Performance Analysis of Voice Call using Skype

M.Pradhan and L.Sun

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

Abstract

The purpose of this research is to investigate how Skype adapts its voice sender bitrate in reaction to network congestions and what Skype's performances are in terms of voice Quality of Experience (QoE) or Mean Opinion Score (MOS) under different network conditions. The voice quality of Skype was evaluated by setting up a VoIP testbed that allowed making voice calls between two Skype clients, each installed on a separate PC. The network conditions were altered as desired and sample speech files from the ITU-T P.50 database were used to play through Skype during a Skype-to-Skype voice call. The resulting audio was recorded along with the call traffic being captured. The audio quality of the recorded speech samples was then evaluated by using the PESQ algorithm and the network parameters were calculated from the captured call traffic using Awk scripts. The results obtained showed that the voice quality of Skype deteriorates with increasing amounts of packet loss in the network. Further, the throughput values that were obtained indicated that Skype tends to reduce its send bitrate when it encounters packet loss in the network which could result in increased delay. The values obtained for the interarrival times indicated that the packet interarrival time increases with increasing packet loss and this could lead to increase in the mouth-to-ear delay. Finally, the jitter values obtained were found to be well within the acceptable limit for VoIP and thus it could be concluded that jitter does not have much effect on the audio quality of Skype.

Keywords

VoIP, Skype, voice quality, PESQ, throughput, interarrival, jitter

1 Introduction

VoIP (Voice over Internet Protocol) is a form of voice communication that utilizes audio data in order to transmit voice signals to the end user. Over the last few years, VoIP has emerged as one of the most vital technologies in the world of communication and is providing stiff competition to the traditional telephone lines due to its lower cost and richer features (Kazemitabar et al. 2010).

Skype is one of the most popular VoIP applications that are freely available on the web today. Skype allows users to make voice and video calls over the internet and also provides instant messaging (IM) services. Skype makes use of a proprietary protocol known as the Skype protocol and its operation is based on P2P (Peer-to-peer) technology while other VoIP clients make use of the traditional client-server architecture. Skype call traffic is generally transmitted in the form of UDP (User Datagram Protocol) packets over IP networks. The payload of the UDP packets is compressed and encrypted, thus securing the call data (Speidel and Eimann, 2010).

Existing research on Skype has indicated that Skype provides better voice/video quality in adverse and changing network conditions as compared to its other VoIP competitors. Menezes Filho et al. (2005) analysed and compared Google Talk, MSN Messenger, Skype and Yahoo Messenger using Ethereal software to capture the packets for each. From their study they concluded that Skype was the best software for having a VoIP conversation, providing better voice quality with less phonetic loss since it had the highest speed of them all for transmitting information.

Adopting a measurement-based approach, Chiang et al. (2006) performed a quantitative evaluation of both MSN and Skype. The results obtained indicated that Skype's overall throughput showed an up to 47% improvement over that of MSN and its MOS score showed an over 50% improvement over that of MSN. Also, the variance of interarrival time in Skype was found to be very low.

Ahmed and Shaon (2009) analysed the various performance aspects of three widely used VoIP applications – Skype, GTalk (Google Talk) and Gizmo. Although under ideal conditions it was found that the performance of all three applications was similar, when the network parameters were changed, Skype exhibited better adaptation quality than the other two.

Skype's superior voice quality is possibly due to its supposed built-in adaptation/control mechanism which enables it to adapt its voice/video sender bitrate automatically in order to ease network congestion. However, due to Skype's proprietary nature, details of how this adaptation/control mechanism works are still unknown and needs to be investigated.

A proper understanding of the mechanisms that Skype utilizes in order to provide good voice/video quality in difficult network conditions may help in contributing towards the improvement of other VoIP applications available to the users. For this purpose, the effect of the different network parameters on the perceived voice quality of Skype will have to be studied and investigated in detail. Also, further research in this direction may result in further improving the overall quality of voice/video that can be made available to the users by VoIP providers including Skype.

