have been conducted for four Internet telephony systems, namely, CoolTalk, Net Meeting, Internet Phone, and Speak Freely. These were used to assess the factors of network performance, voice processor and connection method.

INTRODUCTION

Recently, as a result of declining costs of computer hardware, advances in computer technology and phenomenal growth in Internet, a number of research prototypes and commercial products of Internet telephone systems have been developed to offer real-time voice communications and other value added services over the Internet. The growing enthusiasm stems mainly from huge potential cost savings by making it possible to make transcontinental telephone calls at the prices of local telephone calls plus nominal standard Internet connectivity charges.

![Diagram of Internet Telephony System](image-url)
Figure 1 shows the basic components of an Internet telephone system. Two host computers acting as caller and recipient are required. In using the standard Internet Transmission Control Protocol/Internet Protocol (TCP/IP), each host computer is identified by a unique IP address. The host computer can either be a workstation or a personal computer with sufficient computation power and audio capabilities. The telephone software system which resides on each host computer facilitates the real-time voice communication across the Internet. In the basic communication process, the caller’s software system will acquire the real-time voice data through an audio input device and convert the analogue signals into digitised form which is then compressed and optionally encrypted before being transmitted to the recipient through the Internet using the TCP/IP protocol. Compression is necessary to reduce the bandwidth requirement of the voice data. At the recipient’s end, the software system carries out the reverse process. Incoming data is first decrypted, decompressed and played back in real-time on the audio device of recipient’s computer.

There are many factors that affect the performance of an Internet telephony system. However, it is also not easy to establish a set of performance measurements that can be objectively determined. A number of factors have been identified that affect the performance significantly:

- **Network performance.** An Internet Telephony system will not be able to perform satisfactorily if the network does not meet the requirements of the system. High rate of packet lost will cause the system to be unable to recover the original signal, and long delay will lead to unpleasant conversation.

- **Computing resource.** A fast enough signal processor must be available to process the voice signals at real-time so that data will not overrun. Most of the Internet telephony products state a minimum requirement of an Intel 486/60Mhz processor or equivalent. A number of products utilise Digital Signal Processors for better performance.

- **Voice processor.** One of the factors that makes an Internet telephony system outshine another is the engine for voice processing. A good voice processor will allow higher compression ratios, use less computing resource, least affected by data packets lost, and reproduce quality signals at the receiving end. There are currently two major systems of coding, namely, waveform coders and speech coders.

- **Signal acquisition.** Voice signals are normally captured using microphones and digitised with Digital-Analogue-Converters. The quality of captured information depends on the rate of sampling and resolution of data. However, having a higher sampling frequency or higher resolution of data, will increase the size of information to be transmitted. In general, a configuration with sampling frequency of 8 kHz and 8 bit conversion is suitable for real-time voice communication.

- **Output transducer.** A good output transducer is generally required in order to minimise noise in the output signals.

- **Methods to establish connection between users.** Knowledge of the IP addresses of the two parties is necessary before a connection can take place. This poses no problems for users with static IP addresses. However, for users having dynamic IP addresses (for cases of Internet connections via an Internet Service Provider (ISP)), there is a need to publish and exchange
information of their current IP addresses through some means, such as servers or other gateways. The ability of the Internet telephony system to establish connection between users quickly becomes another performance issue.

PERFORMANCE EVALUATION

A performance evaluation of a number of Internet telephony systems has been carried out using both quantitative and qualitative methods. The reader should refer to (Lee 1997) for details of the qualitative methods used and the results of the tests. In terms of quantitative tests, these were evaluated for three factors: network performance, voice processor and methods to establish connection.

Since we are using a same standard set of equipment for performance evaluation, the factors of computing resource, signal acquisition and output transducer are kept constant and therefore will not affect the results of the tests. In addition, computing resource used is assumed to satisfy at least the minimum system requirements of all the systems tested.

Four Internet telephony systems, namely, Internet Phone Release 4 (Vocaltec Inc. 1996), NetMeeting 2.0 Beta (Microsoft Corp. 1996), CoolTalk 1.0 (Netscape Communication Corp. 1996), Speak Freely (Walker 1996), were used for conducting these tests. They were selected on the basis of common system requirements and its availability on the Internet. All tests were conducted in a controlled and similar environment using the same equipment to eliminate inconsistencies and other factors. The equipment used include:

- 2 x Pentium 100MHz PC with 32MB RAM running Microsoft Window 95
- 2 x Creative Sound Blaster 16 PnP sound card with full-duplex software driver
- 2 x US Robotics Sportster 28.8k modem
- Signal Generator

A. NETWORK PERFORMANCE

UNIX’s ping (Mark, 1996) is a tool used for network testing, measurement and management. It utilises the Internet Control Message Protocol's (ICMP) ECHO_REQUEST datagram to elicit an ICMP ECHO_RESPONSE from a host or gateway. In carrying out the tests, 20 packets of 1 kilobyte size were transmitted to emulate the transmission of audio packets across the Internet. Network performance was gauged by computing the average round-trip delays and percentage of packets lost. Network performance must be a controlled factor for the tests to be accurate. The results of the tests are shown in Table 1.

