Wireless VoIP Performance Analysis

A.Karawita and L.Sun

Centre for Security, Communications and Network Research,
Plymouth University, Plymouth, UK
e-mail: info@cscan.org

Abstract

During the past few years we have been seeing an increasing popularity of Wireless VoIP (Voice over IP) services, mostly due to the increasing availability of the software driven smartphones running Google Android and Apple iOS, where VoIP applications can be downloaded and installed on to the mobile device and then simply start using it similar to how you would have done in front of a computer. However there are still many major challenges with regard to quality that need to be overcome in order to offer customer an excellent service. In this paper we provide a detailed analysis on the performance of Skype's SILK speech codec. We setup a wireless VoIP test bed on an Android driven mobile platform and investigate the impact of packet-loss and background traffic, on the performance of SILK. We used recommended speech samples from the ITU-T database, and three different experiments were conducted. We used data extraction and reporting tools, to measure and evaluate the network performance parameters such as jitter, inter-arrival time, packet loss, and variation of bit rate on all these experiments. A relationship between packet loss vs. jitter and inter arrival time were seen. With increasing levels of packet loss jitter and inter-arrival time were affected badly. It was seen for 5% and 10% packet loss jitter can be tolerated as shown by the MOS score of 4.44 and 3.88 respectively. Informal subjective quality measurements showed SILK was able to tolerate packet loss of up to 10% before it showed signs of degradation. We believe that this study can be helpful for the research community who are currently doing performance analysis on SILK.

Keywords

Wireless VoIP, Android, SILK speech codec, Jitter, Inter- arrival, Bit rate, User perceived quality, MOS, PESQ.

1 Introduction

Traditionally Voice over Internet Protocol (VoIP) were transmitted via wired networks, but gaining rapid success to the telecommunication industry are Wireless LAN based VoIP technologies, popularly known as VoWLAN, that are emerging and at its early infancy, are expected to be the future of wireless communications. The popularity is due to the low deployment costs, ability of being more scalable and easy access to wireless hot-spots. In VoIP the quality of speech are notably the most important Quality of Service (QoS) requirements that govern end to end communication of this technology (Yamamoto and Beerends, 1997). VoIP transmits voice traffic via packet switched networks and packets get routed through the best and most efficient paths, virtual connections are setup for the duration of the call, between the caller and receiver, which gets terminated when one party hangs up.

Since SILK is a relatively new open source voice codec, there has not been much talk about it in the research community, however many publications consider the user perceived quality of Skype. Hence in the following we give an overview of some of the interesting work that has been done with regards to Skype and SILK. In Schlosser et al (2010), they provided a detail analysis on SILK speech codec and compare it to its predecessor iLBC and GSM, using objective quality measurements based on PESQ. They conducted experiments with bulk and random packet loss, applying error patters directly to the encoded VoIP frames. Their results show SILK performs well in all of these conditions. Other results showed if loss is applied to shorter speech samples the degradation is more. In Ramo and Toukomaa (2010), their research they consider three relatively new open source speech codecs, SILK, CELT and Broad Voice and compare with G.718 and some of the other ITU-T standardized speech codecs. Their results indicate SILK codec was the best performing codec compared to the rest, which received high MOS scores. Also how SILK was able to achieve different bit rates and change bit rate on frame by frame basics was look at. Other publications tried to compare Skype and MSN, and experiment were done with regard to packet loss, NAT scenarios; cross traffic and available bandwidth to come to a conclusion which VoIP application performs better (Chiang et al, 2006). According to Chen et al (2009), on their experiment on jitter buffer behaviour on Skype they mention that Skype’s playout buffer remains within the range of 250 and 350ms, and does not adjust the buffer size according to the magnitude of the jitter.

The main objectives of this research paper are to (1) investigate the impact of packet loss and background traffic on the performance of SILK codec, and analyse key network performance parameters such as jitter, inter-arrival time, bit rate and packet loss. (2) To evaluate the performance of the recently published open source Skype SILK codec on two Android driven Smart Phones, and measuring the user perceived quality between the SILK sender and the receiver using informal subjective tests and objective measurements using PESQ.

The paper is structured as follows. In section 2, the testbed architecture and experimental setup is described. The results and discussion of the experiments would be presented in section 3. Finally we draw conclusions and future work in section 4.

