San Jose State University
SJSU ScholarWorks

Master's Projects Master's Theses and Graduate Research

2009

Objective Measurement of Speech Quality in VoIP over Wireless LAN during Handoff

Nidhi Marwaha Gambhir
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_projects

Part of the Computer Sciences Commons

Recommended Citation
Gambhir, Nidhi Marwaha, "Objective Measurement of Speech Quality in VoIP over Wireless LAN during Handoff" (2009). *Master's Projects*. 47.
DOI: https://doi.org/10.31979/etd.85z3-hacv
https://scholarworks.sjsu.edu/etd_projects/47

This Master's Project is brought to you for free and open access by the Master's Theses and Graduate Research at SJSU ScholarWorks. It has been accepted for inclusion in Master's Projects by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

Objective Measurement of Speech Quality in VoIP over Wireless LAN during Handoff

A Writing Project

Presented to

The Faculty of the Department of Computer Science

San Jose State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Nidhi Marwaha Gambhir

Fall 2009

Copyright © 2009

Nidhi Marwaha Gambhir

All Rights Reserved

ABSTRACT

Quality of Service is a very important factor to determine the quality of a VoIP call. Different subjective and objective models exist for evaluating the speech quality in VoIP.

E-model is one of the objective methods of measuring the speech quality; it considers various factors like packet loss, delay and codec impairments. The calculations of E-model are not very accurate in case of handovers – when a VoIP call moves from one wireless LAN to another. This project conducted experimental evaluation of performance of E-model during handovers and proposes a new approach to accurately calculate the speech quality of VoIP during handovers. A detailed description of the experimental setup and the comparison of the new approach with E-model is presented in this report.

ACKNOWLEDGEMENTS

I would like to thank Dr. Soon Teoh for his support and valuable suggestions all through my research work. I would not have been able to complete this project without his guidance. I would also like to thank Dr. Robert Chun and Professor Frank Butt for being a part of my committee and taking out time to review my project. Lastly I would like to thank my family for motivating me to complete the project.

Table of Contents

List of Tables

List of Figures

1. Introduction

Voice over IP (VoIP) services has become very popular since the advent of internet and is now replacing the traditional PSTN networks and the cellular networks. The main reason for choosing VoIP calls over cellular calls is its inexpensive nature. Long distance cellular calls tend to be very expensive to setup, whereas the same call can be setup at no cost over the internet. Google Talk, Skype and Yahoo Messenger provide free voice services over the internet and thus are banned in some countries. Nowadays, VoIP phones are available that can be used even when the user is travelling to setup a call to another VoIP phone, land-line or a mobile number at a bare minimal cost.

1.1 Problem Description

Voice over IP services uses the traditional Internet Protocol (IP) to send the voice packets. It breaks the voice call into small packets that are routed over the internet. Due to the unreliable nature of the internet, these packets might get lost in the network which results in missing packets at the receiver end. As a result, the receiver would hear the speaker's sentence incomplete and may not understand it. It is very essential to monitor the quality of these voice calls to achieve user satisfaction.

To measure the speech quality various network factors like delay, packet loss, jitter etc. are considered. The measured speech quality is then mapped to a user satisfaction level. Nowadays, many people make VoIP calls when they are travelling, thus moving from one network to another. It is very important that the user experiences a good call quality when the VoIP call gets handed off from one network to another. The process of handoff consists of temporarily disconnecting from one network and then establishing a connection with the new network, this could result in dropped calls or heavy packet loss if not performed smoothly. Thus, it is very important to measure the speech quality of VoIP during handovers to achieve high user satisfaction.

1.2 Related Work

Speech quality has always been a concern for VoIP calls. The traditional subjective method to measure the speech quality is by performing the Mean Opinion Score (MOS) tests. To replace the expensive subjective test, objective methods are developed that base the calculations by considering varied network parameters. The most popular objective methods to calculate speech quality are Perceptual Evaluation of Speech Quality (PESQ) and E-model. The effect of handoff on voice quality of VoIP is studied in [4]; the paper studies the various network parameters that are affected during a call handover.

In this project, a new approach is proposed to measure the speech quality of VoIP calls during handover based on the existing E-model method.

1.3 Outline

The brief outline of the report is below:

Chapter 2 provides information on VoIP networks. The Transport Control Protocol (TCP), Internet Protocol (IP) and Real Time Protocol (RTP) are discussed in detail. It also discusses the coding technology used in VoIP namely codec's G.711 and iLBC.