<table>
<thead>
<tr>
<th>Type of Network</th>
<th>Average Round Trip Delay</th>
<th>Packet Lost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local area network</td>
<td>&lt;1 ms</td>
<td>0%</td>
</tr>
<tr>
<td>ISP - ISP</td>
<td>0.5 sec - 1.5 sec</td>
<td>0 - 40%</td>
</tr>
<tr>
<td>Modem - ISP router</td>
<td>1 sec</td>
<td>0%</td>
</tr>
</tbody>
</table>

It can be seen that network performance between ISPs has a great impact on the performance of Internet telephony systems. Our tests have indicated that a network with no more than 20% packets lost and 1 second delay must be attained in order to make it acceptable for human hearing perception.
Packet lost of more than 20% results in unacceptable forms of communication regardless of system used. However, as network performance is an external factor to the systems being tested, it cannot be used to differentiate performance among systems. It does, however, indicate that a local area network must be utilised to carry out all other tests in order to isolate network fluctuations which might otherwise affect the test results.

B. VOICE PROCESSOR

In order to investigate the efficiency of the voice processor of system been tested, 300 Hz and 600 Hz square waves were transmitted across under near ideal network environment (via a Local Area Network). The signals at the receiving ends were observed and compared with the original waveform on the oscilloscope. The lower and the upper cut-off frequencies were also investigated since they determine the quality of the playback audio signal. Fast Fourier Transform (FFT) was used to analyse the source of noise to determine whether it was induced during the stage of signal acquisition or generated by the voice processors.

Square waves were used for the test since they can be observed and compared easily. Noise signals riding on the square waves can also be noticed. Using Fourier analysis (Ramirez 1985), square waves can be decomposed into fundamental frequency and its harmonics as shown in Equation 1 or approximated using power series of Equation 2.

\[ x(t) = A_0 + \sum_{n=1}^{\infty} \left[ A_n \cos(n \cdot t) + B_n \sin(n \cdot t) \right] \]

where

\[ A_n = \frac{1}{\pi} \int_{-\pi}^{\pi} x(t) \cos(n \cdot t) \, dt \]

\[ B_n = \frac{1}{\pi} \int_{-\pi}^{\pi} x(t) \sin(n \cdot t) \, dt \]

(1)

\[ f(x) = \frac{4}{\pi} \left[ \sin(x) + \frac{\sin(3x)}{3} + \frac{\sin(5x)}{5} + \cdots \right] \]

(2)

However, as there is a limitation of bandwidth, Gibb’s phenomenon (Washington 1995) will be observed as higher frequencies are cut off as shown in Figure 2.
In addition, 300 Hz & 600 Hz square waves were chosen for these tests as they can emulate the range of frequencies of human vocals. The results of these tests are shown in Table 2.

Table 2. Voice Processor Test Observations

<table>
<thead>
<tr>
<th>System</th>
<th>Characteristic of Recovered Signal</th>
<th>Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet Phone</td>
<td>Resembles original wave with Gibb’s phenomenon.</td>
<td>54 Hz - 3.3 kHz</td>
</tr>
<tr>
<td>CoolTalk</td>
<td>Resembles original wave with Gibb’s phenomenon.</td>
<td>20 Hz - 3.9 kHz</td>
</tr>
<tr>
<td>Net Meeting</td>
<td>There is a 12ms signal loss for every 200ms of speech transmission. One of the possible reasons is inefficiency of voice processing that results in data over run.</td>
<td>70 Hz - 3.7 kHz</td>
</tr>
<tr>
<td>Speak Freely</td>
<td>Simple compression. Resembles original wave with Gibb’s phenomenon.</td>
<td>30 Hz - 2kHz</td>
</tr>
<tr>
<td>(Simple compression)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speak Freely</td>
<td>Resembles original wave with Gibb’s phenomenon.</td>
<td>30 Hz - 3.4kHz</td>
</tr>
<tr>
<td>(GSM compression)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speak Freely</td>
<td>Resembles original wave with Gibb’s phenomenon.</td>
<td>30 Hz - 3.5 kHz</td>
</tr>
<tr>
<td>(ADPCM compression)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From this series of tests, we can conclude that almost all systems, except Net Meeting, were able to reproduce the original square wave with noticeable Gibb’s phenomenon been detected. All systems were tested similarly under ideal network environment. In the test, CoolTalk was observed to have the least signal distortion and offers a wider bandwidth than other systems. This indicates CoolTalk’s capability to offer better audio signal recovery.