2 Testbed Architecture and Experimental Setup

The figure 1 below illustrates a schematic diagram of the wireless testbed which would be setup to conduct all of the scheduled experiments that include measurement based experiments and subjective, objective based tests.

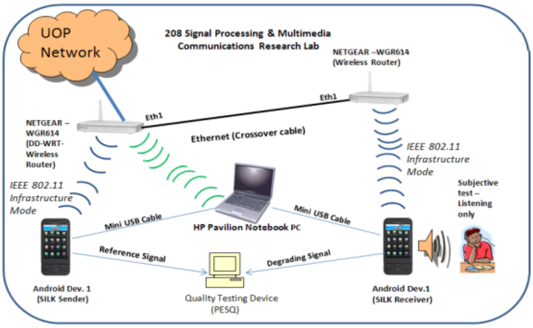

Figure 1: Testbed Architecture

2.1 Speech samples

As a clear guidance ITU-T P.862 should be referred when considering the selection of a reference speech signal. The recommended length of the reference signal for PESQ tests should be between 8 and 30 seconds, this includes any silence before after and between utterances (ITU-T, P.862). The ITU-T recommended reference speech samples can be obtained from the (ITU-T P.50).The converted .wav speech samples need to be sampled at 8000Hz and copied to the SD card of the SILK sender. A section of code was modified in the original SILK sender which will now read the wav files directly from the SD card of the phone (sender side).The degrading signal was captured to conduct objective quality measurements using PESQ, and the packets transmitted over the network were captured for network performance analysis. There were certain limitations that came up when capturing the degraded signal; eventually a voice recorder was used to capture the degrading signal from the speaker of the receiver. This may have some impact on recorded speech signal quality.

2.2 Measurement based experiments

TCPDUMP was used to collect voice packets under different packet loss rates at SILK sender and SILK receiver; these captured data would then be analysed using AWK scripts to calculate the performance parameters, which include, inter-arrival time of packets, jitter, packet loss and variation of bit rates. Three experiments were carried out. In the first experiment we would be introducing packet loss of 0%, 5%, 10%, 20%30% and 40% at a time for the duration of the call. In the second experiment, Gilbert model was used to simulate unconditional loss probability of 0%, 40%, 5%, 25%, 10% 0%, 30% interchanging every 10 seconds. In the third experiment, two streaming applications on both SILK sender and receiver would be played in the background along with SILK call simultaneously. NetEm was used to simulate packet-loss for these experiments except the third.

2.3 Objective and Subjective Testing Method

Objective measurements were conducted to measure end-to-end speech quality using PESQ. Packet loss would be simulated similar to the way subjective tests were carried out. Due to the limitations in the recording mechanism, it was expected that the MOS scores given by PESQ should be far less compared to the subjective scores. Hence, in this case it wouldn't be practical to benchmark PESQ against the subjective score, as mapping would be accurately distributed. The listening-only tests were carried out as an informal subjective test that is the experiments were conducted on a non-sound proof research lab and following the guidelines that was recommended by the ITU-T P.800 – Methods for Subjective Determination of Transmission Quality (ITU-T P.800, 1996).

3 Experimental results and Discussion

In this section we present and discuss the results from measurement based experiments and the overall subjective and objective assessments. The results would be discussed experiment wise.

In experiment 1 we used different packet loss levels of 0%, 5%, 10%, 20%, 30% and 40%. Under 0% packet loss conditions, the average inter arrival time was 20.63ms and during periods of silence we observed the inter arrival times was reduced to 9.999ms as the DTX feature was enabled, that reacts by sending equal size bytes in quick succession until the speech parts are encoded (Rao and Toukomaa, 2010). When packet loss is increased for 0% to 5%, SILK was able to maintain the average inter arrival time, to about 20 to 21ms. But as the packet loss rate increased to 10%, it was clear that the inter arrival time was increasing steadily and when the packet loss rates reached 40%, the average inter arrival time was well above 40ms, which was quite high. This too meant the jitter was getting added up and speech frames that arrive later than the length of the jitter buffer can be discarded, which meant further packet loss. We examined the relationship of inter arrival time and packet loss illustrated in the figure 2. It shows the inter-arrival time from 40% packet loss, and the points plotted reflect the times the packet loss has occurred. As seen most of packet loss can be seen when Inter arrival time is more than 30ms. Most of the spikes that have inter arrival time of more than 100ms, was due to packet loss which can be clearly seen. Also looking closely we are able to observe packet loss occurring during periods of silence, below 10ms; this certainly would not have a significant effect on the quality of speech. The location of loss within the speech was looked at (Sun et al, 2001) and it highlighted unvoiced signals have no impact on overall quality. We can also assume when packet arrive having a high inter-arrival time, and it is more than the time of the jitter buffer, there is a high chance that these packets can also be discarded, shown by almost every tall spike.