Chapter 3 covers the various factors that affect the Quality of Service in VoIP. The method of measuring the speech quality using Mean Opinion Score (MOS), Perceptual Evaluation of Speech Quality (PESQ) and E-model are discussed.

Chapter 4 discusses the impact of handovers on VoIP. It studies the performance of E-model during handoff and introduces the new proposed method to measure the speech quality during handoff.

Chapter 5 covers the whole test setup which includes screenshots of the VoIP client, Wireshark tool and the software developed.

Chapter 6 shows the implementation results suing the traditional E-model and the new proposed model. It correlates subjective MOS test experiment results with E-model and the new objective model. It also compares the two speech codec's iLBC and G.711 for various network parameters.

2. VoIP Networks

2.1 Introduction to VoIP

With the growth of Internet technology, the traditional PSTN – Public Switched Telephone Network is now being replaced by the Voice over IP (VoIP) network. A call in a traditional telephone network is established using circuit switching, where a dedicated channel is setup between the callers. The advantage of this approach is that once a call is setup between the callers the voice quality is very good, but on the other end regular telephone calls requires a lot of resources like network capacity that makes it expensive to setup [18].

Voice over IP establishes a call between the users using the Internet protocol. It breaks the voice data into packets and transfers them over the internet. On the other hand the receiver reassembles those packets to establish a voice call. VoIP calls utilize the bandwidth more efficiently as compared to the PSTN calls and hence are cheaper to setup. This is one of the major advantages of using VoIP over telephony now a-days. VoIP calls can be established PC-to-PC, PC-to-telephone or telephone-to-PC and is also referred as “Internet Telephony” or “IP Telephony” [18].

2.2 VoIP Network Connections

A common VoIP connection is established between phone to phone, phone to PC or PC to PC as shown in figure 2.1 [20].

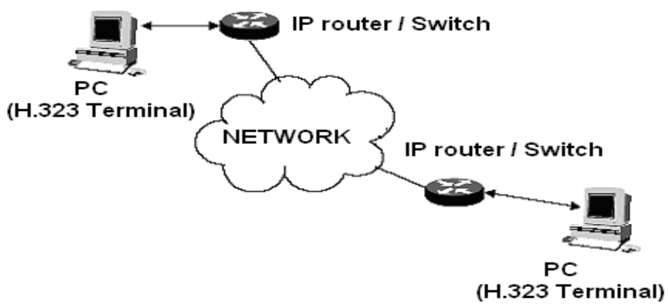


Fig 2.1. PC-to-PC Architecture [20]

The basic process of VoIP includes the following steps:

1. Conversion of analog voice into digital packets
2. Conversion of digital packets into Internet protocol packets
3. Transmission of the packets over IP based network
4. Conversion of digital packets into analog voice at receiver's end

The process of compression is carried out by a voice encoding algorithm called *Codec,* which allows the call to be transmitted over the IP network.

2.3 TCP/IP Protocol

Transmission Control Protocol and Internet Protocol is a stack of protocols that sets up an agreement for two computers to communicate. TCP/IP protocol consists of four layers as shown in figure 2.2.

