Performance Evaluation of VoIP in Different Settings

Tom Christiansen, Ioannis Giotis and Shobhit Mathur

{tomchr,giotis,shobhit}@cs.washington.edu

Abstract

The internet is fast evolving into a universal communication network and it is contemplated that soon it will carry all types of traffic, including voice and video along with data. Among them, telephony is an application of great importance, particularly because of the significant revenue it can generate. **V**oice **o**ver **I**nternet **P**rotocol is a technology that allows telephone calls to be made over computer networks like the Internet. There are several implementations of VoIP in the internet today. Many major telephone companies like AT&T[1] have moved over to VoIP completely. It is still not clear how the performance varies with different network conditions, considering that the Internet does not provide QoS guarantees. In this project we try to answer the following questions:

- How does VoIP work in the Internet today?
- How is VoIP quality measured?
- What are factors which affect the performance of VoIP?
- How does the quality of VoIP depend on these factors? Can we explicitly quantify this?
- Can we improve the existing VoIP technology? If so, how?

We ran several hundred experiments over a period of two weeks. Based on the data obtained, we looked for answers to the above questions. Answering these questions helped us understand the VoIP technology better and gave us an insight into how it can be improved in future.

1 Introduction

VoIP is a revolutionary technology that has the potential to completely rework the world's phone systems. VoIP providers like Vonage [2] have already been around for a while and are growing steadily. Major carriers, like AT&T[1], have set up VoIP calling plans in several markets around the United States, and the FCC is looking seriously at the potential ramifications of VoIP service. A major advantage of VoIP and Internet telephony is that it avoids the tolls charged by ordinary telephone service. As the internet evolves into a ubiquitous communication infrastructure and provides various services including telephony, it will be expected to meet the quality standards achieved in the public switched telephone network (PSTN).

VoIP sends voice information in digital form in discrete packets rather than in the traditional circuit-committed protocols of the public switched telephone network. It converts analog voice signals into digital data packets and supports real-time, two-way transmission of conversations using Internet Protocol (IP). The voice blocks are encapsulated in a sequence of voice packets using the Real-time Transport Protocol (RTP) and delivered by the User Datagram Protocol (UDP). The RTP header contains timing information and a sequence number that allow the receivers to reconstruct the timing produced by the source. Note that RTP itself does not provide any mechanism to ensure timely delivery or provide other QoS guarantees, but relies on lower-layer services to do so. It does not guarantee delivery or prevent out-of-order delivery, nor does it assume that the underlying network is reliable. To help VoIP applications deal with unpredictable network performance, the Real-time Transport Control Protocol (RTCP) is developed to monitor the performance of RTP packets and provide feedback to the VoIP applications. The feedback on packet delay, jitter, and loss rate enables the applications to adapt to network conditions to maintain a certain level of voice quality.

In this report, we first look at the codecs used in VoIP in Section 2. Codecs play an important role in determining the quality of a VoIP call. In Section 3 we look at the various VoIP quality measurement metrics. We describe how to measure the overall call quality in Section 4. Having looked at the metrics, Section 5 describes the experimentation methodology, followed by the analysis of the experimental results in Section 6. Section 7 discusses what we achieved from the project and in Section 8 we propose possible future directions.

2 Codecs used in VoIP

CODEC is an acronym for COder/DECoder. A codec is responsible for converting a voice signal into a format suitable for transport and receipt over a network. The codec at the sending end compresses (COdes) the voice signal for transmission over the network. At the receiving end, the codec decompresses (DECodes) the signal for the listener. Codecs vary in the sound quality, the bandwidth required, the computational requirements, etc. Each service, program, phone, gateway, etc typically support several different codecs, and when talking to each other, negotiate which codec they will use.

The general requirement for VoIP codecs is low bandwidth usage. As seen in Figure 1 a typical VoIP codec uses 10-30 kbps data rate. For example, Skype's iLBC [3] uses 13 kbps. Another important requirement is the ability to encode/decode in real time, which is not much of an issue nowadays. Codecs specifically built for VoIP have specific optimizations to encode human voice in a typical speech pattern. Encoding is usually based on a fixed frame size. Larger frames allow for more efficient encoding but introduce larger delays and higher sensitivity to packet loss. Therefore the choice of a good frame size is equally important as the choice of codec. Keep in mind that a codec's bitrate is not equal to the actual IP traffic, as typical VoIP packets are small and the IP and UDP headers impose a significant overhead. Figure 2 illustrates the overhead for common codecs.

