EFFICIENCY ENHANCEMENT OF AUDIO AND VIDEO CHAT APPLICATION OVER THE INTERNET

Dissertation submitted in fulfilment of the requirements for the Degree of

MASTER OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

By

RAJNISH KUMAR SINGH

11207403

Supervisor

ISHA

School of Computer Science and Engineering

Lovely Professional University

Phagwara, Punjab (India)

January – May, 2017

LOVELY PROFESSIONAL UNIVERSITY

Transforming Education, Transforming India

TOPIC APPROVAL PERFORMA

School of Computer Science and Engineering

Program : 1202D::B.Tech -M.Tech (Dual Degree) - CSE

COURSE CODE : CSE546 **REGULAR/BACKLOG :** Regular **GROUP NUMBER :** CSERGD0277

Supervisor Name : Isha **UID :** 17451 **Designation :** Assistant Professor

Qualification : ____________ **Research Experience :** ____________

SR.NO.	NAME OF STUDENT	REGISTRATION NO	BATCH	SECTION	CONTACT NUMBER
1	Rajnish Kumar Singh	11207403	2012	K1209	9779328905

SPECIALIZATION AREA : Networking and Security **Supervisor Signature:**

PROPOSED TOPIC : Efficiency enhancement of audio and video chat application over the internet.

Qualitative Assessment of Proposed Topic by PAC		
Sr.No.	**Parameter**	**Rating (out of 10)**
1	Project Novelty: Potential of the project to create new knowledge	6.60
2	Project Feasibility: Project can be timely carried out in-house with low-cost and available resources in the University by the students.	6.20
3	Project Academic Inputs: Project topic is relevant and makes extensive use of academic inputs in UG program and serves as a culminating effort for core study area of the degree program.	6.60
4	Project Supervision: Project supervisor's is technically competent to guide students, resolve any issues, and impart necessary skills.	6.80
5	Social Applicability: Project work intends to solve a practical problem.	7.20
6	Future Scope: Project has potential to become basis of future research work, publication or patent.	7.00

PAC Committee Members		
PAC Member 1 Name: Prateek Agrawal	UID: 13714	Recommended (Y/N): Yes
PAC Member 2 Name: Pushpendra Kumar Pateriya	UID: 14623	Recommended (Y/N): Yes
PAC Member 3 Name: Deepak Prashar	UID: 13897	Recommended (Y/N): Yes
PAC Member 4 Name: Kewal Krishan	UID: 11179	Recommended (Y/N): Yes
PAC Member 5 Name: Anupinder Singh	UID: 19385	Recommended (Y/N): NA
DAA Nominee Name: Kanwar Preet Singh	UID: 15367	Recommended (Y/N): Yes

Final Topic Approved by PAC: **Efficiency enhancement of audio and video chat application over the internet.**

Overall Remarks: Approved

PAC CHAIRPERSON Name: 11011::Dr. Rajeev Sobti **Approval Date:** 21 Nov 2016

ABSTRACT

Study and analysis of popular VoIP (voice/video-over-Internet protocol) applications (apps) on android smartphone under the same bandwidth has been done. Out of these three popular apps (Skype, Imo & Google Duo), Skype is most popular VoIP applications where as Google Duo has been recently launched by Google last year in august. Also in this paper a new android VoIP application has been developed with the implementation of adaptive filter. The traffic is captured with tcpdump and analysed with tools (Wireshark & SteelCentral Packet Analyzer). Skype is based on closed source and is a proprietary project whereas Google Duo, Imo, and developed application are based on WebRTC (an open source project maintained by Google Chrome team). The results indicates that Google Duo voice/video quality is best among the apps whereas Imo having the worst voice/video quality. In this research it is found that most of the variation in voice/video quality arise due to noise in the form of echo in the signal. So to deal with the noise and echo adaptive filtering algorithm need to be used to enhance the existing voice/video quality. Normalized Least Mean Square (NLMS) is used in the developed application to deal with noise and enhancement of VoIP calls. Developed application when tested and compared with popular it shows good quality VoIP call than IMO application that suffer significant packet loss.

DECLARATION

I hereby declare that the research work reported in the dissertation entitled "**EFFICIENCY ENHANCEMENT OF AUDIO AND VIDEO CHAT APPLICATION OVER THE INTERNET**" in partial fulfilment of the requirement for the award of Degree for Master of Technology in Computer Science and Engineering at Lovely Professional University, Phagwara, Punjab is an authentic work carried out under supervision of my research supervisor Mrs. Isha. I have not submitted this work elsewhere for any degree or diploma.

I understand that the work presented herewith is in direct compliance with Lovely Professional University's Policy on plagiarism, intellectual property rights, and highest standards of moral and ethical conduct. Therefore, to the best of my knowledge, the content of this dissertation represents authentic and honest research effort conducted, in its entirety, by me. I am fully responsible for the contents of my dissertation work.

Signature of Candidate

Rajnish Kumar Singh

Registration No. 11207403

SUPERVISOR'S CERTIFICATE

This is to certify that the work reported in the M.Tech Dissertation entitled "**EFFICIENCY ENHANCEMENT OF AUDIO AND VIDEO CHAT APPLICATION OVER THE INTERNET**", submitted by **Rajnish Kumar Singh** at **Lovely Professional University, Phagwara, India** is a bonafide record of his / her original work carried out under my supervision. This work has not been submitted elsewhere for any other degree.

Signature of Supervisor

ISHA

Date: ______________

Counter Signed by:

1) **Concerned HOD:**

   HoD's Signature: _________________

   HoD Name: ______________________

   Date: ____________________________

2) **Neutral Examiners:**

   External Examiner

   Signature: _________________

   Name: ____________________

   Affiliation: ________________

   Date: _____________________

   Internal Examiner

   Signature: _________________

   Name: ____________________

   Affiliation: ________________

   Date: _____________________

ACKNOWLEDGEMENT

First and Foremost I would like to thank almighty for giving me courage to bring up this dissertation. Before getting into thick and thin of this dissertation I would like to show my gratitude to some of the people who have helped me in this project. Firstly I would like to purpose a word thanks to my mentor Isha who has encouraged me to get through this dissertation. Secondly I would like to thanks my friends who gave me unending support and helped me in numerous ways from the stage when the ides of the thesis was conceived. I am very thankful to all of them for making my work complete successfully under their guidance.

Rajnish Kumar Singh

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Video chat is not a recent invention. In 1927 the inventor of telephone, Alexander Gramham Bell, quoted in the article of New York Times as saying "the day would come when the man at the telephone would be able to see the distant person to whom he was speaking" [1]. First commercial video calling was introduced by AT & T in 1964. In the early days videophone was used for voice/video calling. It is a telephone with screen on the top as shown in Figure 1. 1 for making video calling. It is also called image phone [2].

Figure 1. 1 Earliest Videophone

With the development of technology and computer advancement video calling are now performed with the help of computer as shown in Figure 1. 2.

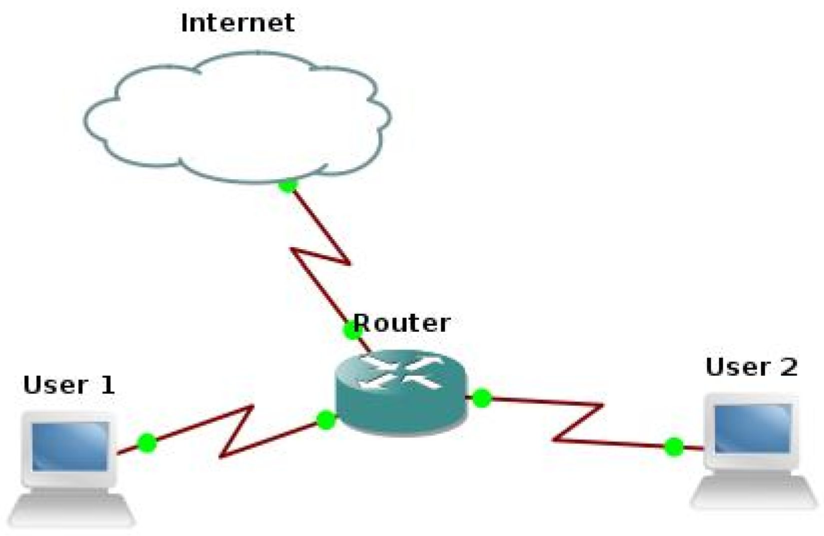

Figure 1. 2 General Architecture of Voice/Video calling

But nowadays smartphone and wireless services (3G/4G) has changed the above scenario and has taken the place of computer to perform voice/video calling as shown in Figure 1. 3.

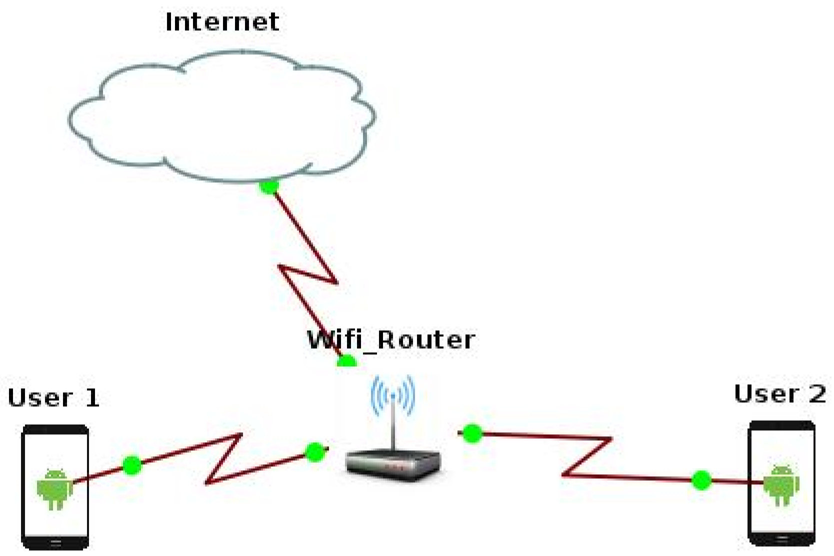

Figure 1. 3 Voice/Video calling via smartphone

In place of Wi-Fi router in Figure 1. 3 mobile network (3G/4G) can be used for making the VoIP calling and any smartphone with these service enabled can be used.

1.2 VoIP/Video PROTOCOL STACK

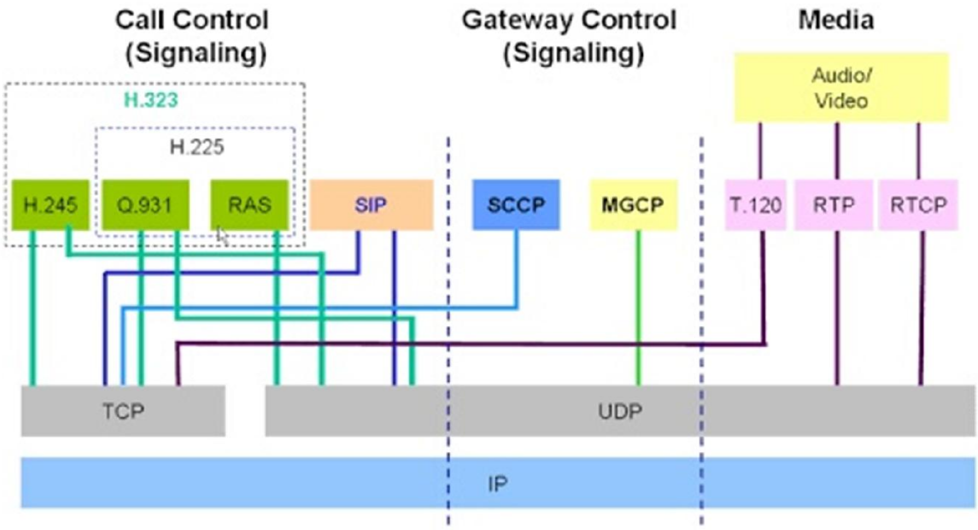

Figure 1. 4 VoIP/ Video Protocol Stack

Figure 1. 4 shows the architecture of VoIP protocol stack. It has three main parts:

- Call Control Signaling
- Gateway Control Signaling
- Media.

1.2.1 H.323

To connect public telephone network with computer to the internet 1TV designed a H.323 standard as shown in Figure 1. 5 [3], [4]. Gateway is responsible for translation of message from one protocol stack to another. Telephone message is transformed to internet message. It is responsible for registration. There are number of protocol used by H.323. Message transmitted between H.323 endpoints are defined by protocol H.225.0 and H.245. These protocols are used in any network architecture as they are generic protocols.

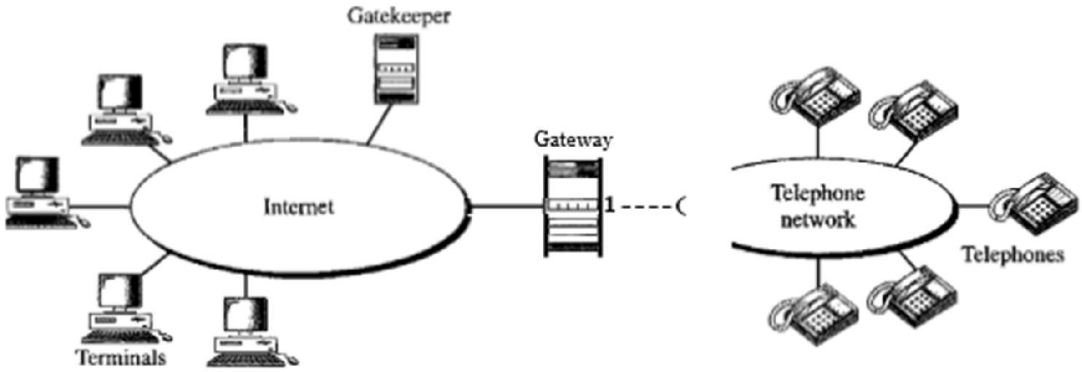

Figure 1. 5 Architecture of H.323

1.2.2 H.225.0

It has two parts: Q.931 and RAS. Q.931 is used to establish and terminate the connection between H.323 endpoints. This is called Q.931 signaling or call signaling. RAS (Registration, Admission and Status) is used between the endpoints and gatekeeper for registration with the gatekeeper. It is the responsibility of gatekeeper to allow or deny the registration with the network.

1.2.3 H.245

It is a control protocol between endpoints. It is responsible for managing media streams between H.323 sessions. It ensures media that is to be sent can be understand by another endpoints. These there protocols: RAS, Q.931, and H.245 is responsible for establishment and maintenance of VoIP call. It controls and negotiate media parameters with the endpoints for media transmission.

1.2.4 SIP (Session Initiation Protocol)

SIP is an alternative to H.323 protocol. It is more flexible, and easy to use than H.323. VoIP community thinks SIP can be used in future although H.323 is widely used for VoIP communication till today. SIP does not take packet itself, it uses RTP (Real Time Transport Protocol). It supports any type of communication (Voice, Video, or instant messaging). It is used today by many enterprise applications. SIP is similar to HTTP.

1.2.5 RTP (Real Time Protocol)

For end-to-end delivery of real-time data like: video and audio RTP [5] is used. It includes payload type, sequence number, timestamp, and delivery monitoring. It runs on the top of UDP (User Datagram Protocol) as shown in Figure 1. 4 and Figure 1. 6 and supports multicasting for data transfer. It does not itself provide any mechanism to ensure on time delivery and quality of service rather it depends on lower level to do so. It does not delivery packets in sequence nor guarantee delivery of packets. At the receiver end using the sequence number packet is reconstructed by the receiver. It is designed for video conferencing. Figure 1. 7 shows RTP carrying media packets. Here MP2T is a media encoding technique that encode both video and audio packets.

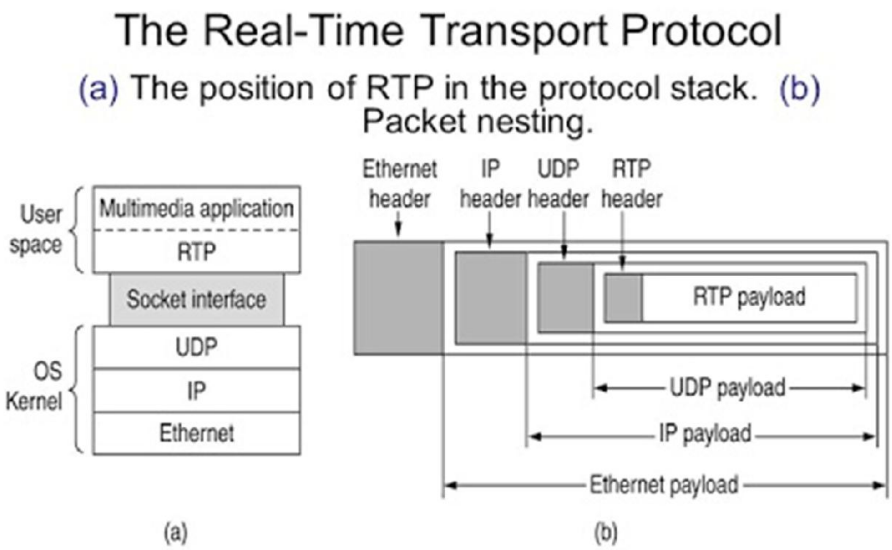

Figure 1. 6 Real Time Protocol Stack

1.2.6 RTCP (Real Time Control Protocol)

Media packets are sent using RTP over UDP Figure 1. 7. This requires two UDP ports, one for the media packet and other for control protocol (RTCP) [6]. It is used to get the feedback of the sent media over RTP to ensure quality of communication. It carries identifier for RTP source as in case SSRC identifier changes receiver can use this identifier to track participants.

1.2.7 Media Codec

Analog signals (voice/video) captured during VoIP calling is need to transfer over the network so they are converted to digital signals with the help of different compression and decompression algorithm called codecs (Coder-Decoder).