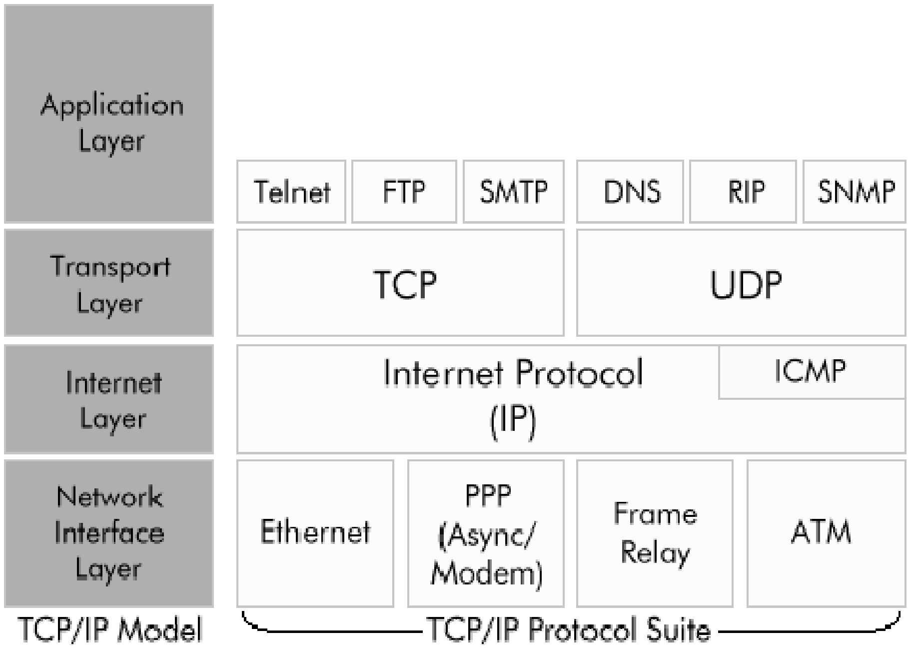


Figure 2.2. TCP/IP Protocol Stack [9]

Application layer is the uppermost layer of the TCP/IP protocol stack, it is used by programs like SMTP (for email), HTTP (web browsing) and FTP (file transfer protocol). After processing the request at the Application layer, the protocol at the Application layer now interacts with Transport layer, also referred as Transmission Control Protocol (TCP). TCP gets data from the Application layer and breaks it into packets and sends it to the Internet Layer [9].

The Internet layer, also known as Internet protocol adds a virtual address, known as IP address to the data packets. The IP address of the sender and that of the receiver is added to the data packet. The lowest layer is the Network layer and the data is sent as datagram's in this layer [9].

2.3.1 Transport Layer

The Transmission Control Protocol (TCP) is responsible for sending the packets over the Internet protocol and also reassembling the packets at the receiver end. If the packets arrive out of order, then TCP reorders the packets and incase of lost or corrupted packet received, it sends an acknowledgement to the transmitter. The packet is then resend by the transmitter. The TCP protocol ensures reliable delivery of packets over the internet.

The Transport layer also consists of the User Datagram Protocol (UDP) which is known as the unreliable protocol. Unlike TCP, UDP protocol is not in charge of reordering the packets and checking for the lost ones. As UDP protocol does not acknowledge or reorder data, it transmits data fast as compared to TCP. UDP protocol is mostly used when data sent over the internet is not so important. In Transport layer, TCP protocol is known as Connection-oriented protocol, whereas UDP is known as Connection-less protocol [11].

2.4 Real Time Protocol

Real Time Protocol (RTP) is used to transport delay sensitive applications like audio and video streams. It uses the User Datagram Protocol (UDP) rather than TCP to minimize the delay for these real time applications. RTP packet encapsulates its payload to include a sequence number and a timestamp to keep track of the real time information [26].