- GIPS - 13.3 kbps and up
- GSM - 13 kbps (full rate), 20ms frame size
- iLBC - 15kbps,20ms frame size: 13.3 kbps, 30ms frame size
- ITU G.711 - 64 kbps, sample-based. Also known as A-law/μ-law PCM
- ITU G.722 - 48/56/64 kbps
- ITU G.723.1 - 5.3/6.3 kbps, 30ms frame size
- ITU G.726 - 16/24/32/40 kbps
- ITU G.728 - 16 kbps
- ITU G.729 - 8 kbps, 10ms frame size
- Speex - 2.15 to 44.2 kbps
- LPC10 - 2.5 kbps
- DoD CELP - 4.8 kbps

Figure 1: Most common VoIP codecs and their bitrates [4]

Codec	Bitrate	IP bitrate
G.711	64 kbps	87.2 kbps
G.729	8 kbps	31.2 kbps
G.723.1	6.4 kbps	21.9 kbps
G.723.1	5.3 kbps	20.8 kbps
G.726	32 kbps	55.2 kbps
G.726	24 kbps	47.2 kbps
G.728	16 kbps	31.5 kbps
iLBC	15 kbps	27.7 kbps

Figure 2: Actual IP bitrate used by various codecs [4]

The presence of many different codecs (see Figure 1) showcases that there isn't an obvious choice, but variable network setups should be handled with different codecs or bitrates. A valuable feature of many commercially available VoIP products is the ability to switch bitrates while a conversation is in progress, delivering better quality on-the-fly.

Each codec provides a certain quality of speech. The quality of transmitted speech is a subjective response of the listener. A common benchmark used to determine the quality of sound produced by specific codecs is the Mean Opinion Score (MOS) [5] .

One interesting point of research is alternative encoding codecs. By analyzing human speech, it is concluded that speech can be synthesized using a relatively small (less than 8-10 taps) Infinite Impulse Response (IIR) vocal tract filter excited with either a quasi periodic impulse train or white noise [12]. To transmit synthesized speech through a network only the filter coefficients and a gain factor are transmitted. The resulting data stream will require a bandwidth of 2.0 to 2.7 Kbit/s (assuming 8-bit IIR coefficients). This provides a significant reduction in required bandwidth compared to currently used VoIP protocols. However, implementing a voice synthesizing codec to test the sound quality is beyond the scope of this project.

3 VoIP Quality Metrics

There are three important measures of VoIP quality: (1)Signaling quality, (2)Delivery quality, and (3)Call quality. In this section we discuss how to measure each of them.

(1) Signaling Quality: It is a measure of the call setup performance. Before a conversation can begin, a call must be setup. Both sides need to be able to find and reach one another, consent to talk, and agree how the call is to proceed. Call setup is a complicated process. Over the years, quite a few signaling (or setup) protocols have been developed. SIP[6], or the Session Initiation Protocol, is currently popular and used in many VoIP services. SIP is defined in RFC 3261. The metrics for measuring the signaling quality are given below. They refer to the time it takes to accomplish the various stages of setting up a call. Assume A initiates the call and the other end is B. There are 2 sets of metrics, one for each direction of the conversation. Basic knowledge of the SIP messages is assumed.

From A to B:

- **Post-Dial Delay** is the time it takes, after A sends the INVITE message for the phone at B to ring.
- **Call Setup Delay** is the full time it takes after A sends the INVITE message for it to receive the 200 OK response from B. Call setup delay includes post-dial delay.
- **Media Delay** includes the full call setup time plus the time it takes for B to receive the first packet of media (conversation). Media Delay includes both call setup delay and post-dial delay.

From B to A:

- **Post-Pickup Delay** is the time that elapses between B sending the 200 OK response and receiving the first packet of media (conversation).
- **Call Setup Delay** is the time elapsed between the INVITE and the ACK messages that B receives from A. The ACK message from A to B confirms that the call has been successfully set up.
- **Media Delay** is the time between receiving the initial INVITE request to receiving the first media packet (conversation). Media Delay includes both the call setup delay and post-pickup delay.

(2) Delivery Quality : Once a call has successfully been setup, latency, jitter, and packet loss effects are important predictors of the call stream performance also called as the delivery quality.

- **Latency** : A measure of the delay in a call. The largest contributor to latency is caused by network transmission delay. With round trip latencies above 350 ms users may experience annoying talk-over effects.
- **Jitter** : Jitter refers to how variable latency is in a network. High jitter, greater than approximately 50 ms, can result in both increased latency and packet loss. Jitter causes packets to arrive at their destination with different timing and possibly in a different order than they were sent (spoken), with some arriving faster and some slower than they should. To correct the effects of jitter, VoIP endpoints collect packets in a buffer and reorder them according to their timing and sequence number before the listener hears them. This works, but it is a balancing act. Processing that buffer adds delay to the call, so the bigger the buffer, the longer the delay. If voice packets arrive when the buffer is full then packets are dropped and the receiver will never hear them. These are called *discarded packets*.
- **Packet Loss** : Some of the voice packets may be dropped by network routers or switches that become congested, such packets are called *lost packets*. Knowing the average packet loss rate for a call gives an overall sense for the quality of the call. Additional information is needed to determine whether the loss was 'random' or 'bursty' to infer the call quality.