Table 1. 1 MOS (Mean Opinion Score)

Mean Opinion Score (MOS)	Quality Rating
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Mean Opinion Score (MOS) shown in Table 1. 1 is used as theoretical value for measuring the quality of VoIP. Some of the codecs used in VoIP is given in Table 1. 2 [7].

Table 1. 2 VoIP Codec Comparison

Codec	Data Rate (kbps)	Typical Datagram Size (ms)	Packetization Delay (ms)	Combined Bandwidth for 2 Flows	Typical Jitter Buffer Delay	Theoretical Maximum MOS
G.711u	64.0	20	1.0	174.40	2 datagrams (40 ms)	4.40
G.711a	64.0	20	1.0	174.40	2 datagrams (40 ms)	4.40
G.726-32	32.0	20	1.0	110.40	2 datagrams (40 ms)	4.22
G.729	8.0	20	25.0	62.40	2 datagrams (40 ms)	4.07
G.723.1	6.3	30	67.5	43.73	2 datagrams (60 ms)	3.87

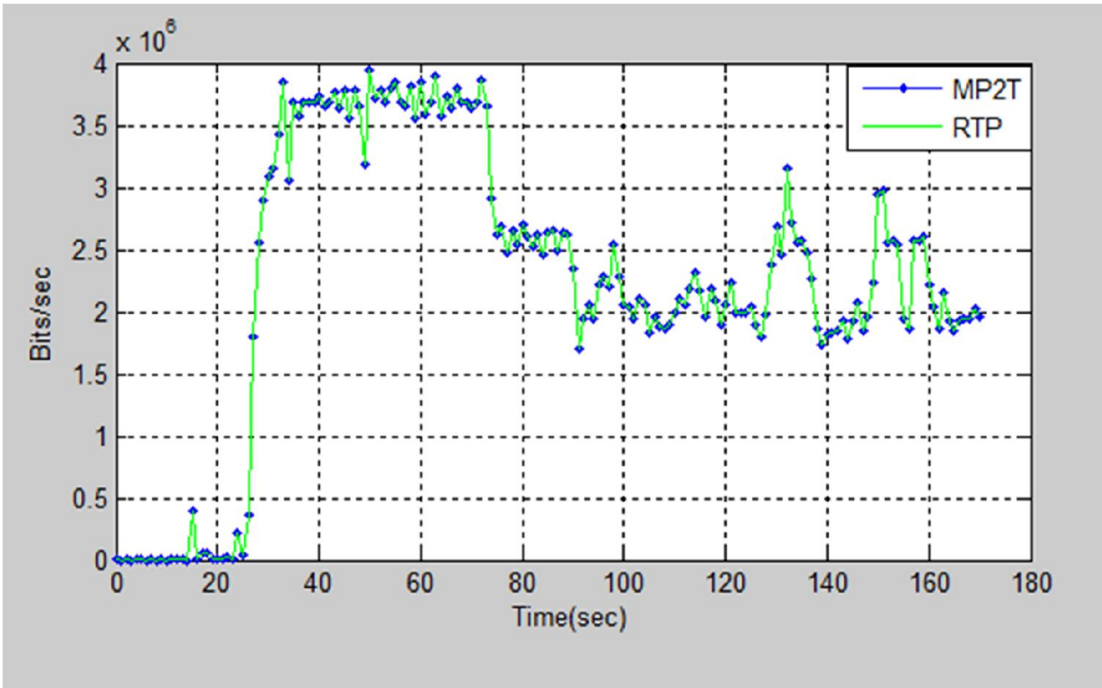

Figure 1. 7 RTP Carrying Media Packets

1.3 WebRTC ARCHITECTURE

WebRTC [8], [9] is an open source project aimed at providing voice/video communication on browser and mobile platform without any third party plugins. It provides peer-to-peer calling and is maintained by Google Chrome team. Major components of Web RTC:

- MediaStream: It allows browser to use camera and microphone of the user system.
- RTCPeerCommunication: It is responsible for establishing VoIP calls.
- RTCDataChannel: It allows to send data via peer-to-peer connections.

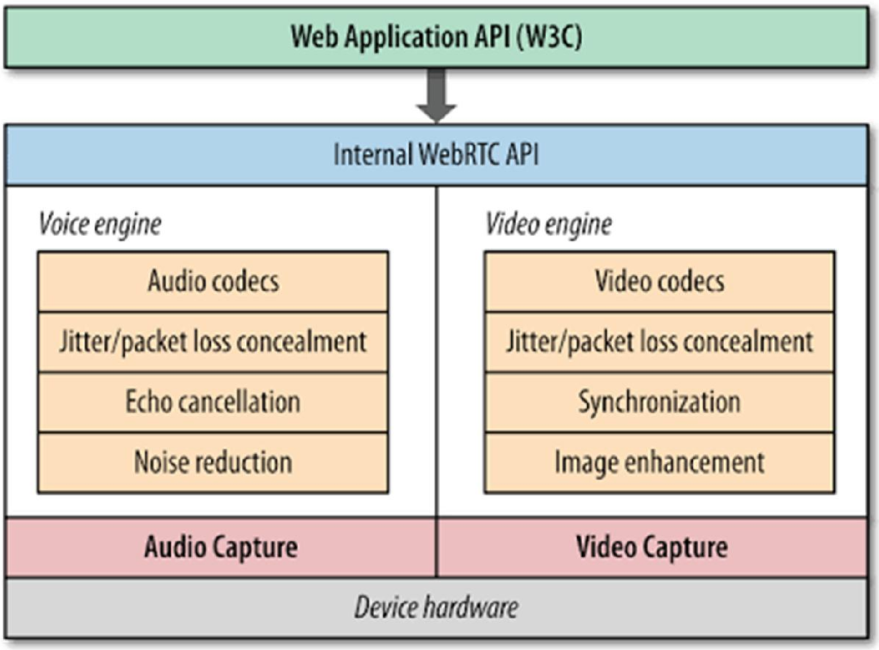

Figure 1. 8 WebRTC Architecture

1.3.1 Web Application API

This is the first layer of the WebRTC architecture where application such browser based on standalone application like desktop application or smartphone application work. This provides user with different control to the audio/video communication. It depends upon the user what he/she wants to do whether proceed for video call or voice call. He/ she can control the video quality and voice quality during the communication according to their need. Also application has some capability to adjust its voice/video quality according to network bandwidth if the user has set the quality preferences as auto mode. There is usually four quality types for video and voice is given in the application like low, medium, high, and HD quality.

1.3.2 Audio and Video Engines

We need an application API to establish communication between the end devices with the same application installed which is provided by the WebRTC API. This API contains two main engines like voice engines and video engines.

Since WebRTC uses browser to perform audio/video communication, it needs to access the system hardware (Camera and Microphone) to capture audio/video streams. Captured streams need to be processed to enhance the quality by removing noise and echo cancellation, bitrate must be able to adjust with variable bandwidth and latency between users.

Once received at the receiver, the process is reverses and audio/video is decoded to adjust with delays. Whole process is complex but WebRTC uses browser feature that process the signals more easily. Different encoding algorithms are used to optimize the media streams and error-concealment is used to hide the effect of packet loss and jitter. Figure 1. 10 shows the architecture of MediaStream and Figure 1. 8 shows its levels in WebRTC architecture. This is done by Voice Engine. Similarly Video Engine process the video by optimizing the image quality and using compression techniques to compress the video packet and error-concealment to reduce the effect of packet loss and jitter.

1.3.3 Audio Codecs

Audio engines provide audio codes (a program) to encode the speech after getting in the form of digital signal. It compress the digital signal and then it is send over the network during the conversation by the sender. On the receiver end audio codec receive the encoded digital signals which then decoded to get the speech signal. According to [10] audio codec used by WebRTC are Opus and G.711. Other similar codecs are iSAC and iLBC.

Opus Audio Codec

Opus is an audio codec which is freely available (open source) Figure 1. 9 and define by IETF RFC 6176 [11] which has bitrate from 6 kbit/s to 510 kbit/s with frame size 2.5 ms to 60 ms having sampling rate 8 kHz to 48 kHz.

G.711 Audio Codec

G.711 audio codec is an ITU-T standard open source widely used voice with sampling frequency 8 kHz. Its bitrate is 64 kbit/s with a delay of 0.125 ms with a built in packet loss management. It allows audio signal in the range of 300 -3400 Hz with a sampling rate of 8000 samples per sec and false under narrowband Figure 1. 9.

iSAC Audio Codec

internet Speech Audio Codec (iSAC) is developed by Global IP Solution as an audio compression which is a wideband and superwideband codec suitable for audio communication over RTP protocol. It has a sampling frequency of 16 kHz or 32 kHz with bitrate 10 kbit/s to 32 kbit/s in case of wideband and 10 kbit/s to 52 kbit/s in case of superwideband. It has a packet size from 30 to 60 ms with delay 3ms due to algorithm Figure 1. 9.

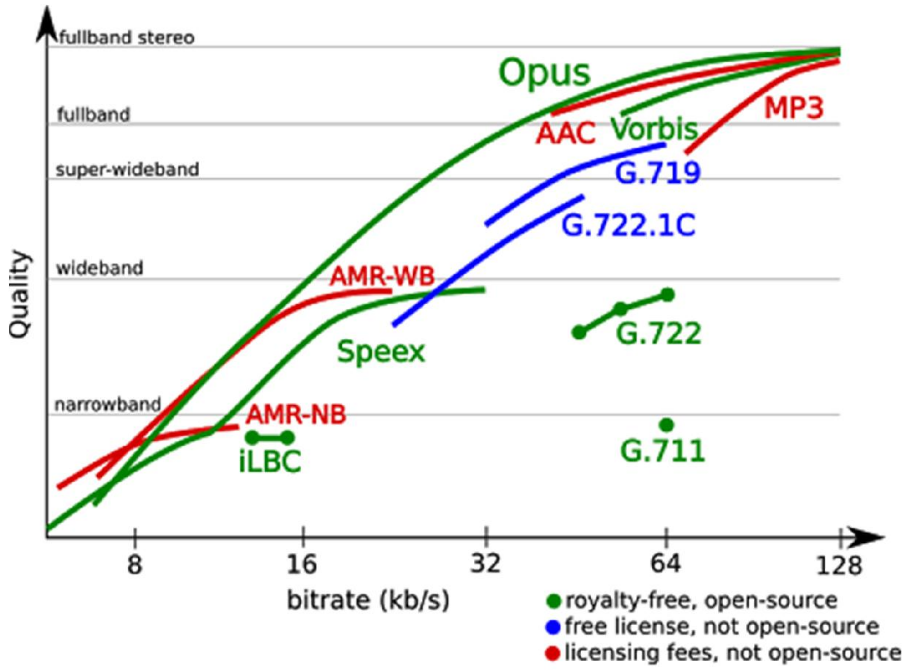

Figure 1. 9 Audio Codec Quality

iLBC Audio Codec

Internet Low Bitrate Codec (iLBC) developed by Global IP Solution as an audio compression which is a narrowband open source audio codec suitable for audio communication over RTP protocol. It can handles lost frames due to degrade speech quality which occurs due to packet loss and packet delay. It has a sampling frequency of 8 kHz with bitrate 15.2 kbit/s for 20 ms frames and 13.33 kbit/s for 30 ms frames with mean opinion score 4.14 Figure 1. 9.

1.3.4 Jitter and Packet Loss Management

The difference between the packet delay due to various factors like: - congestion in network, wrong way of queuing model, or configuration error. Normally delay remain constant. It does not interfere with communication in case of TCP/IP but for real time communication like VoIP network this can cause a problem. Due to delay packet may not reach the end devices in real time causing degrade VoIP quality. During the communication in real time packet loss and jitter are the serious problem of concern as they both cause improper communication quality. In WebRTC audio engines and video engines are responsible for managing jitter and packet loss or packet delay caused by the network to avoid quality loss in real time communication.

1.3.5 Echo Cancellation

In VoIP conservation when user speak to send voice over the network it is transmitted to the next user on the conservation meanwhile same voice is also hear by the sending user causing echo. Echo is generated by following ways: -

- Audio from the loudspeaker is received by the device microphone also called acoustic echo.
- Signal travelling through wire is received by the neighbour wire as these signals are in the form of electrical signals.

Echo is normally occur all the times but problem arise when it interferes with speech signals causing noise and conservation quality degrades causing unable to listen the user sound clearly. Sometimes user hear their own voice during the conservation first after that sending user voice is heard this is due to delay between the words send by the user. Delay less than 25 ms does not make any problem as it is not noticed by our ear but if it increases user face echo problem as sound bounces back to the sender itself.

WebRTC audio codec engines have some algorithm (filtering algorithm) based on mathematical calculation to reduce the echo by cancelling the unwanted noise before sending to the media for transmission. There are many echo cancellation techniques but most of the techniques are proprietary.

1.3.6 Noise Reduction

Process of removing unwanted signals from the original signal or raw signal to get the useful signal is termed as noise reduction. Digital devices as prone to noise signal. It can be of random signal or white signal. Signal when passed over transmission media gets electromagnetic interference causing noise generation in the signal. Signal-to-noise ratio (SNR) is used to measure the quality of voice. If low SNR value then less noise and quality of speech is increased. For this there are many noise reduction algorithms to remove the noise. Audio codec engines have some algorithm to reduce the noise during the conservation.

1.3.7 Video Codecs

Highly efficient video codec called VP8 is used by WebRTC architecture provided by video engines. It is an open source video codec which is the improvement of VP7 codec. Browsers like: Chrome, Firefox, Chromium, and Opera supports VP8 codec in HTML video. It is combined with voice codec Opus to give real time communication over the network during VoIP calls.

1.3.8 Image Enhancement

During video call proper stable image in necessary for good quality of video call. WebRTC supports image enhancement techniques to give proper clear image in real time during communication.

1.3.9 Synchronization

Synchronization of video is necessary to keep video calling in proper sequence otherwise video displayed during the conservation will be broken and blur or no video establishment is possible during the VoIP call. WebRTC provides video synchronization with the help of video engines.

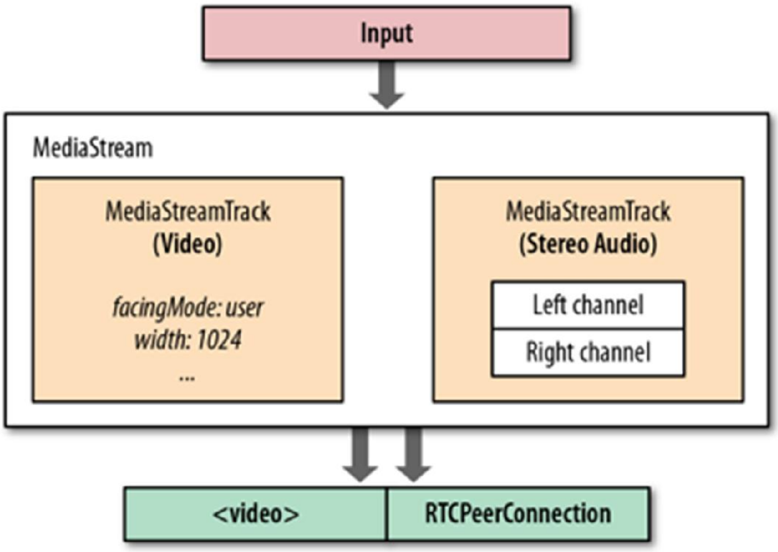

Figure 1. 10 MediaStream

1.4 ECHO

In VoIP conservation when user speak to send voice over the network it is transmitted to the next user on the conservation meanwhile same voice is also hear by the sending user termed as echo. There are different types of echo: -

- Acoustic echo
- Hybrid echo

1.4.1 Acoustic Echo

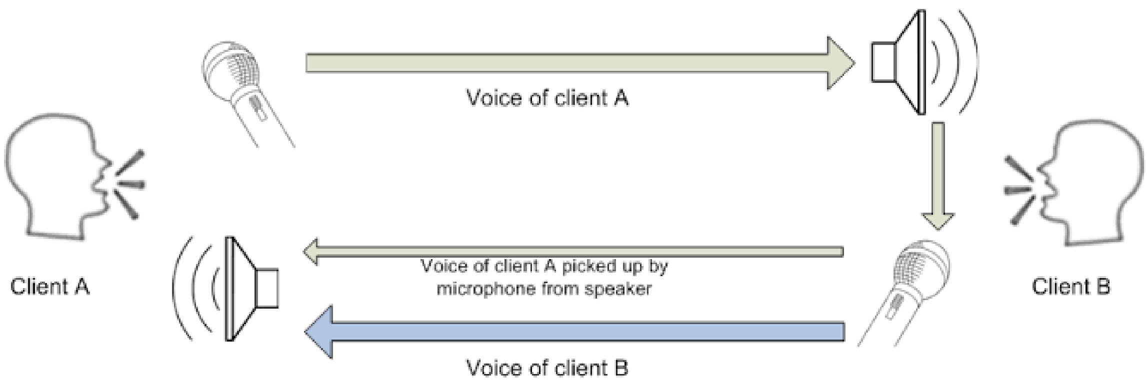

Figure 1. 11 Acoustic Echo Generation

During the communication on VoIP network whenever voice of a user A is transmitted over the network is received by the microphone of the user B and sent back to the user A, acoustic echo is generated as in Figure 1. 11. This type of echo is also generated when microphone is too near to the speaker of the device. It is very common in VoIP calls. User A will hear his/her own voice which can be annoying and degrades the real time communication quality.

1.4.2 Hybrid Echo

This type of echo is generated in telephone where a pair of wire is used to carry communication in both direction. Usually a hybrid circuit is used but due to impedance imperfection echo is generated.