2.4.1 RTP Packet Format

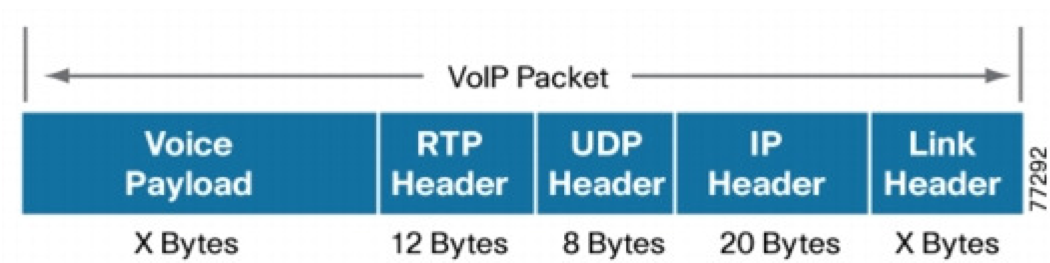


Figure 2.3a. RTP Datagram for VoIP [3]

```
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |V=2|P|X|  CC   |M|     PT      |       sequence number         |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                           timestamp                           |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |           synchronization source (SSRC) identifier            |
 +=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
 |            contributing source (CSRC) identifiers             |
 |                             ....                              |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
Figure 2.3b. RTP Packet Format [17]

RTP protocol has a standard format for all real time applications. The fields of the RTP protocol have the following meaning [26]:

(a) Version (V): This field identifies the version of the RTP. It is 2 bit long.

(b) Padding (P): This is a 1 bit long field; if this bit is set then additional padding octets are set at the end of the packet. Some encryption algorithms use the padded field.

(c) Extension (X): This is 1 bit long and if this bit is set then there is exactly one header extension for the header.

(d) CSRC Count (CC): It is the number of CSRC identifiers that follow the fixed header. It is 4 bit long.

(e) Marker (M): The profile defines the marker; it marks the frame boundaries in the packet stream. Its length is 1 bit.

(f) Payload Type (PT): This field identifies the format of the RTP payload. The length of this field is 7 bits.

(g) Sequence Number: This field is 16 bits in length; it is incremented by one for each RTP packet. The receiver uses the sequence number to detect any packet loss and to restore the original packet sequence.

(h) Timestamp: It reflects the instant at which RTP data packet is sampled. The clock is monotonically incremented to calculate jitter and delays. Its length is 32 bits.

(i) SSRC (Synchronization Source): It identifies the synchronization source. The SSRC identifier is chosen randomly such that no two synchronization sources within the same RTP share the same identifier. Its length is 32 bits.

(j) CSRC (Contributing Source): It is an array of 0 to 15 identifying elements. The CC field lists the number of identifiers and if there are more than 15 identifiers, only 15 might be identified.

2.4.2 RTCP – RTP Control Protocol

Real Time Transport Control Protocol (RTCP), are special messages that are used by RTP to exchange real time reports and statistics. RTCP packets are used for diagnostic purpose and are exchanged timely between the sender and the receiver. It consists of three types of reports: Sender Report (SR), Receiver Report (RR) and Source Description (SDES). The RTCP reports consists of information like total number of packets sent, total packets received and lost and also the inter-arrival jitter. These packets help to monitor the performance and Quality of Service (QoS) of real time data [26].

2.5 Codec

Codec stands for coder/decoder; it converts the analog signal into digital bits and sends it out at a constant data rate. Codec is also referred to as Compressor/Decompressor as it compresses the data bits while sending to save the bandwidth.