(3) Call Quality : The perceived quality of speech is highly subjective. However, several methods for objective speech quality analysis exist. One example is the E-model [7], which is an end-to-end model originally intended for measuring speech quality of telephone networks. The result of an E-model calculation is an R-value which is related to the mean opinion score (MOS). The E-model and MOS are described in more detail in Section 4.

MOS	Quality	Listening Effort
5	Excellent	Complete Relaxation
4	Good	Attention Necessary
3	Fair	Moderate Effort
2	Poor	Considerable Effort
1	Bad	No Meaning Understood

Figure 3: Understanding the MOS scores

4 Measuring Speech Quality

The fundamental concern for VoIP QoS is voice quality. Unfortunately, objective measurements for this have been elusive. Modern communications networks include elements (bad coding, error-prone channels and voice activity detection) that cannot reliably be assessed by such conventional engineering metrics as signal-to-noise ratio. One way to measure customers' perception of the quality of these systems is to conduct a subjective test involving panels of human subjects. However, these tests are expensive and unsuitable for such applications as real-time monitoring. The Mean Opinion Score (MOS) is a subjective number indicating how people feel about the quality of the voice signal. MOS is measured on a scale from 1-5 where 1 is the lowest and 5 the highest. Figure 3 shows the correlation between MOS scores and the listening quality.

In this project we use an online tool [8], to measure the quality metrics described in Section 3. This tool uses the ITU G.107 E-Model. E-Model produces an R-factor score which can be directly correlated to a MOS score. The R-value is calculated as the sum of five variables:

$$R = R_o - I_s - I_d - I_{e-eff} + A,$$

where R_o is the base signal-to-noise ratio (SNR), I_s represents the signal impairment introduced simultaneously with the voice transmission (e.g. quantization noise), I_d is the im-

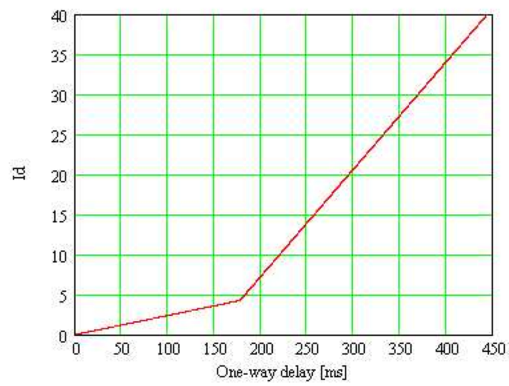

Figure 4: Relation between I_d and delay

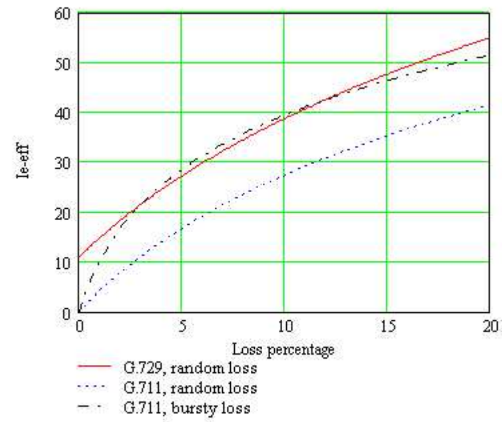

Figure 5: Relation between I_{e-eff} and packet loss percentage

pairment caused by delay (e.g. talker echo, listener echo), I_{e-eff} is the effective equipment degradation factor (e.g. codec loss), and A is an advantage factor. The relation between I_d and delay is shown in Figure 4 and Figure 5 shows how I_{e-eff} varies with the percentage of packets lost.

We see that once the one-way delay is increased above 177 ms, the delay impairment on the speech quality is quite severe. For VoIP applications, this is actually a fairly stringent requirement as IP traffic often experiences significant delays as a result of network congestion, link failures, and just general routing schemes. Also note that the use of jitter buffers and the encoding itself introduces delay (10-25 ms is typical).

The advantage factor, A, is an attempt to quantify the user expectations for a particular communication medium. For example, cell phone users may rate a certain quality level as being of acceptable quality, whereas, for a landline telephone, the same level of quality would be rated as low quality solely because of user expectations of the particular phone service. In other words, cell phone users would be willing to trade some speech quality for the convenience of mobility. Similarly, VoIP users would likely be willing to trade voice quality for a lower price - in some cases even free service. Figure 6 shows the network parameters which influence the R factor. The user satisfaction as a function of R-value and MOS is shown in Figure 7.