1.5 FACTOR AFFECTING VOIP CALL

There are many factors that affects the VoIP call. Some of them are:

- Audio Codec
- Latency
- Jitter and Jitter buffer
- Packet Loss
- Silence Suppression
- Echo
- Other Network Parameters

1.5.1 Latency

There are many factors that affect the speed of the vehicle on the highway. Road traffic, traffic sign, and weather affect the speed of the vehicle. Likewise latency is considered as a speed for data packets which is affected by various network parameters. Time taken by a data packets (VoIP packets) to reach the receiver end and return back to sender is termed as round-trip latency. It created delay which results in echo generation that is caused mostly by slow network. It is measured in milliseconds (ms). Higher the latency increases the packet delay which in turn echo generation. Synchronization between voice and video during communication is affected. Following are the cause of latency: -

- Data packets are transmitted slow over the network during the communication due to insufficient bandwidth
- Sometimes firewall blocks some of the traffic of VoIP communication
- Use of wrong audio and video codecs for encoding and decoding the data packets or due to algorithmic delay.
- Hardware or device used for communication is old enough to be used in VoIP communication.

1.5.2 Packet Loss

During the communication over the network there is always a chance of packet being lost due to network factors. In TCP lost packets are recovered where as in UDP traffic lost packets are not recovered. VoIP communication uses UDP in most of the time. The task of the codec to compensate the lost packets with the help of previous packets. Packet loss of 5% is compensated and user in not able to notice it as there is no such degradation in VoIP call quality. But above 5% packet loss degrades the call quality, causes delay and hence echo may arise.

1.5.3 Other Network Parameters

Sufficient network bandwidth is required for proper VoIP calls. Bandwidth can be used by many application limiting bandwidth for the real time communication like VoIP calls. Increase in network load decreases the call quality. One way to get rid of it is to allow high priority packet to pass through network first which is a tough task as this need to be done at every level like: routers, switches, WAN links, etc. Reserving the bandwidth for real time communication.

1.6 ACOUSTIC ECHO CANCELLER (AEC)

Acoustic Echo Canceller (AEC) [12] [13] is used to remove echo from the data signals to provide good quality speech. Different algorithms are used to filter out the noise signals from the digital data. This system also reduce the background noise and enhance the voice quality. Good AEC system provides high crystal clarity audio under different noisy environment Figure 1. 12.

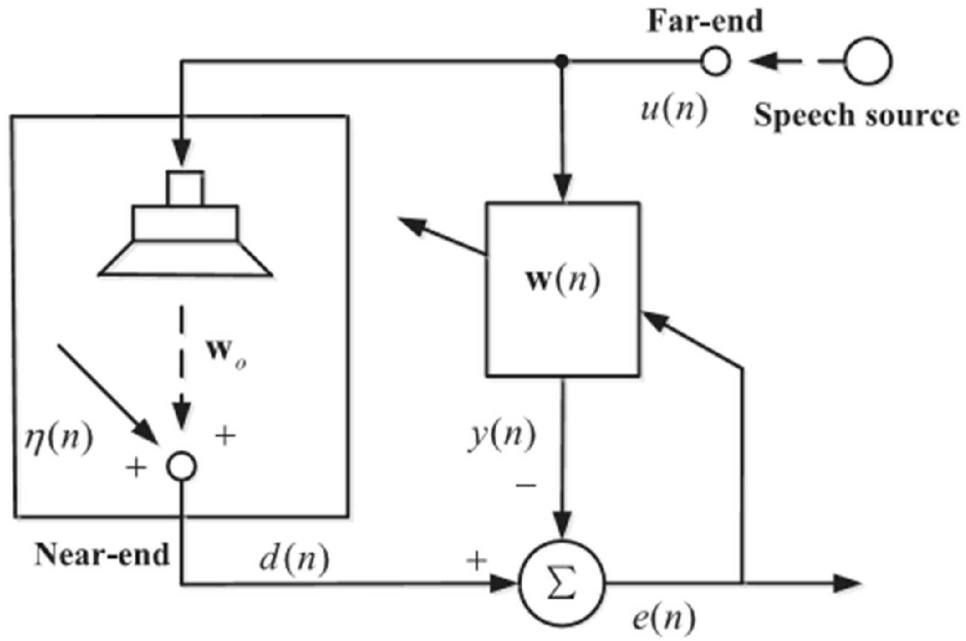

Figure 1. 12 Block Diagram of Acoustic Echo Canceller

Acoustic echo canceller system receive the input signals $u(n)$ from the source which is received by the microphone by the means of loudspeaker and echo signal $d(n)$ is created. Different adaptive filters $w(n)$ are used in the system receive $u(n)$ to give $y(n)$ as the output signals. Error $e(n)$ is estimated by

$$e(n) = d(n) - y(n) \tag{1}$$

This error $e(n)$ reaches to zero if no one is taking on other side of the conservation i.e. and ideal condition which is not in real scenario.

In this paper echo is reduce with the help of adaptive filter like normalized least mean square (NLMS).

1.6.1 Adaptive Filter

Adaptive filter [12] [13] [14] [15] [16] is a system that finds out the relationship between two signals that often uses digital processor for computation or have some complex logical computation working iteratively. It process the signals in real time. It tries to find out the output signal from the input signal. Its parameters are adjusted to make relationship between input signal and output signal. Cost minimization and minimum error is done by adjusting the filter function on next iteration.

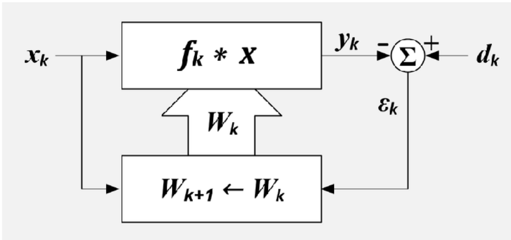

Figure 1. 13 Adaptive Filter Block Diagram

Where k is the sample number, x is the reference input, X is the set of recent value of x, d is the desired input, W is the filter coefficient, Ɛ is the error, f is the filter response, upper rectangle is for linear filter and lower rectangle for adaption algorithm. Following are the types of adaptive filter algorithms: -

- Least Mean Square (LMS) filter
- Normalized Least Mean Square (NLMS) filter
- Recursive Least Square (RLS) filter

1.6.2 Normalized Least Mean Square (NLMS) Filter

It is improvement of LMS filter in which filter is adjusted based on the current error. This paper uses NLMS filter for nose reduction in VoIP calls.

Weights $\underset{w_n}{\rightarrow}$ are updated every step using the error of the previous step.

$$\underset{x_n}{\rightarrow} = (x[n], x[n-1], \dots, x[n-N+1])^T \tag{2}$$

Where N is the number of iteration (samples), $\underset{x_n}{\rightarrow}$ is the input signal. $\underset{w_n}{\rightarrow}$ is calculated as

$$\underset{w_n}{\rightarrow} = \ (w[0], w[1], \dots, x[N-1])^T \tag{3}$$

Then $y[n]$ is calculated as

$$y[n] = \vec{x}_n \cdot \vec{w}_n \tag{4}$$

Weights are updated as

$$\vec{w}_{n+1} = \vec{w}_n + \mu \vec{x}_n \frac{\vec{x}_n^{\,T} \vec{w}_n - d[n]}{\vec{x}_n^{\,T} \vec{x}_n} \tag{5}$$

Where μ is the step size ranging $0 \leq \mu \leq 2$

There are many application of adaptive filter most common application is used for adaptive noise filtering by filtering or reducing out unwanted noise / echo from the speech signal during the communication over the network.

1.7 TOOLS AND SOFTWARE USED

1.7.1 MATLAB

MATLAB (matrix laboratory) [17] Figure 1. 14 is software consisting of many tools developed by MathWorks. It is a high level language for mathematical, scientific, technical computation with inbuilt visualization and graph generation. We can simulate any work using Simulink and also run on hardware with the help of it. It also give code generation in C and C++ language. We can convert our matlab file (file with .m extension) into jar file and other native package for use in other language like Java. This paper uses the same method to generate jar file with the help of matlab file which is later used in android application development. Some of the uses of matlab are:

- Mathematical computation
- Algorithm development
- Modelling, simulation, and prototyping
- Data analysis, and visualization
- Scientific and engineering works
- Application development and Graphical User Interface (GUI) development

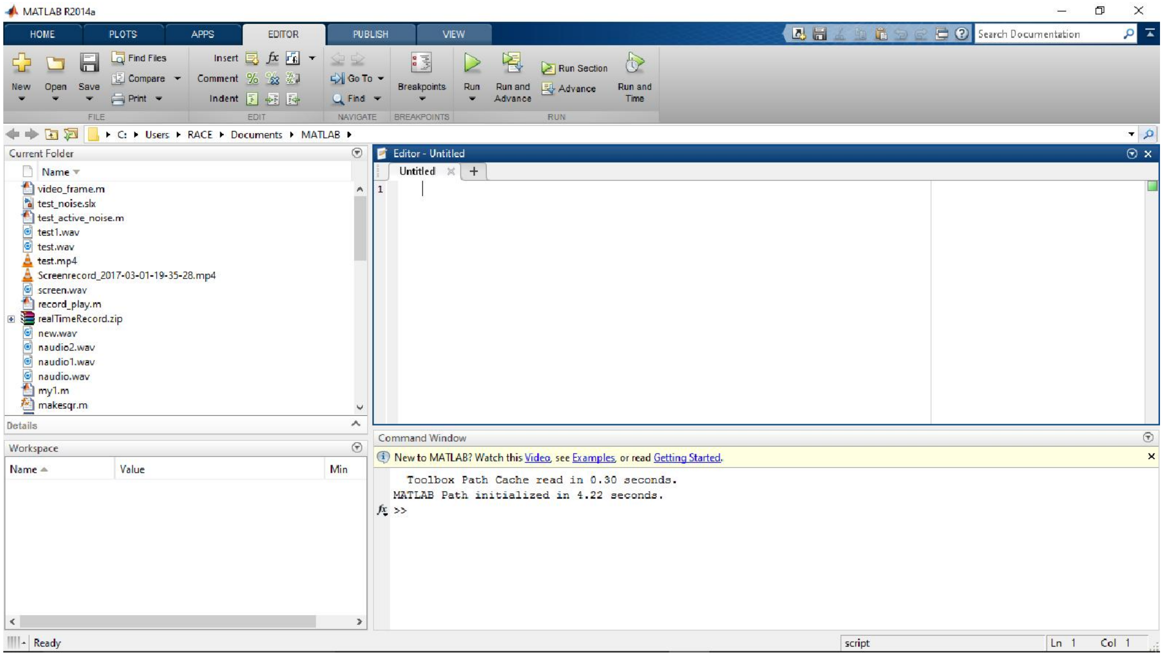

Figure 1. 14 Main Window of MATLAB

1.7.2 Wireshark

Wireshark [18] Figure 1. 15 is an open network analyzer, and packet sniffing tool. We can investigate in depth of the network and its protocol and find out the issue with it.

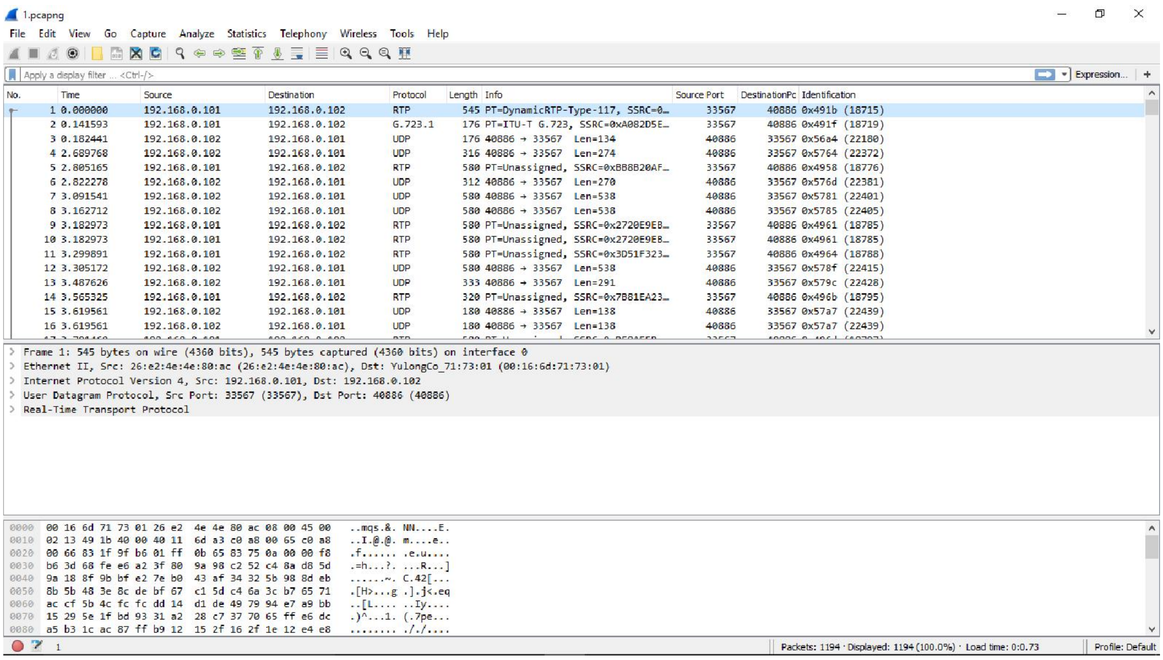

Figure 1. 15 Wireshark Window

1.7.3 SteelCentral Packet Analyzer

Same as Wireshark SteelCentral Packet Analyzer [19] Figure 1. 16 is a network analyzer tool but is not an open source like Wireshark. It gives user an easy interface to get the detail information of the packets no need to filter with the help of command like Wireshark. It has built in module for filtering the packets. We can analyse every detail information of the network with graphs.

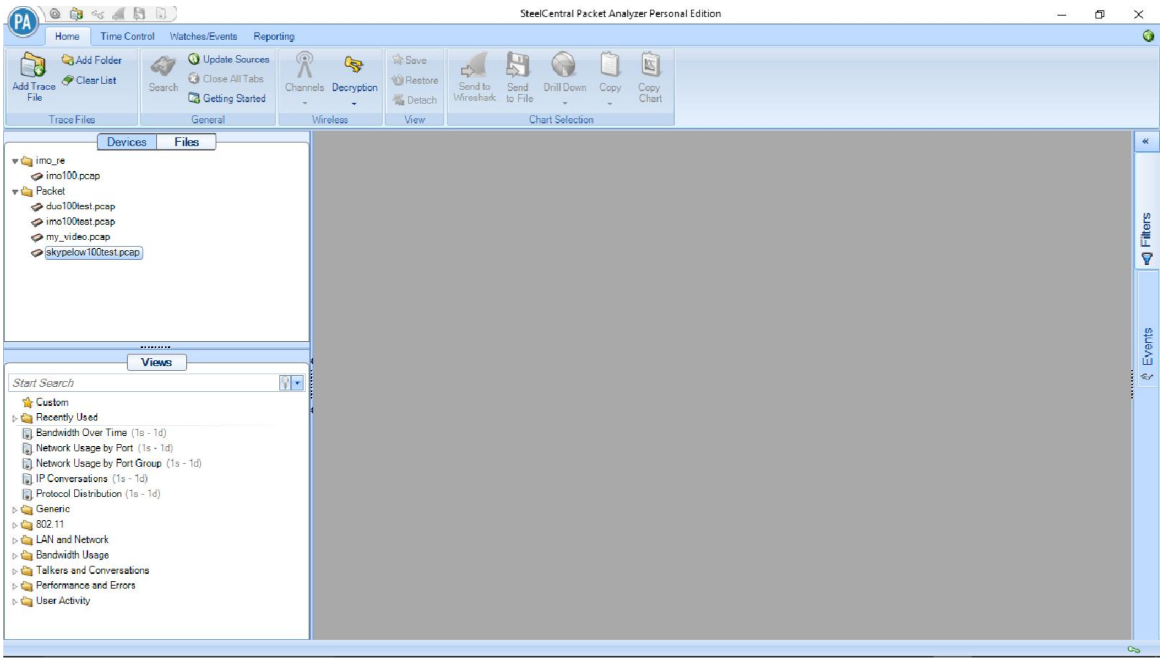

Figure 1. 16 SteelCentral Packet Analyzer Window

1.7.4 Android Studio

To code the application for this research Java as a programming language is used. Android application is developed using android studio [20] Figure 1. 17. There are built is features and GUI supports to develop mobile application with it.

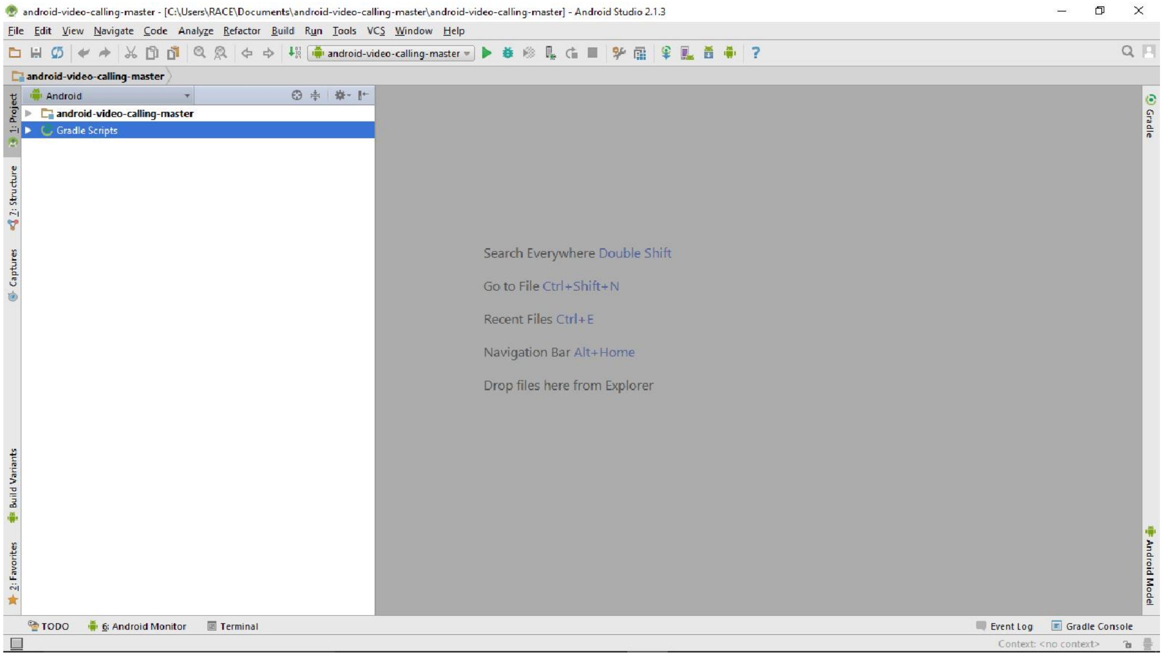

Figure 1. 17 Android Studio Window

CHAPTER 2

REVIEW OF LITERATURE

Daldal et al. [21] investigates WebRTC and found that mostly WebRTC provide peer-to-peer communication. But there are some case where video conferencing is done so they proposed how video conferencing can be achieved on signalling layer with the use of RESTful Web Services.