The process of converting the speech signal into digital values is called Sampling. It represents the amplitude of the input speech waveform. The accuracy of sampling depends on two factors: The number of bits to represent each sample and the maximum sampling rate. In order to get a good quality the maximum sampling rate should be twice the highest frequency as stated by Nyquist's theorem [22].

2.5.1 Sampling

The process of Sampling uses Pulse Code Modulation (PCM). All the frequencies that are below a certain threshold are filtered out using a band-pass filter. Now the analog signal is sampled into values that represent the amplitude for the signal over a particular time period as shown in figure 2.4 below. This sampled signal is now quantized and mapped to fixed values (figure 2.5) that are coded and sent to the receiver. The receiver does the reverse process to obtain the analog signal [20].

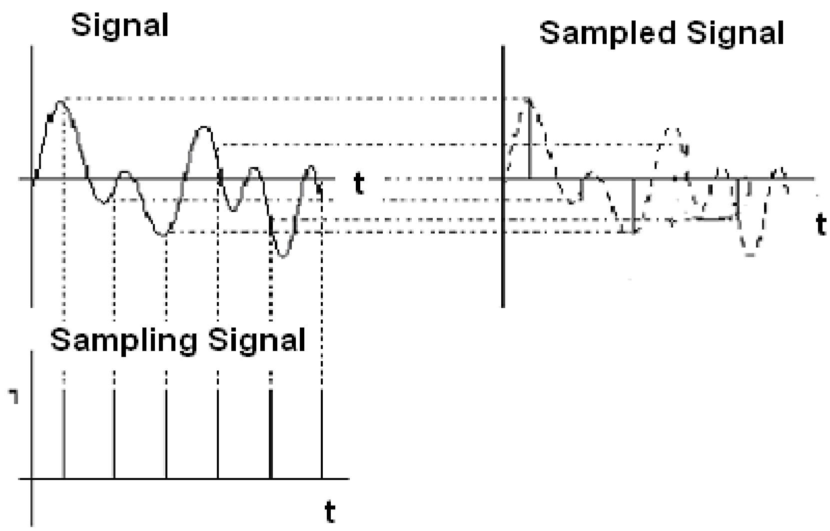


Figure 2.4. Sampling of Signal from Analog to Digital [20]

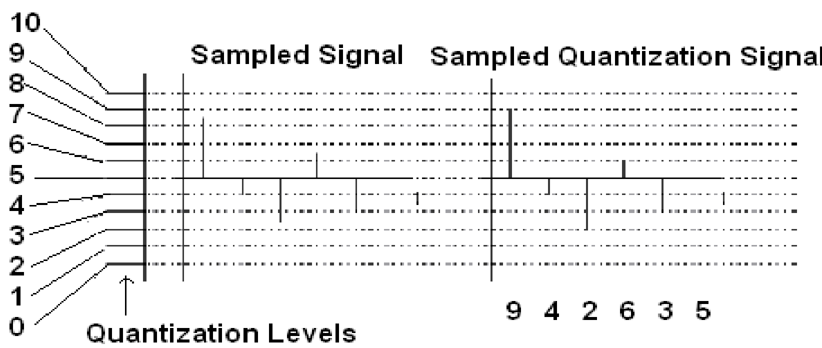


Figure 2.5. Quantized form of a signal [20]

2.5.2 Internet Low Bit Rate Codec – iLBC

Internet Low Bit Rate Codec (iLBC) is developed by Global IP Solutions (GIPs) and it is a freely available codec for VoIP networks. iLBC uses Linear Predictive Coding (LPC) algorithm and is suitable for narrowband speech. It uses Packet Loss Concealment (PLC), where all the coded frames are independent; this makes iLBC perform a little better during packet loss and delay [27].

Specifications	
Implementation	Fixed point ANSI C and DSP[1]. Please visit www.ilbcfreeware.org for access of the royalty-free floating point C code and specification
Frame size	20 and 30 ms
Bit rate	13.3 kbps (30 ms frames) and 15.2 kbps (20 ms frames)
Sampling rate	8 kHz
Basic Quality	Better than G.729A, and comparable to G.729E
Packet loss robustness	Significantly better than G.729A and exceeds G.729E
Packet loss concealment	Integrated example solution

Table 1. iLBC Codec Specifications [13]

2.5.3 G.711 Codec

G.711 is an ITU –T standard and is primarily referred to as telephony codec. Formal name for G.711 is Pulse Code Modulation (PCM) and it represents speech sampled at 8000 samples/second. G.711 uses two main logarithmic algorithms: A- law and the µ – law. The A – law and the µ – law algorithms encode 13 – bit and 14 – bit linear PCM samples into 8-bit logarithmic samples. The µ – law is more suitable for higher level signals while A – law is suitable for lower level signals [2].

Parameter	**Value**
Bit Rate	64kbps
Sampling Rate	8000Hz

Table 2. G.711 Codec Specification [2]

2.6 Summary

This chapter gave an introduction of the VoIP technology and real time network protocol used for VoIP. Also it discussed about the codec's iLBC and G.711 and its coding techniques.

3. Quality of Service

Quality of Service (QoS) is the ability to measure the network's performance such that it delivers predictable results. Quality of Service helps to differentiate between the type of service required and the type of traffic and is a very important tool for VoIP services [28].

3.1 Factors Affecting VoIP Quality

In VoIP, quality means the ability to talk and listen clearly without any unwanted noise. The three major factors that affect the speech quality in VoIP are: Packet Loss, Jitter and Latency.

3.1.1 Packet Loss

Packet loss happens when two users talk on a VoIP phone and the call begins to "break up" during packet transfer over the IP network. This mainly happens when there is a lot of congestion on the network and it results in some part of the conversation being missed at the receiving end. Thus to minimize the affect of packet loss, all VoIP applications should have a module for packet loss concealment [29].