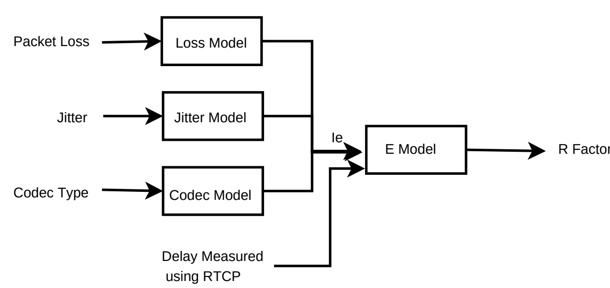

Figure 6: Network factors which influence the R factor

While the theoretical MOS scale tops out at 5.0, practically speaking, a 5.0 score is not achievable regardless of the network connection quality. That is because VoIP codecs introduce some amount of quality loss. For example, the maximum MOS score achievable with G.711 codec is 4.4 and with G.729 codec is 4.2.

R-value (lower limit)	MOS (lower limit)	User satisfaction
90	4.34	Very satisfied.
80	4.03	Satisfied.
70	3.60	Some users dissatisfied.
60	3.10	Many users dissatisfied.
50	2.58	Nearly all users dissatisfied.

Figure 7: User satisfaction as a function of R-value and MOS

5 Experimentation Methodology

The experiments were run over a period of two weeks and at various times of the day. A total of about 600 experiments were run from Seattle. No artificial load conditions (e.g background downloads) were added to the network. The intention was to test VoIP under *normal* load conditions. These experiments were equally divided among four types of networks: (1) Dial-up from home (2) 802.11 b/g wireless network at the UW CSE department (3) Comcast Cable Network (4) UW CSE department switched gigabit ethernet network.

In each of the networks above, calls were made to the following destinations: (1) Boston (2) Helsinki (3) London (4) Montreal (5) San Jose (6) Sydney. Geographically San Jose, Montreal and Boston as closer while Helsinki, London and Sydney are far off from the source location i.e Seattle.

We tested two codecs which are popularly used in VoIP: (1) G.711 (PCM at 64 kbps, 20 ms RTP payload, 80 kbps IP bandwidth) (2) G.729 (CS-ACELP at 8 kbps, 10 ms RTP payload, 40 kbps IP bandwidth). G.729 trades off lower IP bandwidth for voice quality. It uses predictive coding and halves the IP bandwidth.

The following parameters were obtained from each experiment (in both directions): (1) MOS (2) Degradation due to the codec used, latency, packet discards, and packet loss (3) Loss periods (4) Jitter (5) Signaling characteristics: Post-Dial delay, Post-Pickup delay, Call setup time, Media delay (6) Type of packet loss: Burst or Random

6 Analysis of the Experimental Results

In this section the results obtained from the experiments are analyzed. Due to space constraints not all the results are shown. In this report the interesting results are plotted and other results are described in words. We mainly concentrate on the MOS from source to the target destination using G.711 codec. The results for the opposite direction and using G.729 codec are similar. Recall that the maximum MOS that can be achieved with a G.711 codec is 4.4 and with a G.729 codec is 4.2. All results are averaged over the number of experiments run with that setting. We concentrate on the delivery quality (jitter, latency, packet loss and packet discards) and the call quality (MOS). Signaling quality is not very important as it is the characteristic of the signaling protocol used (SIP) and does not affect the speech quality . The graphs are all pie charts, to help the reader visualize the negative contribution of each of the degradation factors (codec, latency, packet discards, packet loss) to the ideal MOS score of 5.0.

6.1 Home Dialup Network

One would expect that VoIP would not perform well over dialup networks. On the contrary, VoIP delivers a fairly good performance over dialup networks. This can be attributed to the fact