The solution they proposed is that instead of carrying media description to the endpoints carry a pointer of the actual media. This pointer contains a link of RESTful Web Services which provide the media description to the endpoints when communication is established. Both endpoints communicate with Web Service to make a call and get the media description understanding it as WebRTC endpoint.

Wuttidittachotti et al. [22] discuss two mathematically models to measure the quality of VoIP. In the first method which is simple too is conducted by informal interview with 180 Thai native participants. Skype is chosen for the test with SILK codec. They performed test at different packet loss rate. Network simulation is used for this test to generate packet loss. In second method subjective results are subtracted from computed results using Simplified E-model Formula. This is tested on different group of 36 native Thai participants using MAPE (Mean Absolute Percentage Error) technique. Factors like: packet loss, delay, and jitter are the key influential parameters to affects the quality on the user side. Quality of Experience (QoE) also differ from user perspective regardless of application or services. So, subjective measure or objective measure i.e. QoE is used in measuring the VoIP system. Subjective measure include: Conversation-opinion tests and Listening-opinion tests whereas non-intrusive methods and intrusive method comes under objective method. Both method provide higher accuracy and reliability. MOS (Mean Opinion Score) is used for both methods to evaluate the quality at different packet loss rate.

Azfar et al. [23] performed a survey on Android VoIP apps to examine data security and privacy. They consider ten VoIP apps (Skype, Google Talk, ICQ, Yahoo, Nimbuzz, Viber, WebChat, Fring, Tango, and Vonage) to perform experiment to test encryption of voice/text in different communication channel (Wi-Fi network to Wi-Fi network, mobile network to mobile network, mobile network to Wi-Fi network, and Wi-Fi network to mobile network). They take sample of songs and captured the network traffic. With the help of histogram analysis and entropy distribution he analysed the captured traffic. They categorise the voice/test as encrypted if bytes are uniformly distributed with no cluster and steady entropy distribution with no sudden change in entropy else not encrypted.

Sredojev et al. [24] discuss about WebRTC technology and how to implement it. Signaling and protocol method used in WebRTC is not specified so they design and implement signaling by using WebSocket. WebRTC server is implemented with WebSocket server and client with WebRTC API to perform the communication. Implementation of signaling with the help of WebSocket avoid redundancy and maximize compatibility with the technology established. This method opens peer-to-peer communication session until communication is closed. Four message control method is proposed: initialize, initiator, got user, and peer channel. First initialize message is send by peer1 to the server and waits until peer3 reply. Peer2 reply and look if it is an initiator or not. If it finds it is not an initiator it gets peer channel and then got user media. Finally both sides come to establish a peer-to-peer communication.

Yoon et al. [25] focus on voice packet quality on wireless network (wireless broadband and high speed downlink packet access) on mobile platform. Degradation in radio channel result in delay resulting in quality degradation of voice service. For measuring the voice quality they use software that they have developed in research [26]. With the help of measurement function as in the software RF quality and VoIP quality is calculated. Different network parameters: one-way delay, packet loss ratio, and jitter are taken into consideration. For quality of VoIP E-model [27] is followed that derives R-Score and mean opinion score (MOS) with the help of obtained values. They consider two codec: G.711 and G.729 in the test. Stress test is performed to measure RF channel quality which is the cause of voice quality degradation. To adjust the radio channel emulator is used. Different results are obtained and compared between WiBro and HSDPA network.

Yu et al. [28] conduct a survey on popular video chat apps: Face-Time, Google Hangout, and Skype over both Wi-Fi and cellular network. They study encoding and decoding techniques, smoothness of video over various wireless network, video quality under different network condition and architecture that provides quality of VoIP call. All these apps uses proprietary protocol and encoding technique to encode media so, they use reverse engineering as a design as discuss in [29] and compare their performance over various wireless network. According to this test, they conclude that Google Hangout only offer multi-party conferencing on mobile platforms. They perform communication between wireless user (3G or Wi-Fi) and wired PC user. Communication is sniffed with the Wireshark at both ends. They use iPhone for Face-Time and perform jailbreak to install third party applications. With the help of remote login via SSH they run tcpdump on iPhone to capture the packets. This experiment is done at different location in different network conditions. Study has been done on a variety of network elements: delay, packet loss, bandwidth for all the applications and comparison has been done.

Johnston et al. [30] discuss the issues that arise with the use of WebRTC for enterprise purpose. Some of the issues are: firewall, traversal access control, peer-to-peer data flow, recording, logging, policy enforcement, integrating with existing communication infrastructure. It is easy for the developer to develop consumer application and use WebRTC by the means of simple JavaScript on website to enhance the feature of the site with communication. But for enterprise there are some issues in use of WebRTC. They explain different approaches used by enterprise firewall: Symmetric RTP, ICE (Iterative Connectivity Establishment) and SBC (Session Border Controller). Out of all these approaches SBC is mostly used. WebRTC has no Signaling channel standard and SBC uses Signaling channel to authenticate media channel. SBC learns when media flow starts and ends by adapting itself in the control path but WebRTC uses HTTP or WebSocket between browser and server over Transport Layer Security (TLS) which is encrypted. So, to use WebRTC one approach is to convert WebRTC session to SIP (Session Initiation Protocol) session and converting back to original on the other user side. This approach is practically not possible. In the rest of the paper they discuss about how WebRTC can cross enterprise firewall and be integrated with existing infrastructure.

Liu [31] discuss on VoIP network impact and QoE (Quality of Experience) under different network conditions. Study is done on Skype video calls to get rate control and video quality. They measure Skype video calls by controlling different network parameters: varying packet loss rate, propagation delay, and available bandwidth and observed its video rate, frame rate, sending rate and FEC redundancy. Network emulator is used to control various network parameters. Skype is working fine in mild packet loss and propagation delay. It utilize the available bandwidth more efficiently and is TCP friendly i.e. works fine with TCP traffic. It works even if network is overloaded i.e. congestion control is manage efficiently. During the research it is observed that Skype used codec: VP6, VP7, and VP8.

Zheng et al. [16] proposed an adaptive filtering algorithm called robust set-membership normalized subband adaptive filter (RSM-NSAF) which is the improved version of set-membership normalized subband adaptive filter (SM-NSAF). Use of new membership set error bound which results in improved against noise signal. For this purpose acoustic echo cancellation (AEC) simulation of the algorithm is designed to confirm the improvement of proposed algorithm. The SM-NSAF algorithm performs well in many application but its performance degrades in the presence of impulsive noises. Algorithm is proposed with the help of robust set-membership affine projection (RSMAP) and set-membership normalized least mean square (SM-NLMS) considering robust error bound.

As discussed about conventional NSAF algorithm and SM-NSAF algorithm. Computational complexity of proposed algorithm show increased in computational complexity then SM-NSAF algorithm but update ratio is decreased. Implementation of algorithm is done with AEC application system. The two unknown echo path of 512 with adaptive filter length is considered. Multiplication by -1 is done in the middle of the iteration for impulsive response echo path to track the capability of the algorithm. The Gaussian random sequence of first order and speech signals are used as an input signals. The output system is mixed with signal-to-noise ratio (SNR) of 30dB by Gaussian noise. The output is added with impulsive noise. Simulation is done over 50 trials and obtained results are compared with the SM-NSAF algorithm which shows algorithms performs better than the SM-NSAF algorithm in term of steady-state misalignment and robustness against impulsive noises.

Qadeer et al. [32] proposed a technique that can recover a true signal from the raw signal and also compare different other recovery techniques. Due to SNR and ergodic nature of raw signal a true signal is difficult to recover. Intrusion detection in surveillance system is useful to detect the unauthorized person in the area. One of the most famous perimeter intrusion detection system (PIDS) is widely used to detect any unauthorized entry in the area if detected an alarm starts as a response to alert the security person. But a smart intruder can enter the area as the PIDS technique is inconsistent in the events like: - fence cutting, crawling under, high wind, and heavy rain. To make more accurate detection algorithm technique for detection should be remap to evaluate stochastic and adaptive signal processing techniques. To study this large data is collected under different environmental conditions. Study of different noise and its properties, adaptive filtering techniques (like: - Wiener Filtering, Kalman Filtering) and stochastic tools like power spectral density and cross correlation is done. MATLAB is used to test the data with different filtering algorithms. Based on low SNR it is find out that Wiener filter gives the best results. The Wiener filter gives response time of 8 to 10 sec with 90% of probability of detection. In similar manner novel approach takes the raw data from the signal processing and estimated the energy which is then pass through band pass filter and then signal is amplified to improve the SNR value. This approach reduces the false alarm. Case study is done with 120 different cases under three algorithm techniques (LC, Wiener and Novel) in which novel approach works best with 116 times detection events in 1-3 sec response time with 95% of probability of detection whereas LC gives poor results among all algorithms with increased false alarm

Jianjiang et al. [15] illustrates that digital devices are prone to strong radio frequency (RF) signal. So this type of RF signal must be cancelled out before the signal is received by the devices. Signal can be composed of two orthorhombic signals. Algorithm is proposed that can cancel out strong RF signal. For this purpose Analog circuit is used. This algorithm is based on Least Mean Square (LMS) adaptive filtering algorithm. Every devices are disturbed by the electromagnetic signal present in the environment. When the receiving device is near to a strong RF signal, it blocks the signal that is to be received which is sent by the sender causing device failure. Some times when strong RF signal needs to be send the receiving device is close to make it work normally but both sending and receiving devices cannot work simultaneously. So, they uses Analog circuit or Analog devices (Like: - phaser, multiplexer) to build a strong RF cancellation circuit and uses LMS adaptive filtering to cancel the strong signal at the receiving end.

Since LMS can process digital signal so to cancel the strong RF signal they modify the algorithm. This does not reduce the performance of the devices but enhance the electromagnetic compatibility of end devices. Experiment setup has an antenna which accepts two signals with strong RF signal of 70MHz with power 23dB while useful signal of 70.5MHz with power -60dB. The attenuator is set to 0.2dB with sampling frequency of 350MHz. The phase difference is random between strong RF signal and reference signal.

MATLAB is used to plot the results when phase difference is 35º. This system has the ability to cancel up to 45dB. Its ability is always above 35dB. Circuit can changed with different demands

Madhavi et al. [13] proposed a novel approach for voice quality enhancement for real time VoIP application. Background noise reduction (BGNR) and Acoustic Echo Cancellation (AEC) are the main concern for this approach to enhance the speech quality. With the help of LMS algorithm they cancel out BGNR and with the help of NLMS algorithm they remove adaptive AEC by the help of ADSP BF533-EZ KIT processor and experiment is simulated with the help of MATLAB. Figure 2. 1 shows the experimental approach for the test.

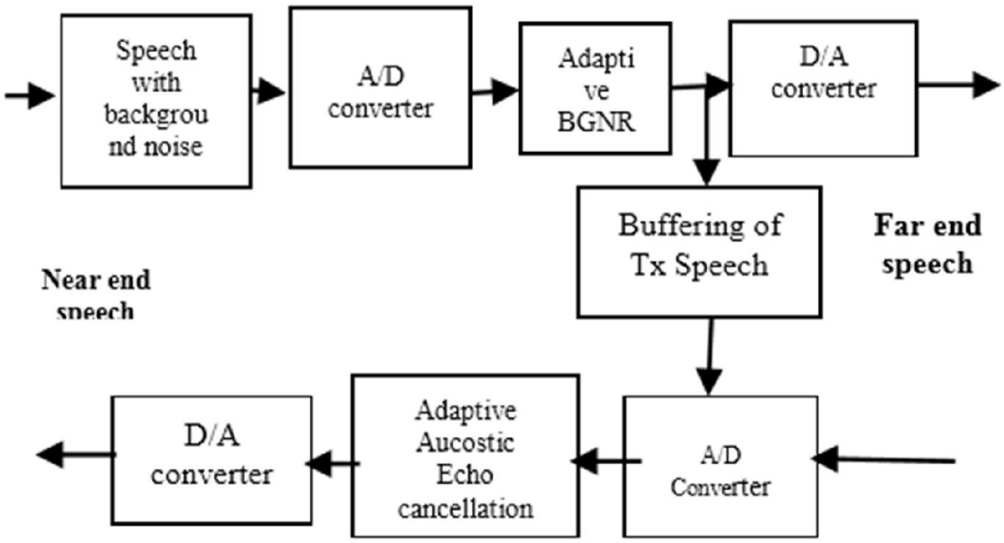

Figure 2. 1 Block Diagram of Experimental Approach

To process the speech enhancement noise is filtered out before sending for transmission as well as before receiving. To remove background noise DSP like hardware based headphones is used for good solution before sending the CODEC. The acoustic echo is the cause of loudspeaker of the phone. The sound from the speaker bounces the objects present in the room like wall, ceiling, fans, etc. and again receive by the microphone causing noise. This type of noise is created anywhere anytime like in car and also in the case of hands free. In the experimental approach speech with background noise at near end speech is converted to digital signals by the use of A/D converter. Adaptive BGNR is applied to the obtained signals and the final with D/A converter obtain the signal at far end speech without background noise. Buffering of the signal after applying BGNR with the help of A/D converter to get the digital signal from the far end speech. AEC is applied to the obtain signal and finally by D/A converter the noise reduced signals is received. DSP processor is used in this process to get the good speech quality.

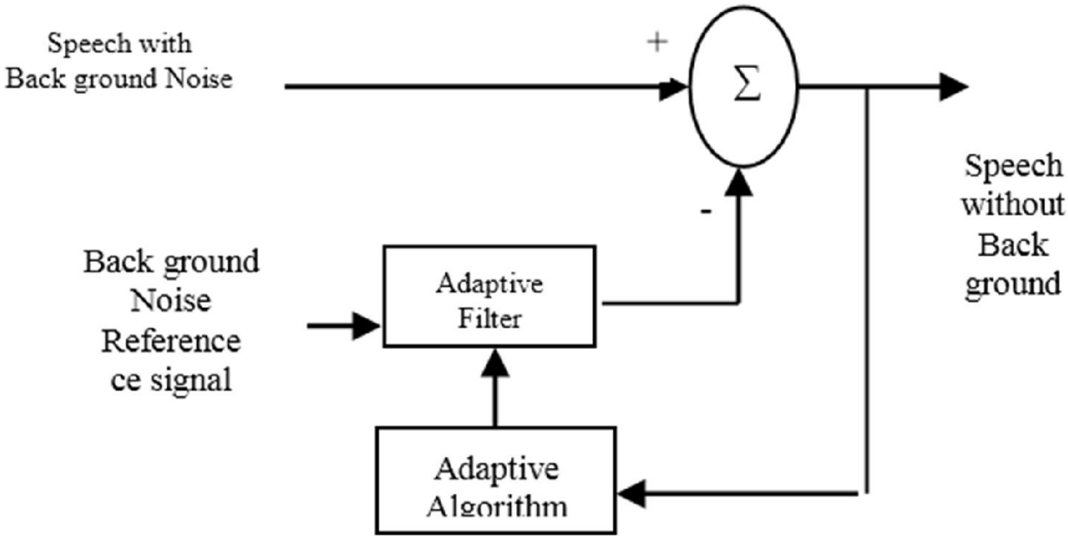

Figure 2. 2 Block Diagram of ABGNR

Figure 2. 2 shows how background noise removal setup is prepared. Signals with background is added with reference signal with background noise the adaptive filter LMS algorithm is applied to which after optimization of error signal without background noise is obtained. This method is call Adaptive Background Noise Reduction (ABGNR).