Speech coders were attempted for testing, but the resulting waveforms were beyond comparison with original signals. Thus, this method of testing can only be applied to waveform coders, as speech coders do not attempt to regenerate the original signals. The performance of speech coders can only be measured subjectively.

C. METHOD TO ESTABLISH CONNECTION

From the user’s point of view, the efficiency of establishing a connection between two parties basically translates into how fast the system can resolve the IP addresses. In this instance, a simple stop watch for measuring the time involved would suffice. Other factors such as resource overheads (e.g. hard disk requirements) and ease of use are not considered in this evaluation.
Net Meeting & CoolTalk uses their **own user directory servers** for the resolution of dynamic IP addresses. Upon startup, the system will register the user’s identification and IP address with the directory servers. The information in these directory servers are downloaded into the local computer system which uses a query and retrieval operation to obtain the IP address of the recipient. It is found that on average, it takes less than five minutes to download a typical directory server comprising 1000 users under peak network conditions. This time reduces accordingly as the network conditions improve.

In contrast, Speak Freely uses **public user directory servers** to store the users’ information. The time required for downloading the users’ information is similar to those using user directory servers. However, once the user directory is downloaded, it takes almost no time to resolve the IP address of the other party. This method of connection is slow and only reflect a static snapshot of the information at the point of download. Users that join or leave the system are not reflected subsequently. A complete download is necessary to get the most-to-date list of active users.

Internet Phone uses the concept of **Inter Relay Chat (IRC) servers** (Oikarinen 1993) which can be linked with one another. Users are not required to download the entire listing but can join channels directly. Under normal conditions, this process takes less than 1 second although it can span up to 10 seconds under adverse network conditions. Once inside these channels, the IP address of the recipient can be resolved almost instantaneously. Another advantage that this mode of connection exhibits over the directory server method is the ability to indicate immediately if the users have logged off.

Another method that can be used for resolving IP address is the **electronic mail method**. This method utilises the concept of fixed email addresses of the recipient and uses an electronic mail to send the IP address of the caller together with a request for connection. The recipient upon receiving the mail and agreeing to the request, will have the necessary information to initiate a connection.

The time stamp in the mail header can be used directly to compute the time delay in using this method for establishing a connection. This time delay will depend on the frequency at which the users’ Standard Mail Transfer Protocol (SMTP) (Postel 1982) server sends mail messages, the traffic of the network, and how fast end users retrieve their electronic mail from the Post Office Protocol (POP) (Rose 1993) servers. It was found that the average respond time can range anywhere between 10 seconds to 2 hours.

Lastly, the **dynamic Domain Name Service** (DNS) (ISOTRO 1996) is a new concept offered by some ISPs for the resolution of dynamic IP addresses. These servers will map the users current IP addresses to their fixed personal domain name. This approach is universal and is compatible with any Internet telephony system. A caller only needs to specify the recipient’s personal domain name for connection establishment. It takes almost no time (less than 1 second) to resolve the IP addresses of the recipient using this method.

From these various methods, it becomes evident that the dynamic DNS is the best method to establish a connection between users. It is fast and easy to use and will probably be the standard technique of the future.

**CONCLUSION**
There are many Internet telephony systems available, each with its own system requirements, with some systems using special hardware for processing. In order to perform evaluation for different systems, there is a need for uniformity, a standard platform for testing, and a need to identify and isolate external factors that can deteriorate or influence the systems’ performance. Only then it becomes possible to conduct quantitative tests on the systems. This paper has presented a series of performance evaluation tests to determine network performance, quality of voice processor and method to establish a connection. These tests results should not be used a stand-alone set of data to rank the systems, but must be combined with a set of qualitative tests (Lee 1997) to finally ascertain the performance of the systems being evaluated. As such, the reader should refer to (Lee 1997) for the final ranking of systems. With the combination of quantitative and qualitative testing, it becomes possible to derive a framework to carry out a proper performance evaluation of an Internet telephony system.

REFERENCES

Mark G. (1996), Ping using ICMP, <URL: http://users.redrose.net/~markg/ieping.html>
Microsoft Corp. (1996), NetMeeting 2.0beta <URL: http://www.microsoft.com/>
Netscape Communications Corp. (1996), CoolTalk 1.0 <URL: http://www.netscape.com/>
Postel, B.J. (1981), Internet Control Message Protocol, RFC 792.
Postel, B.J. (1982), Simple Mail Transfer Protocol, RFC 821.
VocalTec Inc. (1996), Internet Phone 4.0 <URL: http://www.vocaltec.com/>
Washington State University. (1995), Gibb's Phenomenon <URL: http://www.eecs.wsu.edu/~cs445/wwwbook/mma_nbs/Fourier2/Fourier2_5_0_0.html>