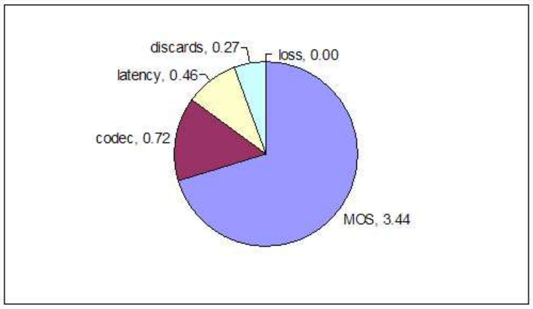

Figure 8: Overall MOS to any destination using G.711 over a dialup network

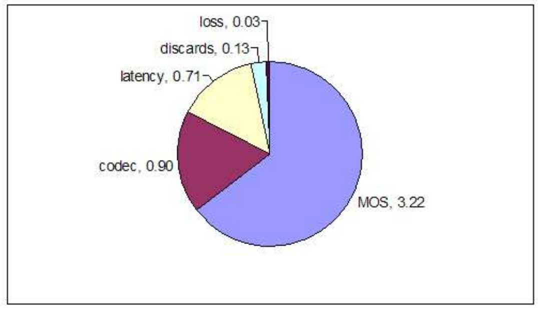

Figure 9: Overall MOS to any destination using G.729 over a dialup network

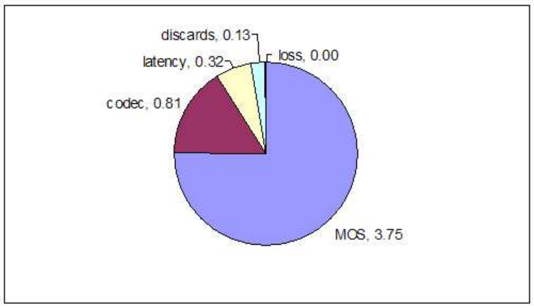

Figure 10: MOS to San Jose using G.711 over a dialup network

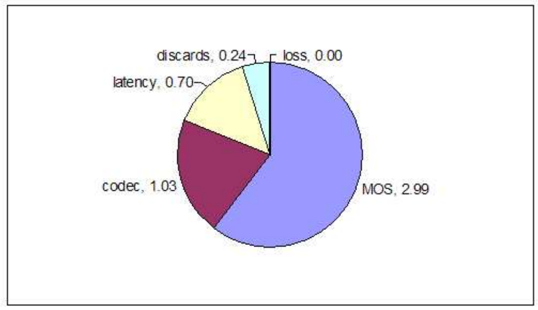

Figure 11: MOS to Helsinki Using G.711 over a dialup network

that VoIP codecs have a minimum bandwidth requirement, which dialup networks can satisfy to a large extent. 24 kbps to 32 kbps is the average bandwidth provided by dialup networks. The G.711 codec tested has a bandwidth requirement of 80 kbps, while G.729 has 40 kbps requirement. If the entire bandwidth was available, these codecs would deliver their best performance, i.e highest MOS. We see in Figures 8 and 9 that even though the bandwidth requirement is not satisfied, the MOS delivered by using the G.711 codec is 3.44 and by using G.729 is 3.22. Even more surprising is the fact the quality degradation is due to discarded packets rather than the lost packets. Hence a dialup network delivers all packets, but the packets do not arrive in time at the destination. Next, let us look at the individual destinations. The MOS to San Jose turns out to be the best, while it is the worst to Helsinki. This can be seen in Figures 10 and 11. The degradation of MOS to Helsinki is mainly due to the latency, jitter (resulting in discarded packets) and lost packets. Our experiments rank the overall MOS (considering both directions) to the destinations as San Jose > Sydney > Boston, London > Montreal > Helsinki. As we can see, geographical distance plays a vital role in determining the end-to-end voice quality in dialup networks.

6.2 802.11 b/g Wireless Network

The wireless network tested also performs fairly well and is suitable for VoIP applications. The average MOS to target locations using the G.711 and G.729 codecs was 3.06 and 3.20 respectively. This is shown in Figures 12 and 13. The voice quality using the G.729 is comparable to the quality delivered by a dialup connection. The quality of the G.711 codec is substantially lower. In Figure 12 notice that the contribution of discarded packets to degradation in voice quality is

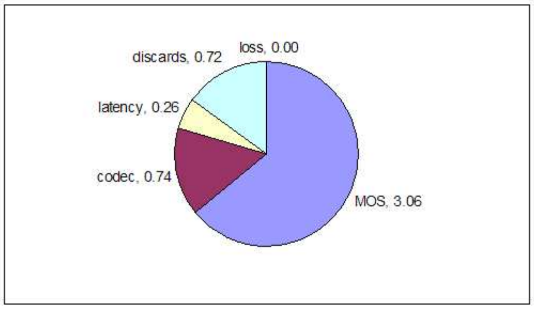

Figure 12: Overall MOS to any destination using G.711 over a wireless network

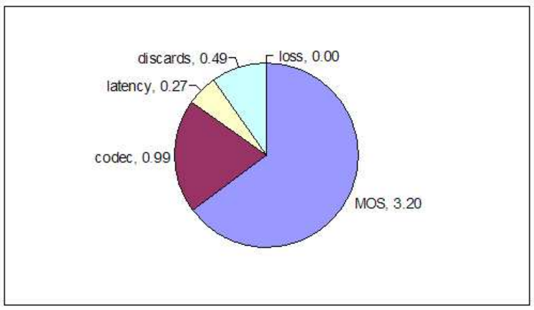

Figure 13: Overall MOS to any destination using G.729 over a wireless network

almost three times that of dialup and that latency is almost half. The discarded packets are due to the jitter experienced by the packets. The maximum jitter observed by packets in the wireless network when using the G.711 codec was observed to be 473 ms while it was just 129 ms in the dialup network. We can conclude that wireless networks are more prone to jitter resulting in discarded packets and consequently lower voice quality. It is interesting to note that this phenomenon is observed even when G.729 codec is used. Hence, jitter experienced in wireless networks is not related to the throughput. Another possible reason for the large number of discarded packets in wireless networks could be transmission errors (i.e collisions resulting in corrupted packets). The latency in a wireless network is almost half that observed in a dialup network, due to the difference in available bandwidths. A dialup network promises a maximum of 56 kbps while the wireless network tested had a maximum capacity of 54 Mbps (802.11b/g). Hence even though the minimum bandwidth requirements of the codecs are satisfied in a wireless network, the jitter plays a vital role. The same phenomenon is observed in experiments to each of the target locations i.e packet discards are the primary cause of voice quality degradation.