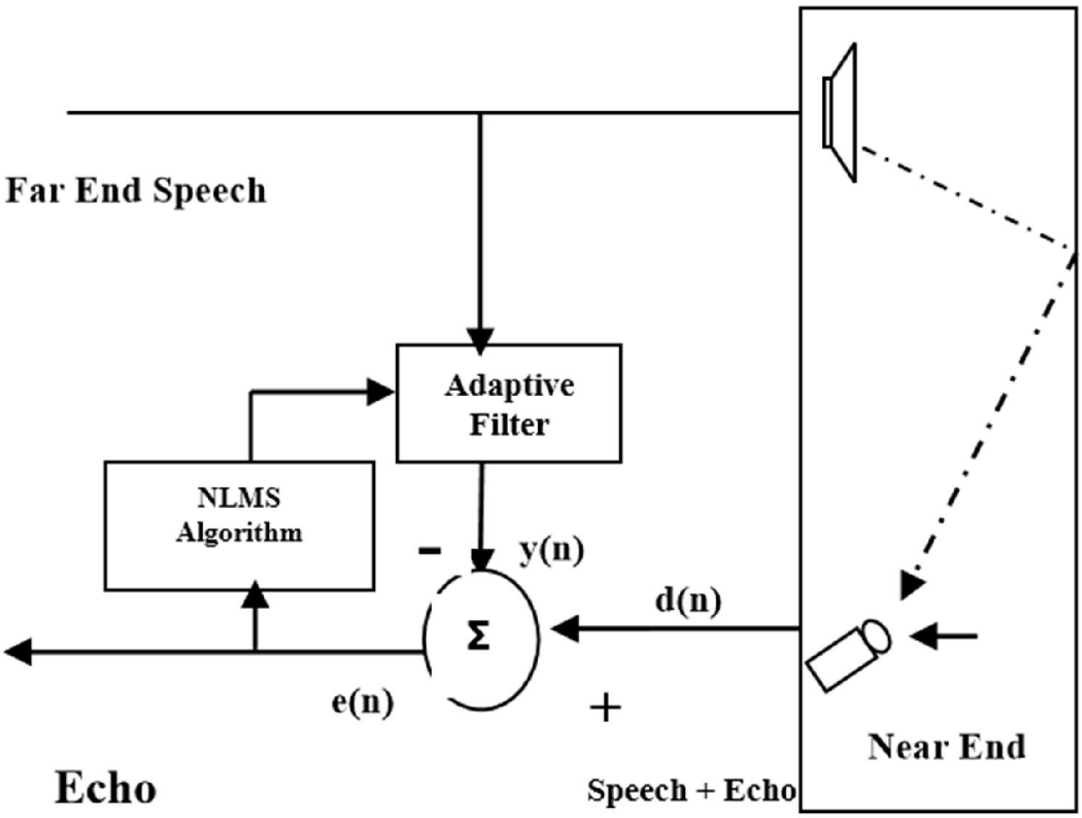

Figure 2. 3 Block Diagram of AAEC System

In case of Adaptive Acoustic Echo Cancellation (AAEC) system Figure 2. 3 adaptive filter NLMS algorithm is used. It filter out noise created by phone speaker and microphone. To perform this experimental test video conferencing is establish. A sampling rate of 8 KHz with 400 filter coefficient is used to cancel the echo. Experiment perform well in different scenario as in Table 2. 1, Table 2. 2 and Table 2. 3.

Table 2. 1 Reduction in Noise Level from FAN in Background

Average Speech Signal Level with Noise in dB	Average Speech Signal Level without Noise in dB	Reduction in Noise Level in dB
-59.78	-66.66	6.88

Table 2. 2 Reduction in Noise Level from Talking People in Background

Average Speech Signal Level with Noise in dB	Average Speech Signal Level without Noise in dB	Reduction in Noise Level in dB
-77.26	-98.78	21.52

Table 2. 3 Reduction in Noise Level from Air Conditioner in Background

Average Speech Signal Level with Noise in dB	Average Speech Signal Level without Noise in dB	Reduction in Noise Level in dB
-47.82	-47.93	0.11

Results show that there is a noise reduction of 21.52 dB in case of people tacking in background during the conferencing whereas in case of fan as a background noise 6.88 dB noise reduction is observed and 0.11 dB is observed in case of air conditioner. Voice quality is improved with a delay of 200ms.

Kadam et al. [12] uses adaptive filtering technique to cancel out echo. Adaptive filter like LMS, NLMS, RLS and dual-H are used to enhance the quality of VoIP over the network. Simulation is done with the help of MATLAB and they tried to remove background noise and tested the effect of packet drop, delay and dual talk of VoIP calls. Studied have been done on echo types, echo generation, echo measurement and various echo cancellation algorithms.

Least Mean Square (LMS) which is based on Wiener filter is widely used in many application. Initially weights are assigned with zero.

$$w(m) = 0 \tag{6}$$

Output of the filter is given by

$$\mathrm{y(m)} = \mathrm{W}^{\mathrm{T}}(m)x(m) \tag{7}$$

Error is calculated by

$$e(m) = d(m) - y(m) \tag{8}$$

Weights are updated on the basis of step size (μ), error and input signal as

$$w(m+1) = w(m) \pm \mu e(m)x(m) \tag{9}$$

Value of step size is $0 \leq \mu \leq 1$. Step size is selected manually which is the main disadvantage of this algorithm. If the step size is small system converge very slowly and if it is large it never converge.

Normalize Least Mean Square (NLMS) is the improvement of LMS algorithms. Weights are updated as

$$w(m+1) = w(m)\frac{\mu}{\mathcal{E} + x^T\,(m)x(m)}\,e(m)x(m) \tag{10}$$

$\mathcal{E}$ is the constant that avoids division by small number. $x^T\,(m)x(m)$ is the power of $x(m)$. So that noise will not occur as in case of LMS. Complexity is same as LMS filter but stability is high with increase in computational cost.

Recursive Least Mean Square (RLS) repeatedly updates the filter weight to minimize the cost.

$$c(m) = \sum_{i=0}^{m} \Lambda^{m-i}\, e^2(i) \tag{11}$$

Value of Λ is between $0 \leq \Lambda \leq 1$ called forgetting factor. It gives more weight to the new sample than old sample. It is more computationally complex than LMS and NLMS so it is not used for high order filter.

Double talk is generated when both parties speak at the same time. This is the main problem in AEC. To reduce the double talk step size is reduced but downside is that for all the time we have to use small step size and quality of voice degrades. To overcome this problem we can use double talk detector that block the double talk. Dual-H is used for this process. Normally VoIP signals are converted to PCM signals by codec and compressed as in Figure 2. 4. AEC block Figure 2. 5 is applied near the far end VoIP gateway to cancel far end echo but this results in delay of the packets at the receiver side. Second solution is to place near the end gateway that filter the echo at the receiver side. All the simulation are done using MATLAB. Results shows length of adaptive filter gives better results i.e. better echo cancellation but performance gain decreases. RLS filter reduces complexity so better option is NLMS filter for large filter length. Increasing adaptive filter length increases the packet delay. Hence choice of adaptive filter algorithm vary with the application over VoIP network.

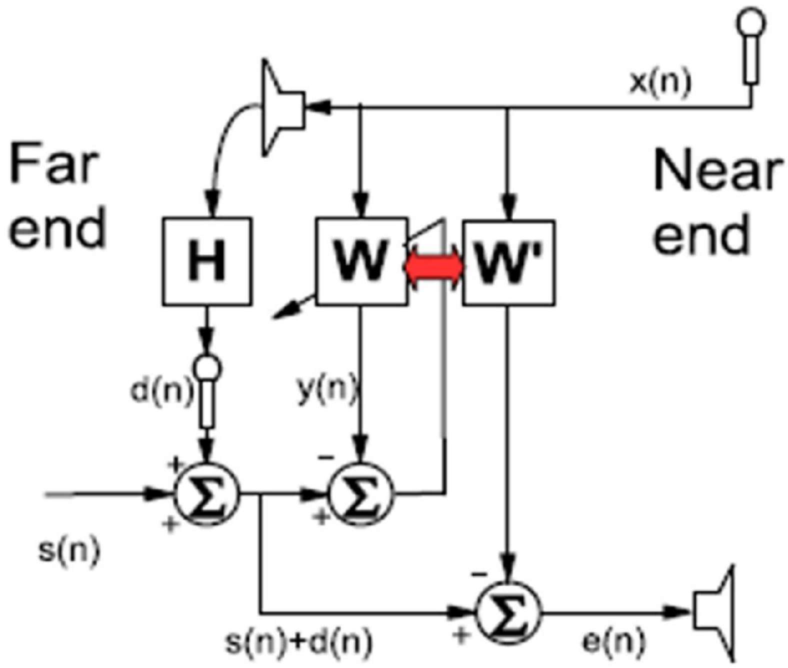

Figure 2. 4 Dual-H System

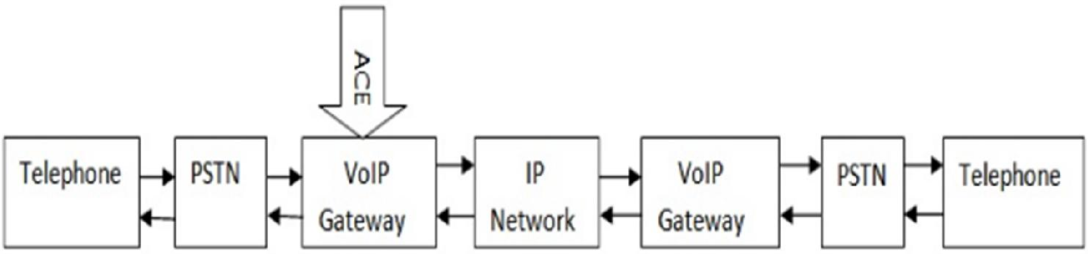

Figure 2. 5 AEC Block in VoIP

Gupta et al. [33] uses different adaptive filtering algorithms (like: - LMS, NLMS and UNANR) to analyse the performance of speech by enhancement in the noisy environment. Every time whenever the communication occur speech signals are buried or degraded by the environmental noise to get good quality of speech different enhancement techniques are used to reduce the noise. They uses adaptive filters to analyse the performance of different algorithm in the noisy environment. Experiment is simulated using MATLAB Simulink. To improve the SNR different algorithms are used along with the proposed Unbiased and normalized adaptation noise reduction (UNANR) algorithm. Response filter of UNANR is given by

$$f(n) = \sum_{m=1}^{M} W_m(n) r(n-m+1) \tag{12}$$

Where $W_m(n)$ denotes UNANR coefficient, $r(n-m+1)$ denotes reference input noise at $(m=1)$ and$(m-1)$, 1≤ m ≤ M, input sample. UNANR coefficient is normalized as

$$\sum_{m=1}^{M} W_m(n) = 1 \tag{13}$$

To estimate noise present in signal UNANR coefficient is modified as

$$W_k(n+1) = \frac{W_k(n+1)}{\sum_{k=1}^{M} W_k(n+1)} \tag{14}$$

Simulating the results of all the algorithm in the noisy environment to remove noise using MATLAB with white noise added to input songs shows that NLMS estimates more near to the actual signal even in the noisy environment and SNR value is also improved. UNANR algorithms perform well which ranks second but LMS algorithms gives poor performance.

Sharma et al. [14] discuss about noise present in Electrocardiogram (ECG) signals. To get the accurate output from the ECG noise reduction with varying input is necessary so he uses adaptive filter like LMS that can adapt any input signal to get low Minimum Mean Square Estimator (MMSE). Other adaptive filter can also be used but LMS gives the best results with minimum MMSE comparing to other algorithms. Simulation of the test experiment has been done in MATLAB. ECG is used to measure the heart information i.e. it plays a vital role so ECG signals must be noise free to good information. Heart information is displayed as an electrical waveform in the ECG. This waveform contains two waveform one embroyo waveform $g(n)$ and heartbeat waveform $h(n)$. $h(n)$ is of the type noise present in the primary input signal Figure 2. 6. So to get pure embroyo waveform $g(n)$ it should be filtered. To do this narrow band and broad band noise are added which is filter by adaptive filter noise canceller system and improved SNR is obtained. ANC system consists of two inputs (primary input and reference input). The primary input signal is disturbed by noise (n). The noise (n_o) is passed through adaptive filter to get the output signal and error is estimated. All simulation are done in MATLAB. Process is repeated up to 15 taps with step size 0.00007, 0.00004, 0.00001, 0.00020, 0.00500, and 0.01 and find that with increase in step size noise also increases and rate of convergence is also increased.

Even with small step size rate of convergence increases and MMSE in decreased as with step size 0.00001.

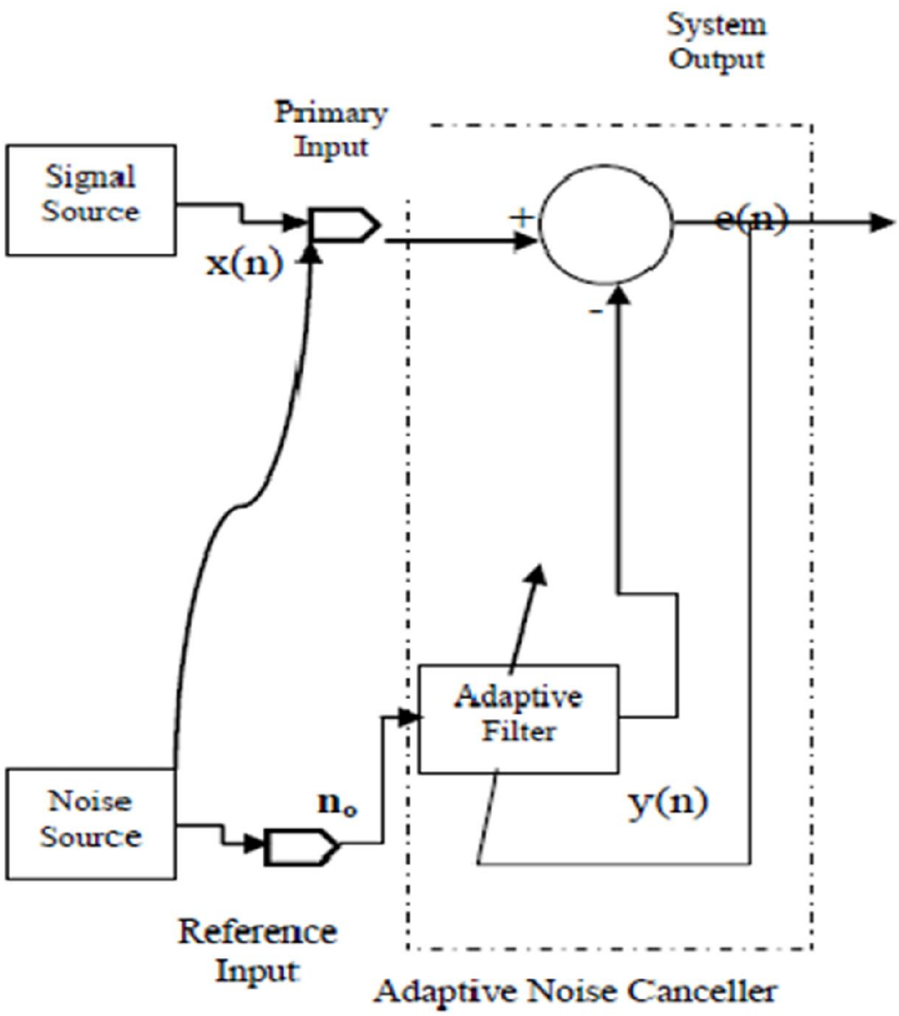

Figure 2. 6 Noise Cancellation System

Daengsi et al. [34] discuss about accuracy and quality enhancement of VoIP calls so survey has been conducted with the Thai users. A subjective method is more authentic way to get the good results so this method is used. In this way accuracy of simplified E-model has been improved and it is then compared with original E-model for codec G.729. This method is called subjective MOS estimation model which is improved and reliable technique for VoIP quality evaluation. Quality of evaluation (QOE) is the main concern for the evaluation of the VoIP call quality test. From Figure 2. 7 we can see that VoIP quality measurement methods are divided into two main categories: - subjective method and objective method. Subjective method is further divided into three types: - conversation opinion tests, listening-opinion test and interview survey test. In all these methods users are given to test the calls and give review based on the quality of the voice and video to find out the MOS and R value to evaluate the quality of VoIP call in the test. Objective method is divided into two types: - non-intrusive and intrusive method. These methods are based on mathematical model to test the VoIP call quality. They choose subjective methods with the help of 220 Thai users to evaluate the VoIP call.

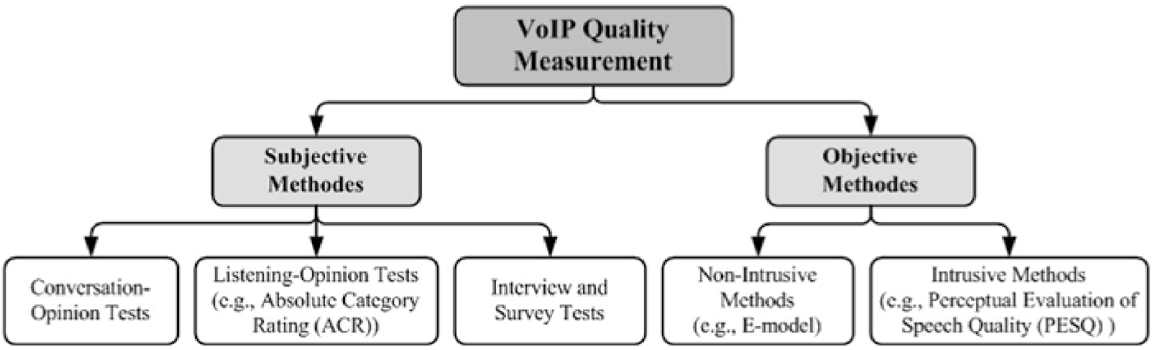

Figure 2. 7 VoIP Quality Evaluation Method

Original E-model for evaluation of VoIP call quality is complex so subjective method also called simplified E-model has been proposed. Results obtain by setting 0-400ms delay with packet loss 0-10%. MOS value is evaluated with Thai users using curve fitting tool in MATLAB as

$$MOS_{G.7\ 29} = 4.056 - 8.164X - 0.06935Y + 15.85X^2 + 0.2415XY \qquad (15)$$

Where X is packet loss rate and Y is packet delay.

Simplified E-model for this evaluation is given in Figure 2. 8 which shows VoIP quality test evaluation has been done by both subjective method as well as objective method. In case of subjective method conversion opinion test is taken as type for the evaluation of MOS value. MOS value is used for the calculation of R value. In case of objective method E-model is used to find out the R value by simplified E-model. R value of Simplified E-model and subjective method is subtracted to get the bias equation which then added with MOS value of simplified E-model to get the enhanced MOS value for VoIP call quality for codec G.729. In this way performance is improved but it is limited to packet loss 0-10% with packet delay 0-400 ms only.