The goal of this research is to investigate and analyse the performance of voice quality of Skype under different network conditions by setting up a VoIP testbed, conducting voice calls between two Skype clients installed on two separate PC's, collecting the call traffic and analysing it. The PESQ (Perceptual Evaluation of Speech Quality) algorithm will be used to evaluate the speech quality since it is one of the most widely used tools in the industry for objective measurement of voice quality. The network parameters such as throughput, interarrival time and jitter will be calculated using Awk scripts.

The rest of the paper is structured as follows. In Section 2, the setup of the VoIP testbed is provided along with the description of the experiments performed using the testbed platform. The results of the experiments conducted are presented and discussed in Section 3. Finally, Section 4 concludes the paper and suggests some future investigations.

2 VoIP testbed platform and the experimental scenarios

In this section we present the VoIP testbed that was setup along with the experiments that were carried out using the testbed platform.

2.1 The testbed setup

Figure 1 illustrates a diagram of the VoIP testbed that was setup in order to perform the various experiments as part of this research.

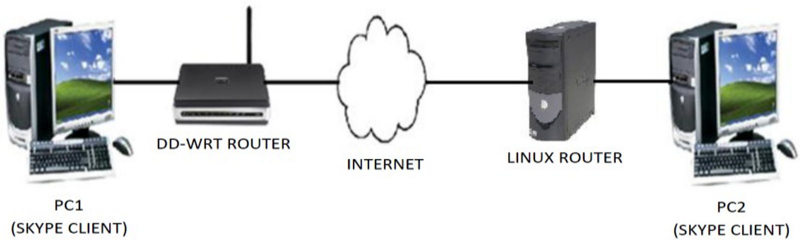

Figure 1: VoIP Testbed Architecture

The main components of the VoIP testbed are the two local PC's – PC1 and PC2, and the DD-WRT wireless router. Besides that we have the Linux router. On PC1, the software applications that were installed and utilized were OPTICOM Opera v3.5, Skype 5.3, Wireshark (Version 1.6) and Windows Media Player 11 (WMP) while on PC2, the software applications that were installed and utilized were Audacity (Version 1.2.6) and Skype 5.3.

The DD-WRT router is a NETGEAR wireless router that has the DD-WRT (Linux-based firmware) installed on it. The DD-WRT router was used in the experiments to modify network conditions such as introducing packet loss with the help of Linux commands like "tc" and "netem". It connected PC1 to the internet via cables. The Linux router is a Linux machine that has Ubuntu installed on it and is configured to act as a router. It connected PC2 to the internet via cables.

2.2 Design and Methodology for the experiments

Two major experiments were conducted as part of this research using the testbed platform that is described in 2.1. The methodology and the procedure for both the experiments is the same except for the use of a different reference speech sample file to be played through Skype for each. A reference speech file of significantly longer length was used in performing the second experiment. The following sub-sections will explain the way in which both the experiments were performed.

2.2.1 Selection of speech samples

Speech samples from the ITU-T P.50 database were used for all the objective voice quality tests that were performed. The P.50 database consists of several speech samples available in different languages and all in the "WAV" file format. For each language, there are 16 speech samples of which 8 are male voices and 8 are female voices. The language that was selected for the experiments performed as part of this research was British English and only the first three files in female voice were used since the gender was not under consideration. The sample rate for the "wav" files was converted to 8000 Hz using Audacity and the files were saved again.

For Experiment 1, the B_eng_f1.wav speech file was selected to be played through Skype. The length of this speech file is 7.2 seconds, the sample rate is 8000 Hz and the audio is in female voice. The speech file consists of three sentences separated by a few seconds of silence. For Experiment 2, three sample speech files, namely, B_eng_f1.wav, B_eng_f2.wav and the B_eng_f3.wav were combined using Audacity to form a single reference speech file of length 21.745 seconds. This reference speech file has a sample rate of 8000 Hz and consists of 9 sentences spoken three at a time in three different female voices, all of which are separated by a few seconds of silence.