6.3 Cable Network

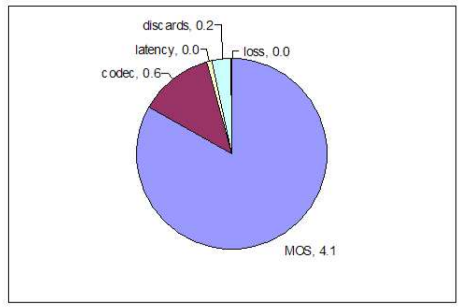

Figure 14: Overall MOS to any destination using G.711 over a cable network

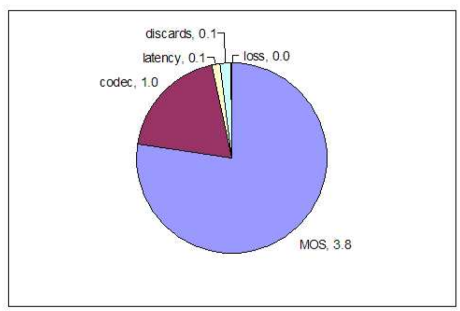

Figure 15: Overall MOS to any destination using G.729 over a cable network

Cable network delivers much better voice quality than dial up and wireless networks. This can be seen in Figures 14 and 15. The average bandwidth available over a cable network is approximately 3 Mbps (downlink) and 300 kbps (uplink). The effects of latency, packet loss and packet discards are almost negligible. The degradation in voice quality is only due to the codec used. This is true for all target locations. The

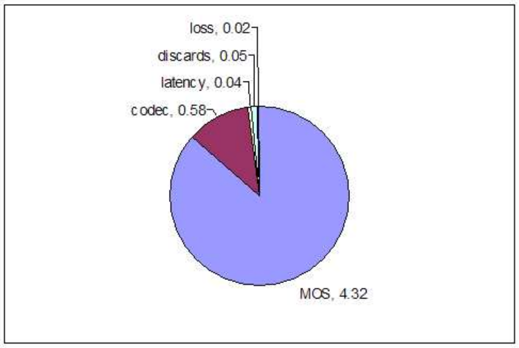

Figure 16: Overall MOS from any destination using G.711 over a cable network

most interesting result we found was, the uplink and downlink bandwidths do affect the voice quality. The MOS *to* any destination was found to be 4.1 using the G.711 codec while the MOS *from* any destination was 4.3. This is seen in Figure 16. The improvement in quality is due to the fewer number of discarded packets. Similar results are observed by using the G.729 codec.

6.4 Gigabit Ethernet Network

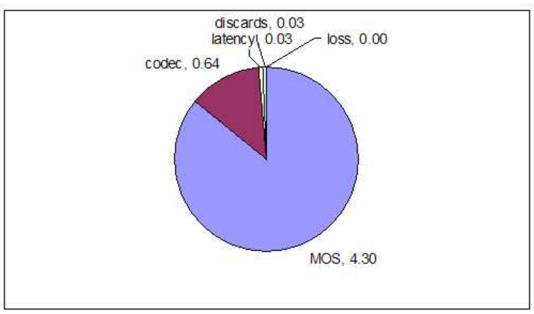

Figure 17: Overall MOS to any destination using G.711 over a gigabit ethernet network

A gigabit ethernet network easily satisfies the minimum bandwidth requirements of the codecs under test. This is seen in Figures 17 and 18. The degradation due to latency, packet discards and packet loss are negligible. The only degradation is due to the codec used. This is inevitable, and can be overcome only by using codecs which provide better voice quality at the

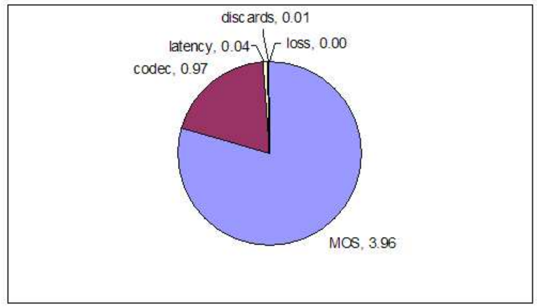

Figure 18: Overall MOS to any destination using G.729 over a gigabit ethernet network

price of higher bandwidth requirements. The same graph is observed in experiments to every location. As expected, gigabit ethernet network provides the best voice quality among all the networks tested. This raises an important question as to whether VoIP implementations can change the codec used according to the type of network. We discuss this and some other possible future directions in Section 7.