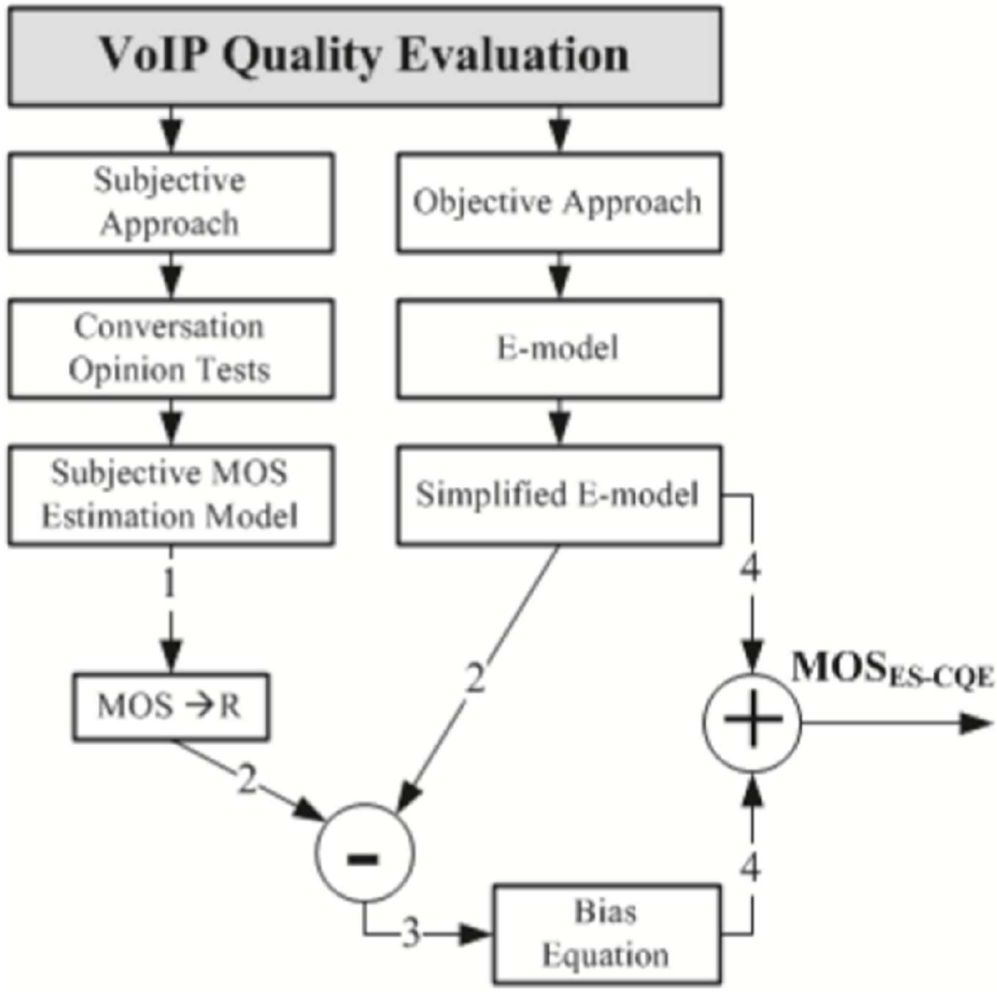

Figure 2. 8 Simplified E-model for VoIP Evaluation

CHAPTER 3

SCOPE OF THE STUDY

The scope of doing this research is to find out the issues in the VoIP chat applications and provide with an enhancement solution with the help of adaptive filtering algorithms. This research can be done by any one with some networking knowledge. It helps the people to get better VoIP calls even in the slow internet connection and also uses less bandwidth. This research has been done considering the scenario that no any other applications other than participating VoIP apps are utilizing the bandwidth. Here one need to provide the same bandwidth to the both users which is not always the case in real world scenario i.e. we can say to perform this experiment near ideal conditions is used.

CHAPTER 4

PROBLEM FORMULATION

A good VoIP application under slow network condition is considered as main demand in the today scenario so that even in the ruler area where maximum time internet connectivity is slow video calling can be possible. In this paper research has been done considering popular video calling application (Google Duo, Skype and Imo) that runs on android phones. It comes as a following results: -

- Although there are many users that are using Imo as their preferred video calling application but it poor video quality and speech quality even under good bandwidth.
- Under my test condition that has been discuss in CHAPTER it is observed that Imo video chat application works very poor and many times it freezes with a lot of packet loss.
- Google duo even under this condition works well it pauses the video whenever the proper video is not able to display allowing the speech in the conservation.
- Skype also performs good ranks second in the test but sometimes it disconnects as proper bandwidth is not given.

All these applications have their own mechanism to handle the noise which are proprietary. In this paper an open source WebRTC video calling architecture is used to build the android application to test the video calling and compared results with the existing popular video calling application (Google Duo, Skype and Imo).

CHAPTER 5

OBJECTIVES OF THE STUDY

Voice/video calling is growing rapidly with the use of smartphone and emerging data service (4G/3G). This research is about enhancing the quality of video calling over the internet by removing the noise with the use of adaptive filtering technique. The main objective of this study are:

- To study and analyse the various VoIP application for voice/video calling
- To optimise the voice/video codec and video conferencing using WebRTC over mobile platform
- To remove noise and improve quality of VoIP with adaptive filter (NLMS algorithm).
- To implement the filtering algorithm by developing an application and compare the results with existing VoIP applications.

CHAPTER 6

RESEARCH METHODOLOGY

6.1 METHODOLOGY I

For the purpose of this study following are the requirements:

- Two Android smartphones
- Internet hotspot (Wi-Fi) and a router (to control bandwidth)
- Wireshark and SteelCentral Packet Analyzer
- MATLAB

6.1.1 Experiment Setup

For the purpose of experiment any one of the Android smartphone is rooted. The whole experimental setup is shown in Figure 6. 1.

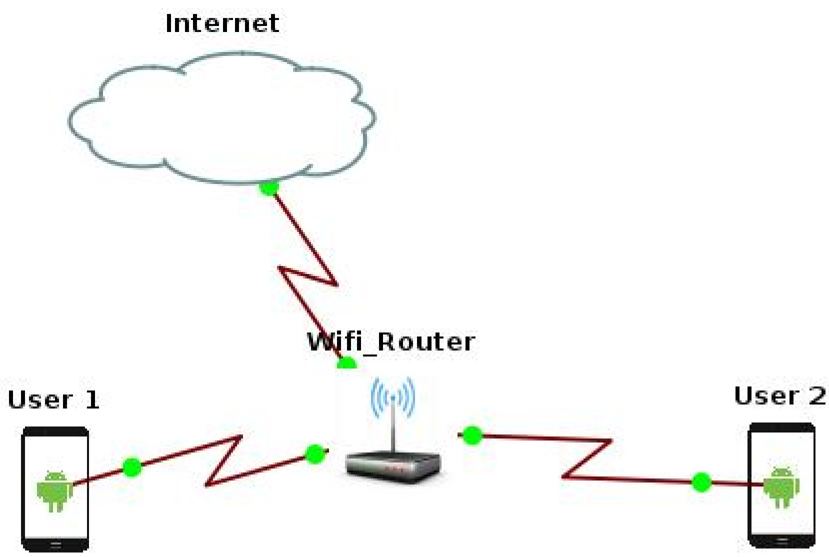

Figure 6. 1 Experimental Setup

Rooted phone is used for sniffing the network traffic with the help of tcpdump. Wi-Fi router is used so that both the phone uses the same bandwidth. Router is configured as shown in Table 6. 1.

Table 6. 1 Bandwidth Control Rules

Description	Egress Bandwidth(Kbps)		Ingress Bandwidth(Kbps)	
	Min	Max	Min	Max
User 1	2	100	2	100
User 2	2	100	2	100
Network	2048		2048	

After configuring the setup VoIP app was run on both the phone and communication was established. For all three apps network traffic was captured by the phone in pcap format. Other apps on the phone have been restricted to established network connection to avoid any bandwidth use. The captured file was analysed with the help of Wireshark and SteelCentral Packet Analyzer. Based on the obtain traffic we draw graphs with the help of MATLAB.

6.1.2 Experiment Process

To perform this experiment two Android smartphones with Android version 5.1. Step by step process for performing this experiment is discussed below:

- Step 1: Reset your both phones
- Step 2: Root the phone: I have used SuperSU_v2.49 for rooting my phone. Any method can be used to root the phone
- Step 3: Now install BusyBox, Terminal Emulator, and VoIP apps (Only one VoIP app at a time for the experiment).
- Step 4: Now download the tcpdump file from the internet.
- Step 5: Open the Terminal Emulator application and type the following commands:

  - su
  - mount -o remount, rw /system

- Step 6: Now remove any tcpdump file present at location /system/xbin with following command:
    - cd /system/xbin
    - rm tcpdump
- Step 7: Now copy the tcpdump file from download location to /system/xbin with following command:
    - cp /sdcard/tcpdump /system/xbin
    - chmod 777 tcpdump
- Step 8: Now open the installed VoIP app on both phones and login via with phone number
- Step 9: Check for VoIP communication by making a VoIP call. If call is successful disconnect the call. If any error occur check for the solution.
- Step 10: Make sure that all other apps are restricted from accessing the internet to avoid any bandwidth sharing among various apps.
- Step 11: Now back to Terminal Emulator and type the following command to start capturing the traffic
    - tcpdump –nnvvSXs 0 –w /sdcard/t_file.pcap
- Step 12: Start the VoIP communication
- After the communication is done stop the tcpdump with command: volume key + z
- This process is repeated for all the VoIP applications.

6.2 METHODOLOGY II

After analyzing the existing VoIP application it is found that echo is the main problem for quality of the VoIP calls so following steps are followed:

- Echo cancellation code is developed with the help of MATLAB.

- The code is then converted in to jar package.
- Android application is developed and jar package is integrated in it that gives provide VoIP calls support.
- After the successful development of the application the application is again tested using METHODOLOGY I.
- Generated results are compared to see whether enhancement has been made or not which is discussed in CHAPTER 7.

CHAPTER 7

RESULTS AND DISCUSSION

One by one VoIP app was used for the communication and their network traffic was captured in a pcap format. To analyse the pcap file Wireshark and SteelCentral Packet Analyzer is used. The obtained data was saved in excel and then used for generating graphs with the help of MATLAB. Bandwidth, packet loss, and frame size was analysed for the existing android application as well as developed application.

7.1 BANDWIDTH OVER TIME

Maximum amount of data that can travel through a channel is termed as bandwidth. Bandwidth in case of VoIP depends on codec used in voice/video. It depends on payload size, IP header, UDP header, RTP header, and header compression technique. It is measured in Bits/sec or Bytes/sec.

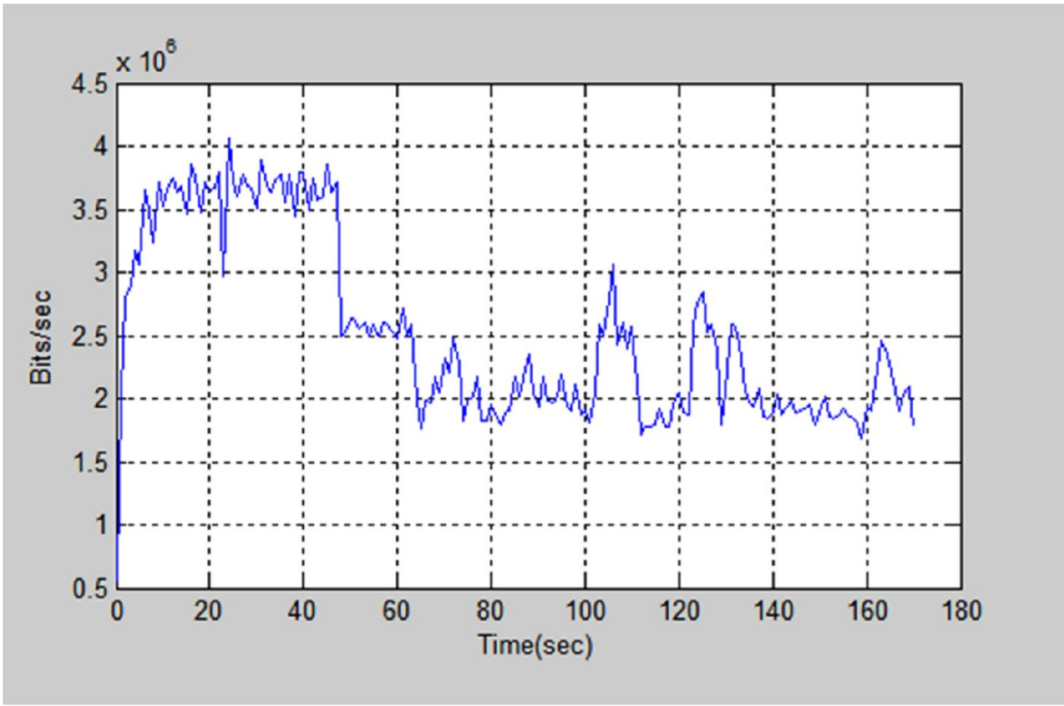

Figure 7. 1 Google Duo Bandwidth over Time

From Figure 7. 1 Y-axis represent Bits/sec whereas X-axis represents Time (sec). It is observed that between 0-50 sec bandwidth was high (between 3.5M bits/sec – 4M bits/sec) due to uncontrolled bandwidth but when bandwidth was controlled with the help of router it fell down and remained in between 1.8M bits/sec – 2.5M bits/sec from 60 sec onwards.

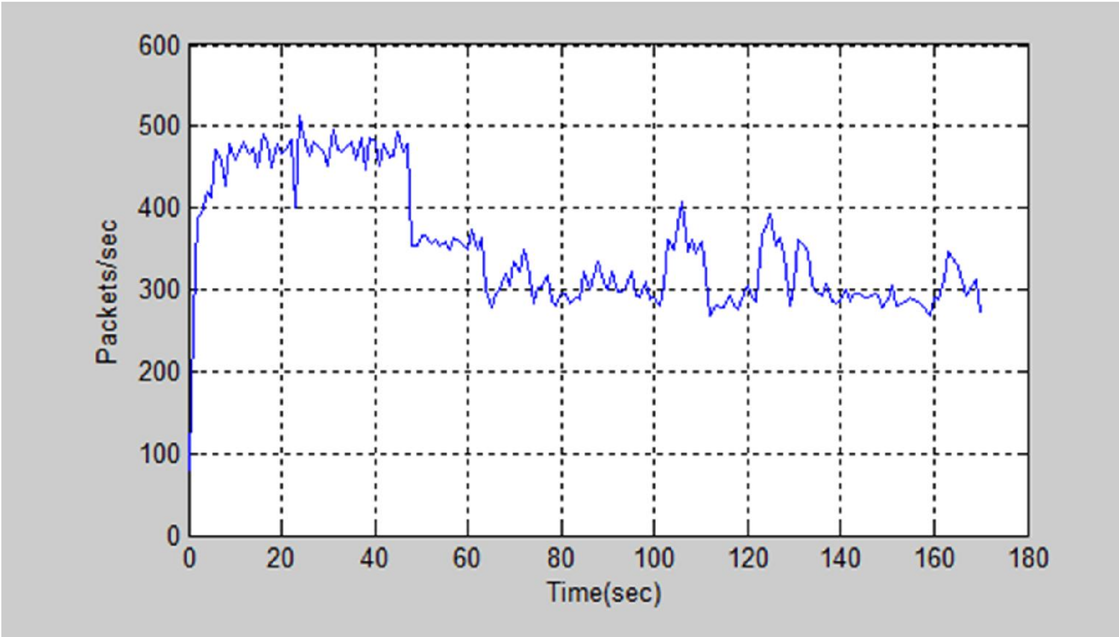

Figure 7. 2 Google Duo Packets over Time

Y-axis represent Packets/sec whereas X-axis represents Time (sec).Packets that was observed in the communication was in between 280 packets/sec to 400 packets/sec from 60 sec onward as in Figure 7. 2.

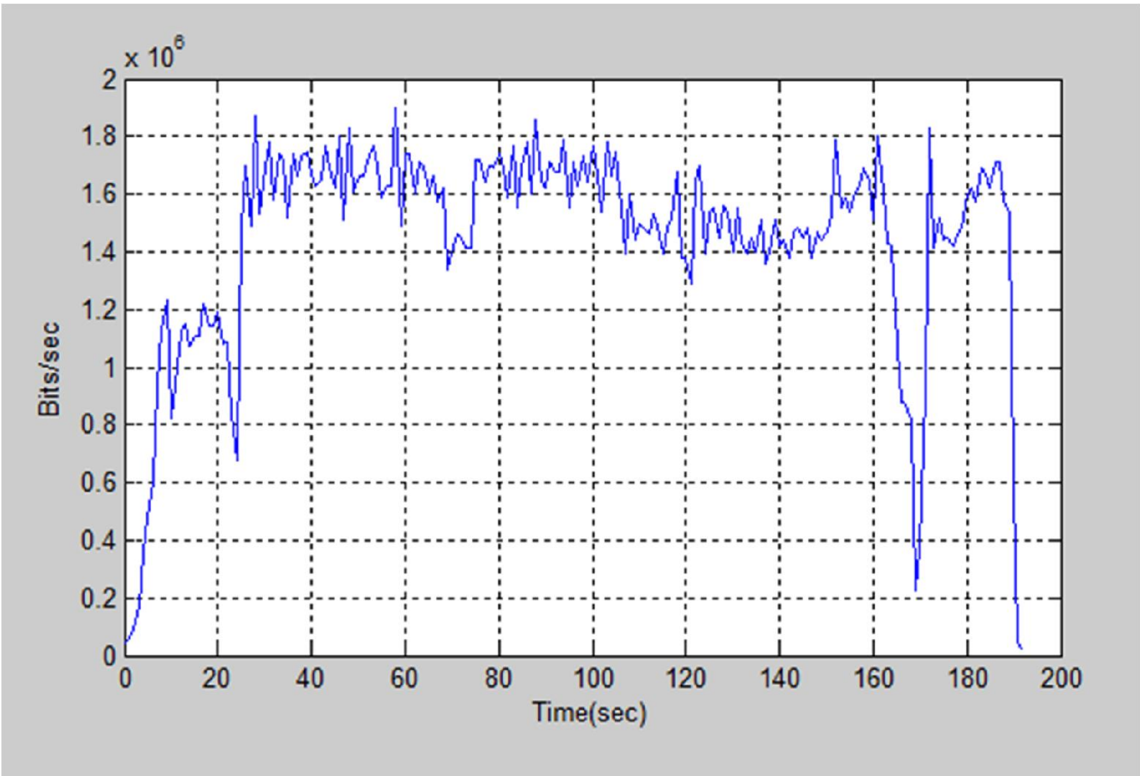

Figure 7. 3 Skype Bandwidth over Time

From Figure 7. 3 Y-axis represent Bits/sec whereas X-axis represents Time (sec). It is observed that bandwidth utilization of the Skype is lower than Google Duo i.e. between 1.4 M bits/sec to 1.8 M bits/sec with certain bandwidth drop in between. Most of the time bandwidth is between 1.6 M bits/sec to 1.8 M bits/sec.