When packloss starts to increase, the levels of jitter values starts to increase significantly as seen in the figure 3, which shows jitter for packetloss of 0% (bottom line), 20% (middle line) and 40% (above line) . Looking at 0% packet loss the line is fairly flat, apart from two few spikes due to the delay in loading and playing the speech from the phone memory, comparatively the lines for 20% and 40% bounces all over the place, and spikes fluctuates rapidly,which indicates there is high jitter on the call and as a result a lower MOS is expected. The average jitter levels for 40% packet loss was close to 35ms which is quite high, compared to the 25ms Jitter for 20% packet loss. We also saw the jitter level for 5% and 10% packet loss can be can be tolerated upto some extent as user perceived quality was not effected significantly shown by the higher MOS scores.

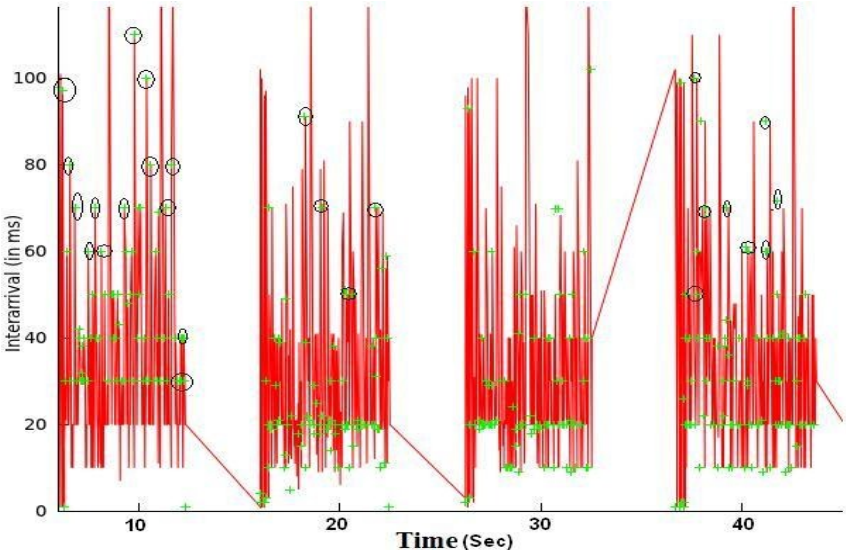

Figure 2: Inter-arrival vs Packet loss

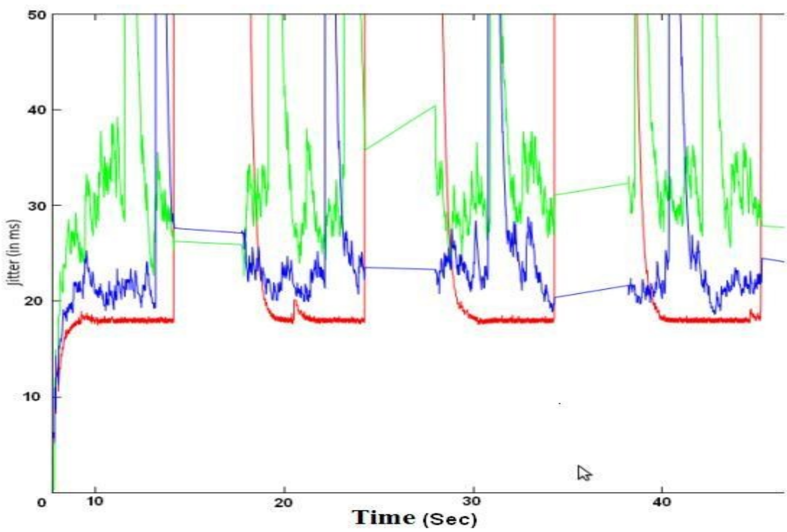

Figure 3: Jitter vs Packet loss

In experiment 2 we hope to understand the effects of random packet loss happening during different time intervals of the call , rather than a constant rate in experiement 1. One of the important result from this experiment was to understand the relationship between packet loss and jitter. Figure 4 illustrates the jitter line after 110 sec on to the call, as seen the line is really messy, with scattered spikes bounching throughout the graph, this leads to loosing the smoothness in a call, with annoying distortions at the receiver. The points marked in the jitter line indicates the time packet loss has occurred. With increasing levels of packet loss, we are able to see jitter is starting to increase and all the packets that were loss were being scattered across the areas of high jitter. Also it can be seen when jitter is very high, it can contribute to a higher loss as shown above. As the packets are randomly dropped by the router, the packets that arrive later than playback can also be dropped (Yamamoto and Beerends, 1997).