7 Discussion

From the analysis in the previous section we can conclude that the *available bandwidth* is an important factor in determining the end to end voice quality. We obtain the following ranking: Dialup Network $<$ 802.11 b/g Wireless Network $<$ Comcast Cable Network $<$ Gigabit Ethernet. It is interesting to note that even though in the ideal case the maximum bandwidth provided by a 802.11 b/g wireless network (54 Mbps) is much more than that provided by a cable network (3 Mbps/300 kbps), cable networks deliver much better voice quality. This is primarily caused by collisions or transmission errors on the wireless network reducing the effective bandwidth, however, the fact that the total bandwidth is shared amongst several users is also a significant contributing factor. Figures 19 and 20 show a sample traceroute analysis of voice packets from London to Seattle. Figure 19 corresponds to a home dialup connection with AOL as the ISP,

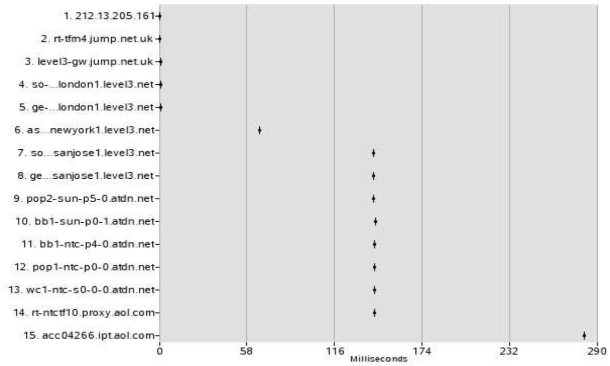

Figure 19: Traceroute analysis over a dialup network

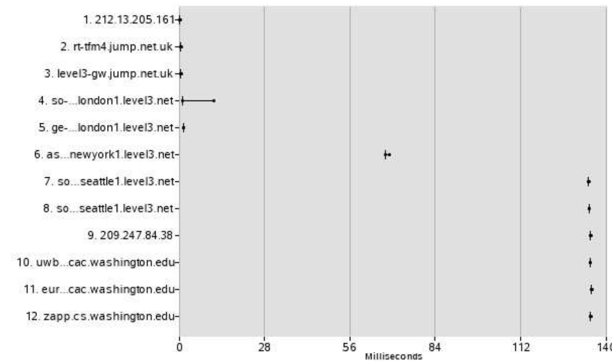

Figure 20: Traceroute analysis over a gigabit ethernet network

while Figure 20 shows the analysis for a connection over the CSE department gigabit ethernet network. The most striking feature of this analysis is that the delay over a dialup connection is more than double the delay over a gigabit ethernet network, the main culprit being the delay from the ISP to the end host which is almost the same as the intercontinental delay. It is this delay at the edge of the network that degrades VoIP quality in dialup networks.

8 Future Directions

Motivated by our findings we now propose possible future directions that address the most significant degradation factors.

8.1 Adaptive VoIP Clients

We have seen in Section 6 that voice quality depends on a number of factors. Depending on the bandwidth available and the type of network, the effect of these factors varies. If the network conditions are known prior to establishing the VoIP call, an appropriate codec can be chosen in order to achieve the best possible voice quality for the given network conditions. However, to provide the best possible speech quality as network conditions change over time, the codec must be adaptive; i.e., either the codec must be swapped out entirely for a different codec, or it must support different modes (frame size, bit rate, etc.) The AMR speech codec presented in [11] is an example of such an adaptive codec. This codec allows for the encoding bit rate to be changed adaptively in eight discrete steps within the range of 4.75 kbit/s to 12.2 kbit/s. The best obtainable MOS range from 3.0 at the lowest bit rate to 3.5 at the highest. As shown in Section 6, this is roughly equivalent to the performance of the G.711 and G.729 codecs when used over a dialup network. However, by using the AMR codec at its lowest bit rate, the network bandwidth requirement is reduced by approximately 40 %. In addition, the AMR codec has higher packet loss tolerance than any of the aforementioned codecs, which for use on *lossy* networks would be an advantage. This, combined with the lower bandwidth requirement, would make it possible to achieve reliable VoIP service on slow networks like dialup connections. However, if the AMR codec is to be significantly better than G.729, its equipment impairment factor, I_e, will need to be lowered while maintaining its high packet loss tolerance.