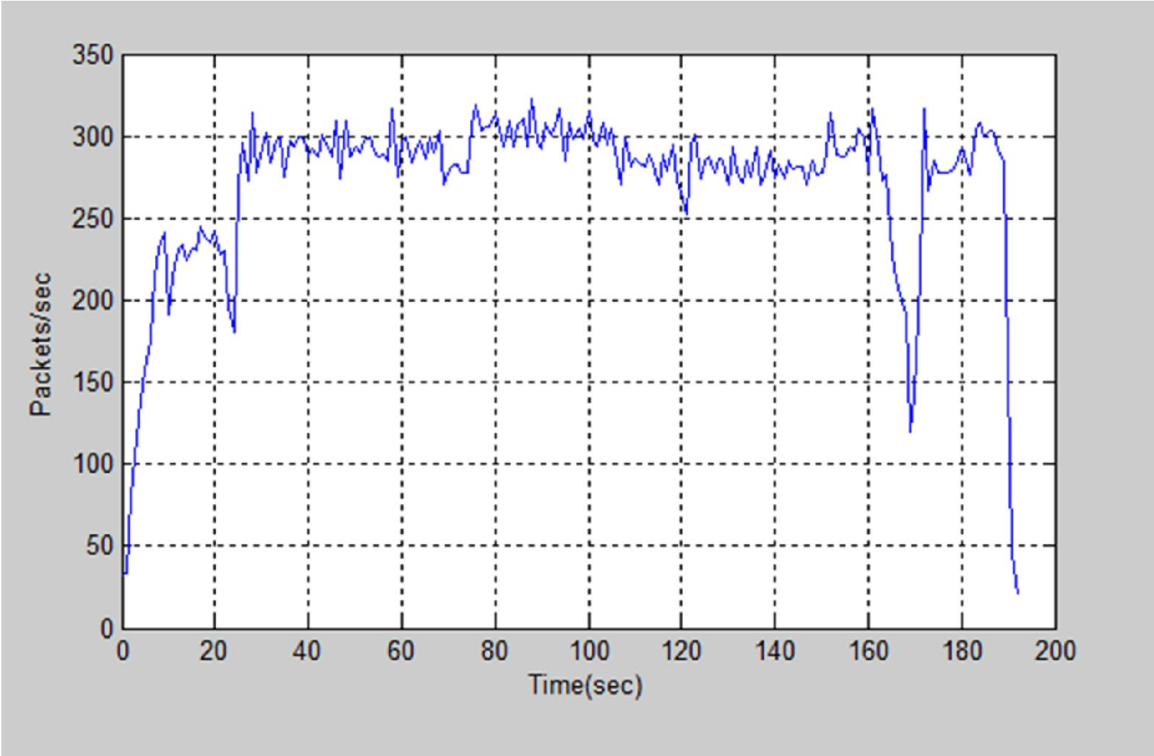

Figure 7. 4 Skype Packets over Time

Figure 7. 4 Y-axis represent Packets/sec whereas X-axis represents Time (sec). It shows packets over time for Skype VoIP call which is ranging from 250 packets/sec to 300 packets/sec.

In this study Imo application perform poor quality in VoIP calling shown in Figure 7. 6 which shows packets as lost with chatting is done with very high noise. Even on seeing the packets of Imo as in Figure 7. 5 we can say Imo video chat is poor in the given bandwidth scenario.

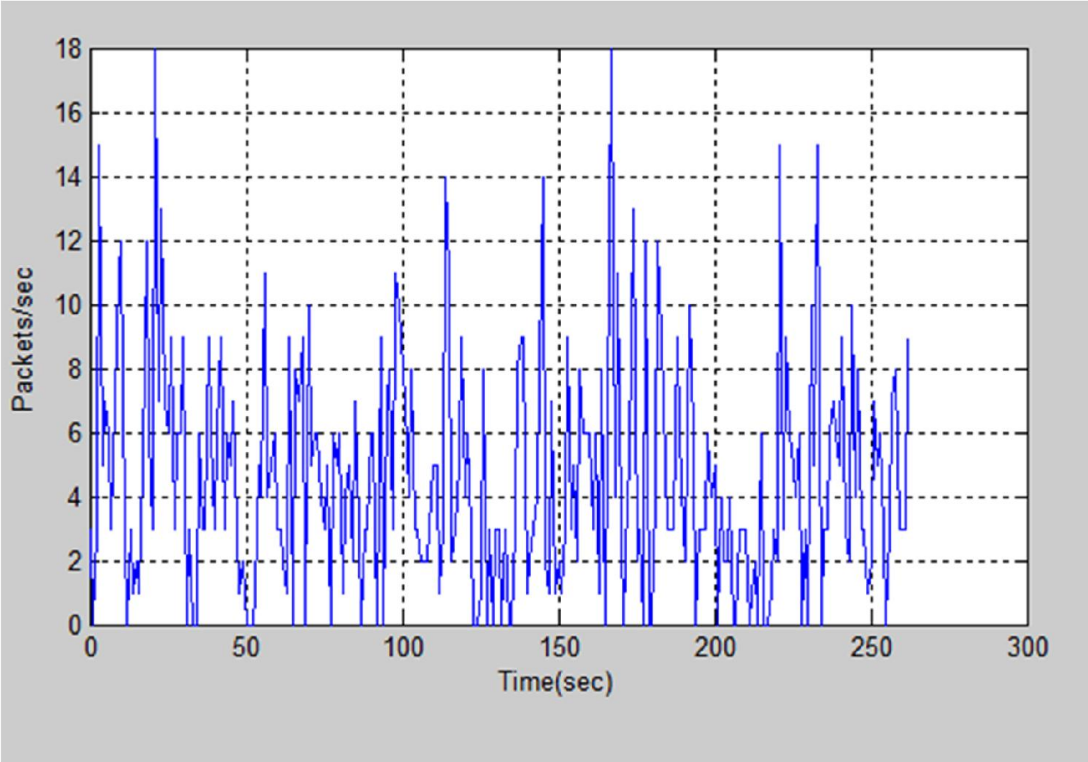

Figure 7. 5 Imo Packets over Time

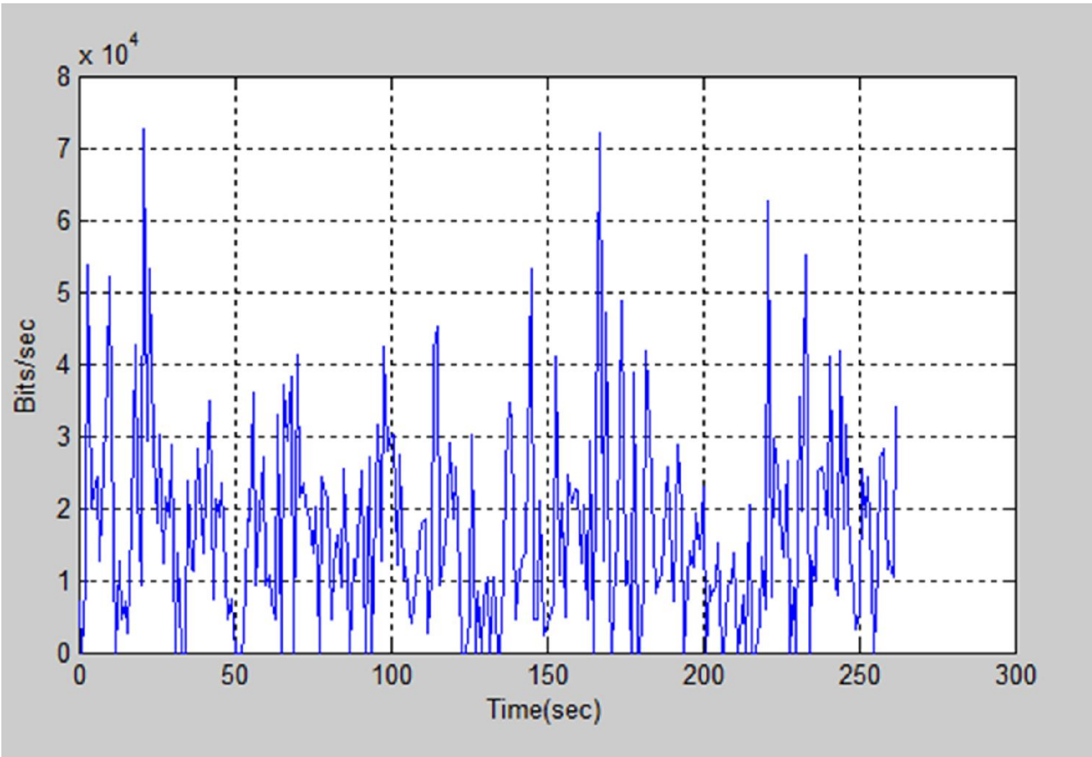

Figure 7. 6 Imo Bandwidth over Time

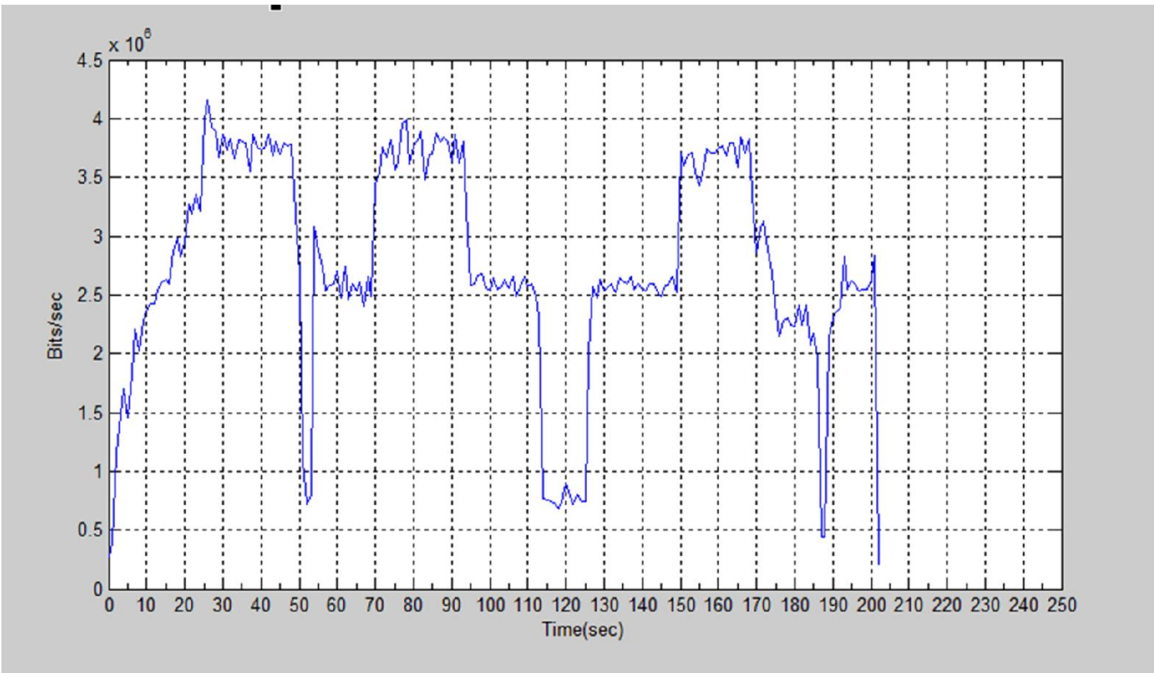

Figure 7. 7 My Video Bandwidth over Time

Developed VoIP when test with METHODOLOGY I results were obtained. Bandwidth graph Figure 7. 7 has X-axis as Time (sec) and Y-axis as Bits/sec shows that initially the bandwidth starts to increase from 0 bits/sec and remain in between 3.5 M bits/sec – 4 M bits/sec from 20 sec to 50 sec and then again it drops to 2.5 M bits/sec to 3 M bits/sec for 20 sec. Again bandwidth increases and remain in between 3.5 M bits/sec to 4 M bits/sec for next 25 sec (from 70 sec to 95 sec) and this seems to continue for rest of the conservation.

Figure 7. 8 shows X-axis as Time (sec) and Y-axis as Packets/sec for developed VoIP application. It is observed that packets/sec initially increases and remain in between 500 packets to 550 packets from 25 sec to 50 sec and then it falls down to 350 packets to 400 packets from 55 sec to 70 sec. Again it rises to 500 packets to 550 packets for next 25 sec and fall down. This process repeats for rest of the conservation but there is a major fell of packets in 115 sec to 125 sec.

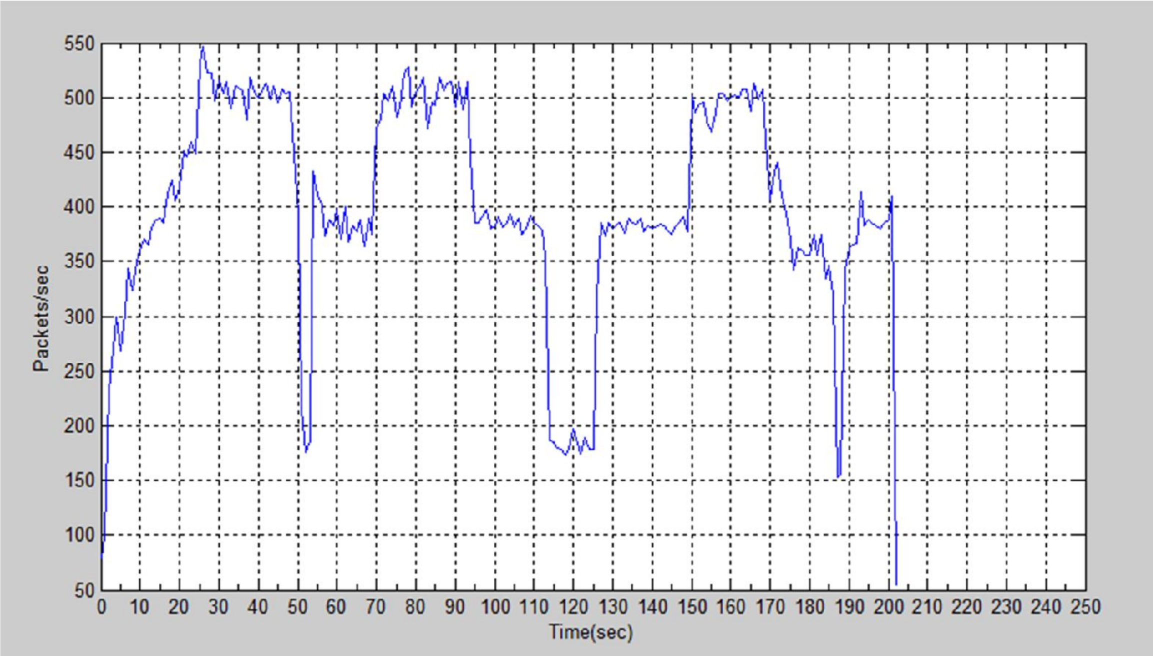

Figure 7. 8 My Video Packets over Time

7.2 PACKET LOSS

Since VoIP applications are using UDP protocol there is no acknowledgement and no retransmission of packets that can be measured for packet loss analysis so packet loss analysis has been done using identification (ip.id) as each packet that pass through network has unique ip.id. Any missing ip.id is the cause of packet loss. Packet loss for all the VoIP applications are measured and is given in **Table 7. *1***. Packet Loss comparison shows that Imo application suffers most in the given test condition as it occurs 87.85% packet loss, Google Duo with 11.49%, Skype with 7.34%, and My Video with 6.32% packet loss respectively.

Table 7. 1 Packet Loss Comparison

VoIP Applications	Total Packets	Lost Packets	Packet Lost Percentage
Google Duo	60367	6939	11.49%
Skype	52219	3831	7.34%
Imo	1194	1049	87.85%
My Video	81105	5124	6.32%

7.3 FRAME SIZE

Frame is the unit of transmitting data over the network and it consists of sequence of bits. Higher the frame size better is the VoIP quality as shown in Figure 7. 9, Figure 7. 10, Figure 7. 11, and Figure 7. 12. It is measured in Bytes over time series.

Google Duo, Skype, and My Video all have same maximum frame size i.e. 1200 Bytes/sec but average frame size of Google Duo is from 800 Bytes/sec to 900 Bytes/sec, Skype frame size range from 600 Bytes/sec to 800 Bytes/sec, and My video frame size range from 800 Bytes/sec to 1000 Bytes/sec. Imo has lowest frame size among all the three i.e. 450 Byte/sec to 500 Byte/sec on an average.