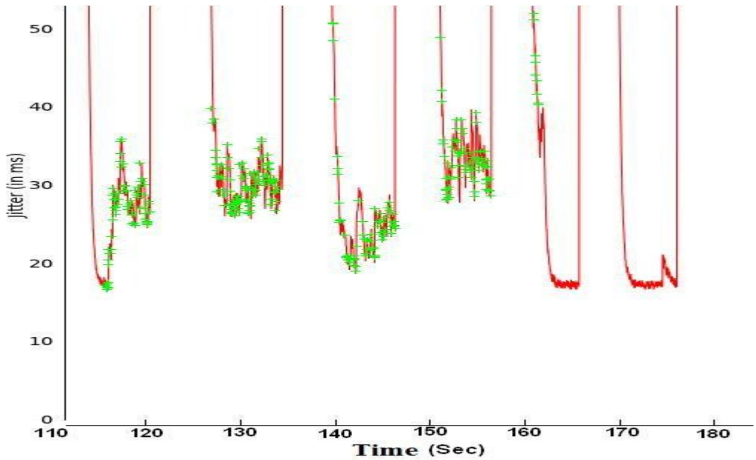

Figure 4: Jitter Vs. Packet loss (40%)

In experiment 3, the SILK encoder and decoder was experimented by introducting background traffic while the call was on progress. No packet loss was simulated for this experiment. Once the call is placed between the sender and receiver, a random video form youtube was being streamed and played at both sender and receiver. It was seen that that jitter was affected by background traffic.The impact of background traffic on voice quality was seen when Jitter was severely affected by background traffic and the bit rate drop below average when background traffic was been played and returned to the average bit rate once the video was finished playing, this is illustrated in figure 5. It was observed that http packets were overwhelming the UDP packets. And the SILK packets sent were severely interrupted and for short durations it was observed no SILK packets were sent at all, as a result the bit rate falls to as low as 9000 bits per second.

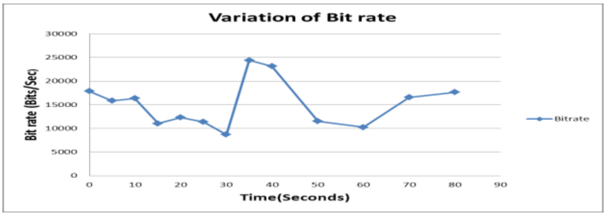

Figure 5: Variation of bit rate for background traffic

3.1 Subjective vs. Objective Quality Measurements

We assessed the user perceived speech quality for SILK codec using objective and subjective measurements and tryed to understand if packet loss had an impact on the user perceived quality. Table 1 shows the summarized average MOS and PESQ scores for different levels of packet-loss.

Packet-Loss (%)	MOS	PESQ
0%	4.55	3.15
5%	4.44	3.08
10%	3.88	2.91
20%	2.87	2.56
30%	1.88	0.90
40%	1.12	-1.00 (Invalid Result)

Table 1: MOS vs. PESQ score

MOS scores gathered for different packet loss scenarios were recorded according to intelligibility and clarity of the user's audio quality perception. The MOS scores are very satisfying for 0% 5% and 10% packet-loss with 4.55, 4.44 and 3.88 respectively which is a very respectable score known as *communication quality*. But after the packet-loss levels are gradually increased to 20% and above the overall speech quality starts to gradually degrade. For 40% packet-loss 88% of the subjects rated as “Bad” as it was impossible to distinguish as to what sample was being played at the receiver.