8.2 QoS Guarantees by the Client Side OS

We have seen that commercial VoIP clients have very low bandwidth requirements typically in the range of 10-30 kbps. A major fraction of this requirement is satisfied even on dialup networks. With the emergence of new codecs like iLBC which has a requirement of only 13 kbps, bandwidth would no longer be a constraint. However, during our experimentation we observed that background downloads affect the speech quality severely, especially on slow connections like dialup networks. This is because the background download process utilizes almost the entire available bandwidth (assuming data is available at that rate) and the VoIP client does not receive its minimum required share. But even on cable connections, the jitter introduced by other traffic impacts the voice quality. This motivated us to look for QoS guarantees from the local operating system. The local OS can prioritize the bandwidth allocation among the contending processes. This can be done in two straightforward ways:

1. Allocation of a minimum amount of bandwidth for applications

2. Giving higher bandwidth priority to time sensitive applications

A combination of these methods can also be used. Since an application running in user mode cannot modify the bandwidth reservation policy of the kernel, these changes have to incorporated into the kernel. It turns out that it is fairly easy to modify the kernel to allow for such requirements. To our knowledge this extension to VoIP has not been studied previously. We strongly believe that incorporating QoS guarantees by the client side OS would boost the user satisfaction especially on slow networks. Moreover, such extensions to the kernel would help the entire range of time sensitive applications, which would soon be widely used.

8.3 Error Correction

VoIP sends voice packets over UDP. Packets may be lost or discarded due to transmission errors, delays, network congestion, etc. Due to the time sensitive nature of voice traffic, the retransmission of damaged, delayed or lost packets is not feasible as a technique for correcting packet transmission errors. A forward error correction algorithm which offers improved voice reliability at the cost of minimum additional delay or bandwidth would be the ideal solution to this problem. The main idea is to transmit redundant packets that will allow the receiver to reconstruct the voice data even when a small fraction of packets are lost. Forward error correction on VoIP has been studied in research extensively and shows a lot of promise in delivering stronger QoS guarantees [9, 10].

Most of the forward error correction schemes use well established codes, such as Reed-Solomon codes. In order to generate the redundant packets, one has to sacrifice a fraction of the actual audio data in effect limiting the optimal quality achieved. However, when packets drop or packet loss occurs, the receiver is still able to reconstruct the audio data, yielding a much better MOS score on average.

Different code rates need to be chosen to deal with different packet loss ratios. The code rate affects the perceived audio quality as it limits the actual audio data transmitted over a fixed bandwidth. Therefore, a careful selection of the code rate needs to be done based on the packet loss ratio. However, the packet loss ratio is generally not known in advance and may vary over time. This naturally leads to the proposal of forward error correction schemes with adaptive rate. Most of the research done in the field suggests a feedback mechanism that monitors the packet loss ratio and adjusts the code rate accordingly to compensate.

Unfortunately, this approach has not been implemented by VoIP providers mainly due to the computational overhead involved in real-time encoding, the additional bandwidth requirements, and larger buffer space needed. However, there are some commercial patents pending in this field. Our best guess is that such systems will become available in the following years.

9 Conclusion

In this project our goal was not to evaluate the performance of various VoIP implementations but investigate the limitations of the technology itself. It is well known that the way the Internet has evolved over the years, it has not evolved into the ideal environment for QoS services. A major open question we faced when we started working on the project was whether we can ever achieve satisfactory VoIP services using the present Internet framework. We listed five basic questions at the beginning of the report. We felt that these are questions anyone interested in VoIP research would face. These questions directed our approach during various stages of the project. We feel that we have been able to answer these questions to the best possible extent. We hope that the analysis of VoIP presented in this report would help in deeper understanding of the technology and fuel future research in the field.

References

[1] AT&T `http://www.att.com/voip/`.

[2] Vonage `http://www.vonage.com`.

[3] ILBC codec. `http://www.globalipsound.com/pdf/gips_iLBC.pdf`.

[4] `http://www.voip-info.org`.

[5] Mean Opinion Score (MOS) terminology. `http://www.itu.int/rec/recommendation.asp?type=folders&lang=e&parent=T-%REC-P.800.1`.

[6] Session Initiation Protocol (SIP) `http://www.ietf.org/html.charters/sip-charter.html`.

[7] The E-Model, a computational model for use in transmission planning `http://www.itu.int/rec/recommendation.asp?type=items&lang=e&parent=T-RE%C-G.107-200303-I`.

[8] `http://www.testyourvoip.com`.

[9] Yvan Calas and Alain Jean-Marie. Audio quality for a simple forward error correction code under bursty traffic conditions. *International Conference on Communications in Computing (CIC 2004)*.

[10] Wenyu Jiang and Henning Schulzrinne. Comparison and optimization of packet loss repair methods on voip perceived quality under bursty loss. *Proceedings of the 12th international workshop on Network and operating systems support for digital audio and video*, 2002.

[11] J. Matta, C. Pepin, K. Lashkary, and R. Jain. A source and channel rate adaptation algorithm for amp in voip using the e-model. *NOSSDAV*, 2003.

[12] Oppenheim and Shafer. *Discrete-Time Signal Processing*. Prentice-Hall, 1999.