2.2.2 Introduction of Packet Loss

In order to introduce packet loss into the network, it was required to gain access to the DD-WRT router and this was done by using the 'telnet' command on the 'default gateway' for PC1. Next, packet loss was introduced at the network interface that connects PC1 to the DD-WRT router by using the Linux command 'tc'. The following command was used to introduce new packet loss or to change the amount of packet loss that was already introduced:

```
tc qdisc add/change dev eth0 root netem loss%
```

'Netem' provides network emulation functionality and is controlled by the 'tc' command. The amount of packet loss to be introduced is specified by the 'loss%' (The Linux Foundation, 2009). In both the experiments, 6 different amounts of packet loss (0%, 5%, 10%, 15%, 20% and 25%) were introduced into the network using the 'tc' command and further tests were conducted with regards to the voice quality of Skype and the estimation and calculation of the throughput, interarrival time and jitter in Skype for each amount of packet loss introduced.

2.2.3 Evaluation of voice quality using PESQ

In order to evaluate the voice quality of Skype, a sample reference speech file was selected to be played through Skype. A different reference speech file was used for Experiment 1 and Experiment 2 as was discussed in 2.2.1. At first, the desired amount of packet loss was introduced into the network. Then packet capture was started on Wireshark on PC1 and a Skype-to-Skype voice call was conducted from PC1 to PC2. This call was answered on PC2 and the 'record' button was hit on Audacity on PC2.

Next, the reference speech file was played in WMP on PC1. Once the file had finished playing - the recording was stopped in Audacity, the voice call was ended and packet capture was stopped in Wireshark. The sound recorded in Audacity and the packets captured in Wireshark were saved in their respective default file formats. This process was carried out 3 times for each amount of packet loss introduced into the network and all the sound files and pcap files were saved.

Now, each of the recorded sound files was edited in Audacity to conform to the waveform and length of the original reference speech file and was exported in “wav” file format. Next, each of these “wav” files was used one by one along with the original reference speech file in OPTICOM’s Opera software and the PESQ algorithm was run. The PESQ results obtained are presented in Section 3.

2.2.4 Evaluation of throughput, interarrival time and jitter

Awk scripts were designed for the evaluation of the throughput, interarrival time and the jitter in Skype. Ubuntu 11.04 was installed on a laptop and ‘tcpdump’ command was used in the Terminal to extract the data from each pcap file in text format. Next, the Awk scripts were run in the Terminal to calculate the throughput, interarrival time and jitter for each pcap file and the results obtained are shown in Section 3.

The Awk script for calculating the throughput was based on the understanding that the throughput was the total amount of data sent in the form of UDP packets by Skype on PC1 to Skype on PC2 during the total time duration of the voice call. The script was also used to evaluate the average payload size for each pcap file. The Awk script for calculating the interarrival time was based on the evaluation of the time difference between the arrivals of two successive packets at the destination. The Awk script for calculating jitter was based on the following formula (Toncar, 2010):

$$J(i) = J(i-1) + (|D(i-1, i)| - J(i-1)) / 16$$

The jitter “J(i)” after every i-th packet that was received, was estimated by calculating the change of interarrival time and dividing it by 16 in order to reduce the noise as well as to reduce the influence of large random changes (Toncar, 2010). The value thus obtained was then added to the jitter value for the previous packet i.e. “J(i-1)”. The jitter was evaluated over the entire duration of each voice call that was conducted, which lasted around 60 seconds.

3 Experimental results

3.1 PESQ results

The average PESQ MOS scores obtained from both the experiments are displayed in Figure 2. The results obtained indicate that the voice quality of Skype deteriorates when there is packet loss in the network. The more the amount of packet loss in the network, the more is the degradation in quality of the voice signal. The PESQ results obtained from Experiment 2 are found to be more accurate in terms of the clearly visible decreasing trend of the PESQ scores with respect to increase in packet loss in the network.