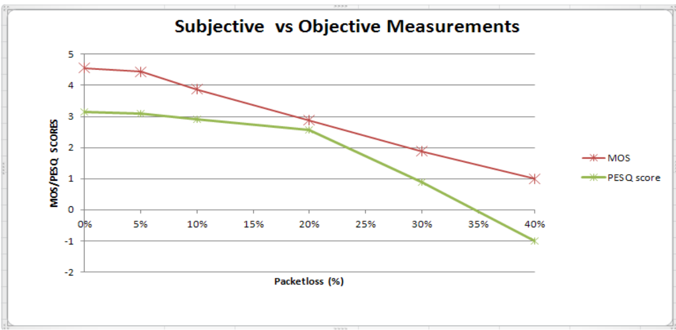

Figure 6: Subjective vs. Objective Measurements

The PESQ score of 3.15 can be viewed as a reference point for no-loss and with increasing levels of packet loss the PESQ scores start to fall gradually, and for 40% packet loss, the degraded signal could not be processed by PESQ. According to Schlosser (2010), SILK is able to achieve PESQ scores of 4.5 for 0% packet loss. Figure 6, illustrates the variation of both subjective and objective scores. The MOS (top) is linearly decreasing after 5% packet loss. At 5% packet-loss the overall speech quality, was hardly affected, but at 10% the MOS was still 3.88 which is still a good communication quality. In the case of PESQ (bottom) a linearly decrease is seen after 20% packet loss. From 0% to 20% packet-loss the difference in PESQ scores was only 0.59, which indicated the quality was quite well preserved compared to 0.67 in subjective tests. Therefore it was a clear indication that subjective results vary depending on user expectations and personal opinions, but in objective measurements these are not part of the algorithm.

4 Conclusion and Future Work

In this research paper we studied the impact of packet loss on user perceived quality of Skype SILK codec. From the results obtained we found out SILK codec does extremely well in packet loss conditions of 5% to 10%, where a high MOS was seen. But after 10%, we observed the speech degraded very rapidly and when it came to 30%, the speech was almost impossible to understand. We studied the relationship between packet loss vs. inter arrival time and packet loss vs. jitter. In both cases we found, packet loss can be a major factor that can influence the values of Jitter and inter arrival time. When it came to jitter, we observed that with increasing levels of packet loss that was being simulated, it caused the jitter line to fluctuate very rapidly and all the packets that were loss were being scattered across the areas of high Jitter. For the case of inter-arrival time, it was observed when packet loss is increased from 0% to 10%. The inter arrival time for 20% packet loss was well above 20ms and at 40% packet loss the inter arrival time was close to 40ms at the receiving end. We also notice SILK initial started with a high bit rate, and after 15-20 seconds the average rate is achieved. The impact of background traffic on voice quality was seen when Jitter was severely affected by background traffic and the bit rate drop below average when background traffic was been played and returned to the average bit rate once the video was finished playing. In our future work, we will pursue measuring user perceived quality on different sampling rates, as SILK supports super wideband up to a sampling rate of 24Hz.

5 References

Chen, K.-T., Huang, C.-Y., Huang, P. & LEI, C.-L.(2009). Quantifying Skype user satisfaction. *In:* SIGCOMM '06 Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications 2006.

Chiang, W.-H., Xiao, W.-C. & Chou, C.-F. (2006). Performance Study of VoIP Applications:MSN vs. Skype.publication. *In:* MULTICOMM 2006 proceedings., 2006. MULTICOMM.

ITU-T (2001). Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs. *T Recommendation P.862, February 2001*

ITU-T (1996). Methods for objective and subjective assessment of quality. ITU-T Recommendation P.800,August 1996.

Raake, A. (2006). *Speech quality of VoIP: assessment and prediction*, John Wiley and Sons.

Ramo, A. & Toukomaa, H. (2010). Voice Quality Evaluation of Recent Open Source Codecs. *INTERSPEECH 2010.* Nokia Research Center, Tampere, Finland: INTERSPEECH 2010.

Schlosser, D., Jarschel, M., Burger, V. & Pries, R.(2010). Monitoring the User Perceived Quality of SILK-Based Voice Calls. *In:* Australasian Telecommunication Networks and Applications Conference 2010 Auckland, New Zealand. ATNAC.

Sun, L., Wade, G., Lines, B. M. & Ifeachor, E. C (2001) Impact of packet loss location on perceived speech quality. In: Proc. Internet Telephony Workshop (IPtel 2001), USA – New York.

Yamamoto, L. A. R. & Beerends, J. G. Y. 1997. Impact of network performanceparameters on the end-to-end perceived speech quality. *ExpertATM Traffic Symp., Mykonos, 1997* . Greece, Sep. 1997.