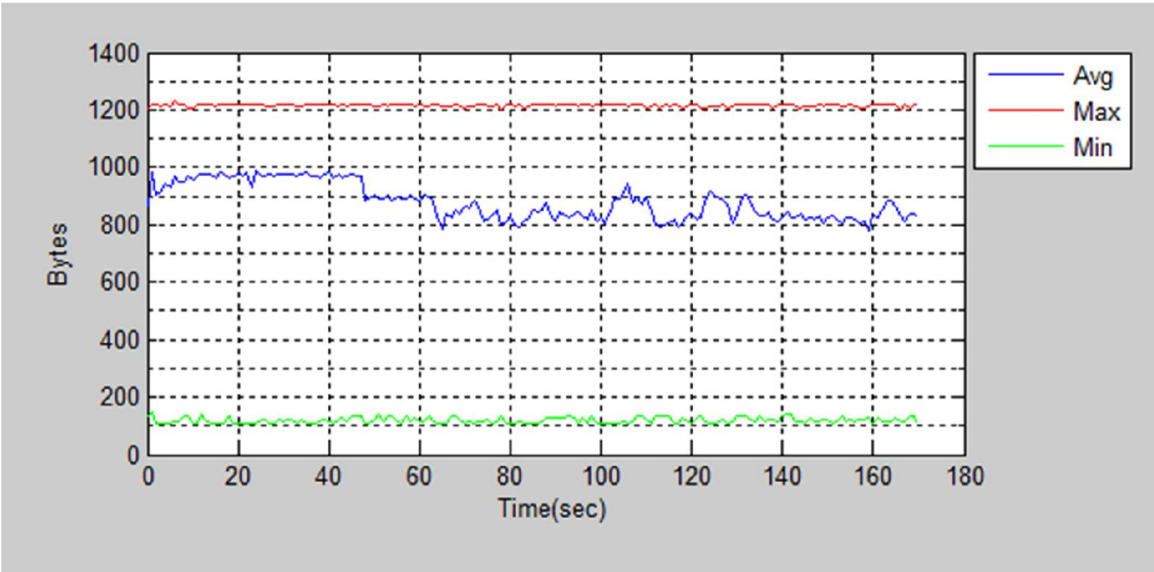

Figure 7. 9 Google Duo Frame Size over Time

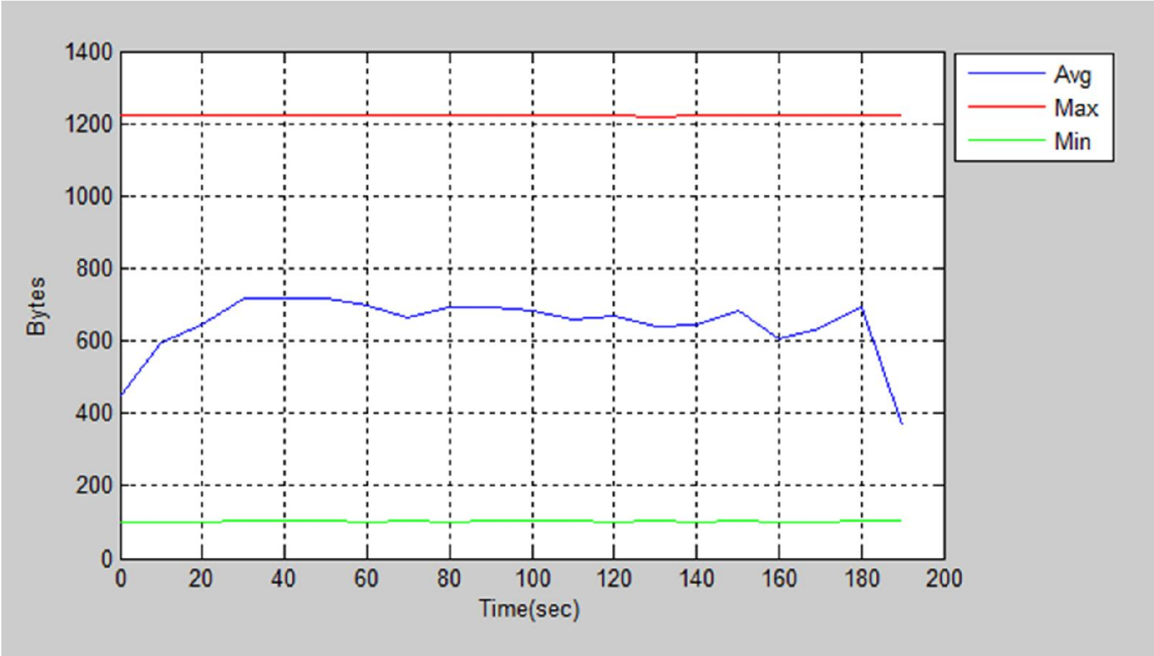

Figure 7. 10 Skype Frame Size over Time

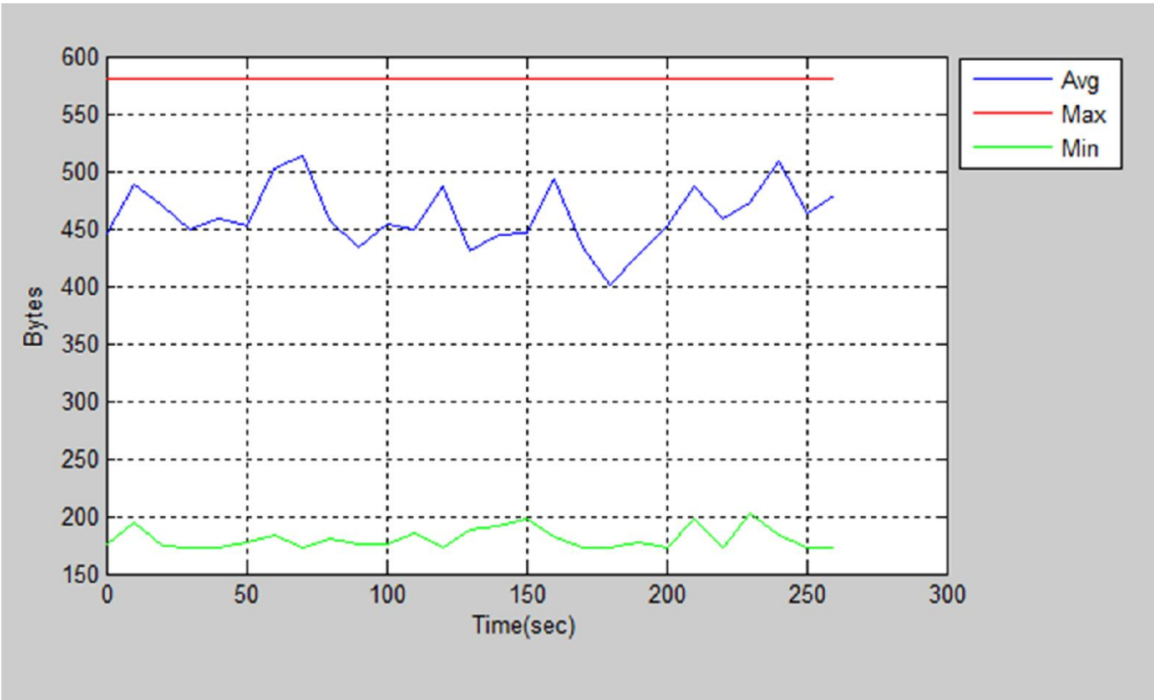

Figure 7. 11 Imo Frame Size over Time

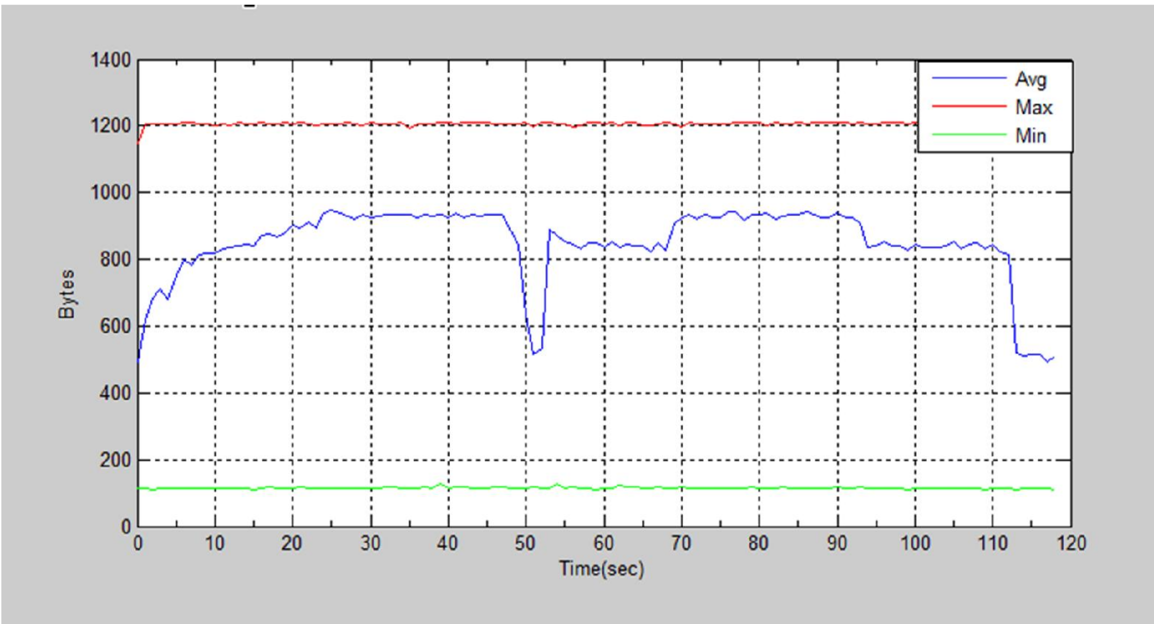

Figure 7. 12 My Video Frame Size over Time

CHAPTER 8

SUMMARY AND CONCLUSIONS

This research focus on enhancement quality of VoIP chat application over the internet. As VoIP packets is equipped with noise and echo as in Figure 7. 6 of test experiment performed this influencing factors must be minimized to enhance the quality of VoIP communication. So this research uses adaptive filtering algorithms (NLMS algorithm) in order to remove/ reduce the noise and echo from the packets. For this purpose proposed application has been developed and test. In the test experiment perform on VoIP apps (Google Duo, Skype, Imo, and My Video) Google Duo performed well even though there is a 11.49% of packet loss and bandwidth over time is between 1.8M bits/sec – 2.5M bits/sec with auto video pausing capability during the call if sufficient bandwidth is provided. Skype shows problem in connecting in low bandwidth but it ranks second in the test with packet loss of 7.34%. My Video (developed application) shows good quality video with sometimes noise at start but later it stabilize itself as adaptive filter coefficient reduces the noise with time. It suffers least packet loss (6.32%) in the test. Whereas Imo shows poor performance with a lot of noise and echo during the test. It suffers 87.85% of packet loss.

REFERENCES

[1] "Alexander Gramham Bell Quote," *New York Times*, 1927. [Online]. Available: https://books.google.co.in/books?id=SdIABAAAQBAJ&pg=PA50&dq=the+day+would+come+when+the+man+at+the+telephone+would+be+able+to+see+the+distant+person+to+whom+he+was+speaking+article+in+new+york+times&hl=en&sa=X&ved=0ahUKEwitwpC-5tfQAhWLN48KHdYDBO0Q6AEIGzAA.

[2] "AT&T Picturephone." [Online]. Available: http://atwiki.assistivetech.net/index.php/Videophone.

[3] B. A. Forouzan, *Data Communications and Networking*, Fourth Edi. McGraw-Hill.

[4] "VoIP: Voice Over Internet Protocol Architecture and Features." [Online]. Available: https://books.google.co.in/books?id=guprsQQMQg0C&pg=PA35&lpg=PA35&dq=h.255.0+protocol&source=bl&ots=NB-BmNlaWd&sig=qNM3d-8hKrEOhKaxMLTnAsByk6s&hl=en&sa=X&ved=0ahUKEwjF7I7R5OLQAhWDuI8KHaIvCeQQ6AEIZTAM#v=onepage&q=h.255.0 protocol&f=true.

[5] "RFC 3550 - RTP: A Transport Protocol for Real-Time Applications." [Online]. Available: https://tools.ietf.org/search/rfc3550.

[6] "RFC 3605 - Real Time Contro Protocol (RTCP)." [Online]. Available: https://tools.ietf.org/html/rfc3605.

[7] P. Shade, "VoIP Analysis Fundamentals with Wireshark," *Sharkfest*, p. 36, 2012.

[8] "WebRTC Tutorial." [Online]. Available: http://www.tutorialspoint.com/webrtc/index.htm.

[9] "Getting Started | WebRTC." [Online]. Available: https://webrtc.org.

[10] "WebRTC Audio Codec and Processing Requirements," 2016. [Online]. Available: https://tools.ietf.org/html/draft-ietf-rtcweb-audio-11.

[11] "Frequent Questions | WebRTC." [Online]. Available: https://webrtc.org/faq.

[12] P. P. Kadam, Z. Saquib, and A. Lahane, "Adaptive echo cancellation in VoIP network," in *Engineering and Technology (ICETECH), 2016 IEEE International Conference on*, 2016, pp. 295–299.

[13] T. Madhavi, B. H. Krishna, and L. P. Kanth, "A novel approach for voice quality enhancement for VoIP applications using DSP processor," in *Electrical, Electronics, and Optimization Techniques (ICEEOT), International Conference on*, 2016, pp. 4946–4951.

[14] I. Sharma, R. Mehra, and M. Singh, "Adaptive filter design for ECG noise reduction using LMS algorithm," in *Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions), 2015 4th International Conference on*, 2015, pp. 1–6.

[15] H. Jianjiang and Y. Zhibang, "The strong RF signal cancellation based on LMS adaptive filtering algorithm," in *Electronic Information and Communication Technology (ICEICT), IEEE International Conference on*, 2016, pp. 493–496.

[16] Z. Zheng, Z. Liu, H. Zhao, Y. Yu, and L. Lu, "Robust Set-Membership Normalized Subband Adaptive Filtering Algorithms and Their Application to Acoustic Echo Cancellation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2017.

[17] "MATLAB." [Online]. Available: http://mathworks.com/products/matlab.html.

[18] "Wireshark." [Online]. Available: http://wireshark.org.

[19] "SteelCentral Packet Analyzer." [Online]. Available: http://www.riverbed.com/.

[20] "Android Studio." [Online]. Available: http://developer.android.com/.

[21] B. Daldal, \. Ibrahim Bilgin, D. Ba\csaran, and S. Metin, "Using Web Services for WebRTC signaling interoperability," in *Network Operations and Management Symposium (NOMS), 2016 IEEE/IFIP*, 2016, pp. 780–783.

[22] P. Wuttidittachotti and T. Daengsi, "Subjective MOS model and simplified E-model enhancement for Skype associated with packet loss effects: a case using conversation-like tests with Thai users," *Multimedia Tools and Applications*, pp. 1–25.

[23] A. Azfar, K.-K. R. Choo, and L. Liu, “Android mobile VoIP apps: a survey and examination of their security and privacy,” *Electronic Commerce Research*, vol. 16, no. 1, pp. 73–111, 2016.

[24] B. Sredojev, D. Samardzija, and D. Posarac, “WebRTC technology overview and signaling solution design and implementation,” in *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015 38th International Convention on*, 2015, pp. 1006–1009.

[25] K.-Y. Yoon and B. Kim, “An Empirical Study on the Quality Assessment of the VoIP Service over Wireless Mobile Networks,” in *Multimedia and Ubiquitous Engineering*, Springer, 2014, pp. 153–157.

[26] Kim, C.-C., Kim, and B., “A Software Implementation for Quality Management of Mobile VoIP Services,” *Springer, Heidelb.*, vol. 353, pp. 73–80, 2012.

[27] ITU, “The E-Model, A Computational Model for Use in Transmission Planning: ITU-T Recommendation G.107,” 1998.

[28] C. Yu, Y. Xu, B. Liu, and Y. Liu, ‘Can you SEE me now?’ A measurement study of mobile video calls,” in *INFOCOM, 2014 Proceedings IEEE*, 2014, pp. 1456–1464.

[29] Y. Xu, C. Yu, J. Li, and Y. Liu, “Video telephony for end-consumers: Measurement study of google+, ichat, and skype,” *Proc. Internet Meas. Conf.*, pp. 371–384, 2012.

[30] A. Johnston, J. Yoakum, and K. Singh, “Taking on webRTC in an enterprise,” *IEEE Communications Magazine*, vol. 51, no. 4, pp. 48–54, 2013.

[31] X. Zhang, Y. Xu, H. Hu, Y. Liu, Z. Guo, and Y. Wang, “Profiling skype video calls: Rate control and video quality,” in *INFOCOM, 2012 Proceedings IEEE*, 2012, pp. 621–629.

[32] M. Qadeer, S. Rizvi, and J. Chaudhary, “A novel adaptive filtering technique for real time acquisition of signals buried in noise,” in *Applied Sciences and Technology (IBCAST), 2017 14th International Bhurban Conference on*, 2017, pp. 843–846.

[33] P. Gupta, M. Patidar, and P. Nema, “Performance analysis of speech enhancement using LMS, NLMS and UNANR algorithms,” in *Computer, Communication and*

Control (IC4), 2015 International Conference on, 2015, pp. 1–5.

[34] T. Daengsi and P. Wuttidittachotti, “QoE modeling: a simplified e-model enhancement using subjective MOS estimation model,” in *Ubiquitous and Future Networks (ICUFN), 2015 Seventh International Conference on*, 2015, pp. 386–390.