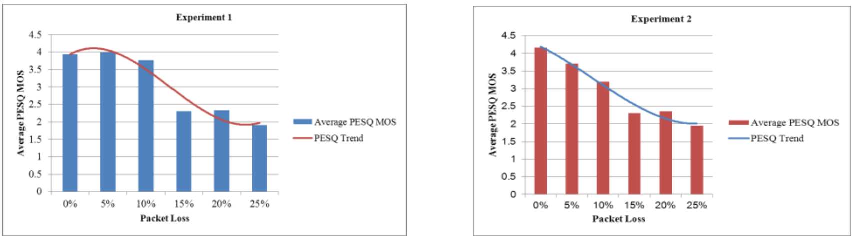

Figure 2: Average PESQ MOS scores obtained from Experiments 1 & 2

3.2 Throughput results

The results for the average throughput of Skype obtained from both the experiments are displayed in Figure 3. These results indicate that Skype's throughput decreases with the presence of increasing amounts of packet loss in the network. This means that Skype reduces the amount of data it sends when it encounters packet loss. The throughput was evaluated in terms of 'kilobits per second' (kbps). The results obtained for the average payload size of the packets sent by Skype indicate that initially, though Skype reduces its voice sender bitrate with increasing packet loss, it transmits the audio packets with increased payload to compensate for the missed data. However, it is observed that when the packet loss is at 15% or more, the payload size gets reduced significantly and this needs to be further investigated.

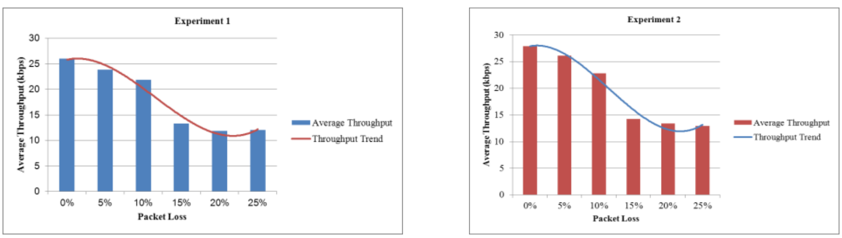

Figure 3: Average Throughput values obtained from Experiments 1 & 2

3.3 Interarrival time results

The results for the average packet interarrival times obtained from Experiment 1 and Experiment 2 are represented in Figure 4.

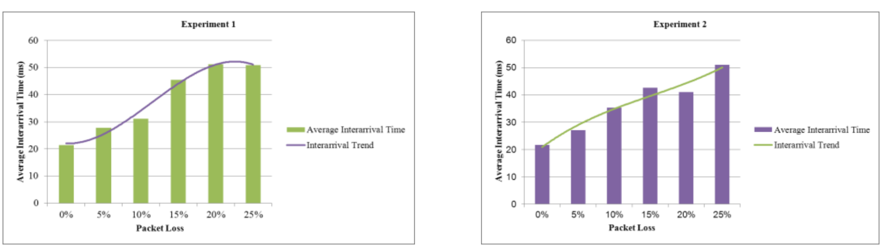

Figure 4: Average Interarrival Time values obtained from Experiments 1 & 2

The above results clearly indicate that the packet interarrival time increases consistently with increase in packet loss in the network and a similar trend was observed in the results obtained from both the experiments. This is because when a packet is lost and does not arrive at the destination, the interarrival time continues to add up until the next packet arrives successfully. The interarrival time was evaluated in terms of ‘milliseconds’ (ms).

3.4 Jitter results

The results for the average jitter values obtained from Experiment 1 and Experiment 2 are represented in Figure 5.

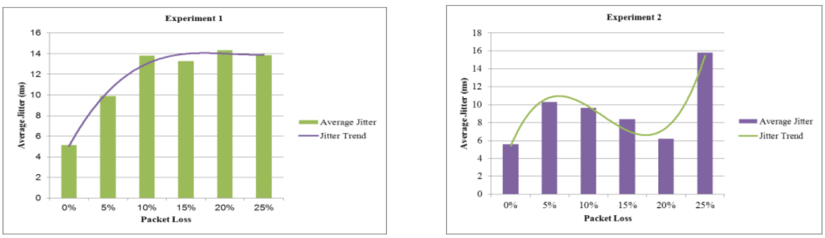

Figure 5: Average Jitter values obtained from Experiments 1 & 2

Although the jitter results obtained from Experiment 1 show a slight increasing trend with increase in packet loss in the network, no particular increasing or decreasing trend is observed in the jitter results obtained from Experiment 2 as can be seen in Figure 5. The interarrival time was evaluated in terms of ‘milliseconds’ (ms). Overall, the jitter values obtained from both the experiments indicate that the voice quality of Skype is not affected much due to jitter as jitter of up to 25 ms is deemed to be acceptable (irockasterisk, 2011).

4 Conclusions and future work

This paper investigated the voice quality of Skype. The main goal of the study was to analyse the behaviour of Skype in different network conditions and to observe how its voice quality was affected in these conditions. For both the experiments conducted, the PESQ MOS scores obtained indicate that the voice quality of Skype tends to deteriorate with increasing amounts of packet loss present in the network. Skype is able to provide decent to average voice quality up until 10% packet loss present in the network. But once the packet loss is at 15% or more the voice quality is affected badly with lots of noticeable gaps in the audio.

The throughput values of Skype that were evaluated under the presence of different amounts of packet loss indicate that Skype tends to reduce its send bitrate when it encounters packet loss and this could lead to increased delay. On investigating the payload size of the packets sent by Skype in different packet loss conditions, it is observed that initially, though Skype reduces its send bitrate with increasing packet loss, it transmits the audio packets with increased payload to compensate for the missed data. However, it is observed that when the packet loss is at 15%, the payload size gets reduced significantly and this needs to be further investigated.

The values for the packet interarrival times obtained from both the experiments clearly indicate that the packet interarrival time increases consistently with increase in packet loss in the network and this could lead to increase in the mouth-to-ear delay, thus affecting the voice quality of Skype that is experienced by the end-user. The jitter values obtained from both the experiments are fairly inconsistent, especially those obtained from Experiment 2. However, the values obtained are well within the acceptable jitter value for VoIP (25 ms) and this indicates that jitter does not have much effect on the quality of voice in Skype.

Overall, it is observed that Experiment 2 provides more substantial, clear and accurate results than Experiment 1 and at the same time helps to verify some of the results obtained from Experiment 1, thus justifying the use of a longer sample speech file in Experiment 2. In future, we would like to conduct research with regards to limiting the bandwidth of the network and then evaluating the voice quality of Skype as well as investigating the TCP-friendliness of Skype with background TCP traffic.

5 References

Ahmed, A.S. and Shaon, R.H. (2009) Evaluation of Popular VoIP Services. Adaptive Science & Technology [e-journal], 58-63. Available through: IEEE Xplore [Accessed 11 August 2011].

Chiang, W., Xiao, W. and Chou, C. (2006) A Performance Study of VoIP Applications: MSN vs. Skype. [online] Available at: http://multicomm.polito.it/proc_multicomm06_3.pdf [Accessed 10 August 2011].

irockasterisk (2011) VoIP quality: culprits and thresholds. [online] Available at: http://irockasterisk.blogspot.com/2011/06/voip-quality-culprits-and-thresholds.html [Accessed 27 August 2011].

Kazemitabar, H., Ahmed, S., Nisar, K., Said, A.B. and Hasbullah, H.B. (2010) A Survey on Voice over IP over Wireless LANs. [online] Available at: http://www.waset.org/journals/waset/v71/v71-63.pdf [Accessed 17 August 2011].

Menezes Filho, L.C.A., da Costa, M.L., Belem, R.L. and Arruda Filho, E.J.M. (2005) Performance and Quality of Service on Free Softwares for VoIP. [online] Available at: http://www3.iesam-pa.edu.br/ojs/index.php/TELECOM/article/viewFile/636/522 [Accessed 10 August 2011].

Speidel, U. and Eimann, R. (2010) How well does Skype compress its call data? [online] Available at: http://busy-byte.de/~raimund/documents/publications/ICAIT_2010_Paper.pdf [Accessed 8 August 2011].

The Linux Foundation (2009) netem. [online] Available at: http://www.linuxfoundation.org/collaborate/workgroups/networking/netem [Accessed 23 August 2011].

Toncar, V. (2010) VoIP Basics: About Jitter. [online] Available at: http://toncar.cz/Tutorials/VoIP/VoIP_Basics_Jitter.html [Accessed 26 August 2011].