Two major methods used for packet loss concealment are: Packet repetition and Zero stuffing. In packet repetition, if a packet is lost then the lost packet is replaced by the previously received packet whereas in zero stuffing the lost packets are replaced by zeroes. However, both the techniques fail to perform when packet loss exceeds 10% [29].

3.1.2 Jitter

In VoIP, when a call is established, the sender sends the VoIP packet at a constant frequency, for example 10ms or 20ms. As the packet travels over the internet it might get delayed or lost, resulting in receiver not receiving the packets with the same frequency. The difference in the expected time of arrival and the actual time of arrival of the packet is called jitter. It is usually experienced in heavily congested networks [29].

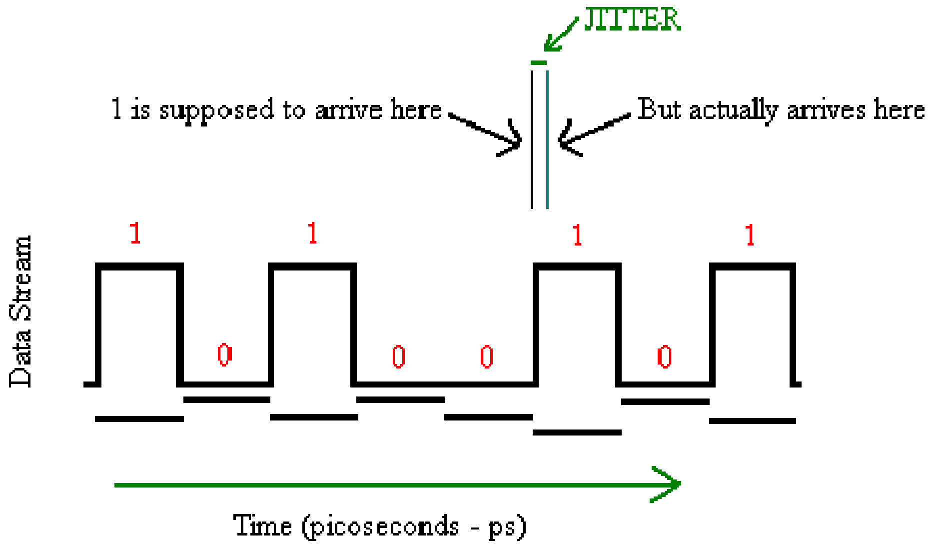


Figure 3.1. Jitter in network [32]

In figure 3.1, it can be seen that a stream of bits is transmitted. The difference between the expected arrival time of bit 1 and the actual arrival time of bit 1 is jitter. In VoIP, jitter can be controlled by the use of jitter buffers.

Jitter buffers are deployed at the receiving end; they receive the VoIP packets, store them in a buffer and then send out the buffered packets at evenly spaced time intervals. This reduces the variations in the packet arrival rate. There are two types of jitter buffers: Fixed size jitter buffer and adaptive jitter buffer. Adaptive jitter buffer is preferred to fixed size jitter buffer as it adapts its size to the network conditions.

3.1.3 Delay or Latency

The total amount of time taken for speech to travel from speaker's mouth to listener's ear is measured as delay or latency. In term of VoIP packets, it is the total time taken by the packet to travel from the source to the destination. ITU-T recommends that one-way delay should not be more that 150 ms for good speech quality [29].

In telephony network the delay is categorized into three types, as shown in figure 3.2: Serialization delay, Propagation delay and handling delay.

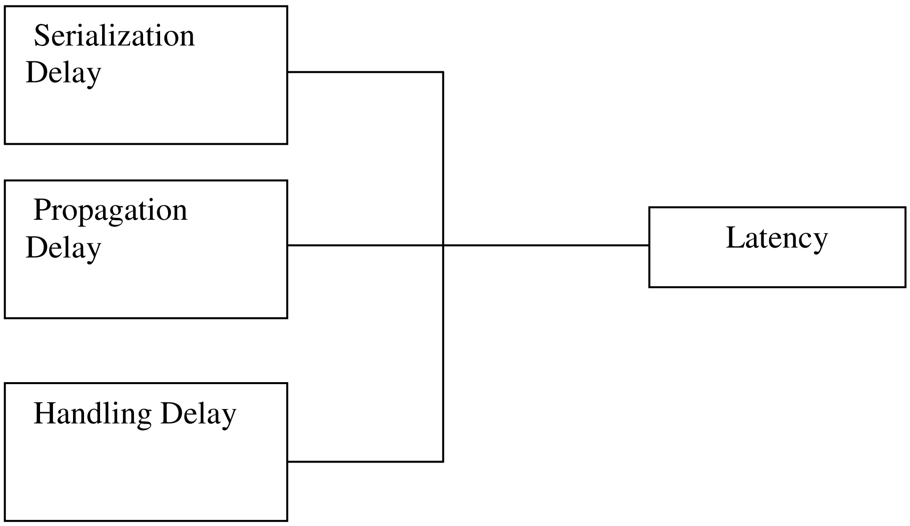


Figure 3.2. Latency in Network

Serialization Delay: It is the amount of time it actually takes to place a bit or byte onto an interface. The effect of serialization delay is very minimal and does not relatively affect the overall delay [24].

Propagation Delay: It is the amount of time it takes a signal to propagate through a copper wire or a fiber optic. It is the total length a signal has to travel before reaching the human ear. This delay might be imperceptible, but when combined with other delay's can cause a noticeable degradation to voice quality [24].

Handling Delay: It is mainly caused by the device that sends the packets on the network**,** it includes factors like packetization, compression, queuing delay and packet switching. In VoIP networks, the time taken by the device to send the packet to the output queue is called packet switching, but when more packets are sent to the output queue than it can handle it results in a queuing delay. Handling delay is also sometimes referred to as **Codec Processing delay** [22]**.**

3.1.4 Tradeoff

A three-dimensional relationship among computation of codec, speech quality and delay is shown in figure 3.3 below [29].

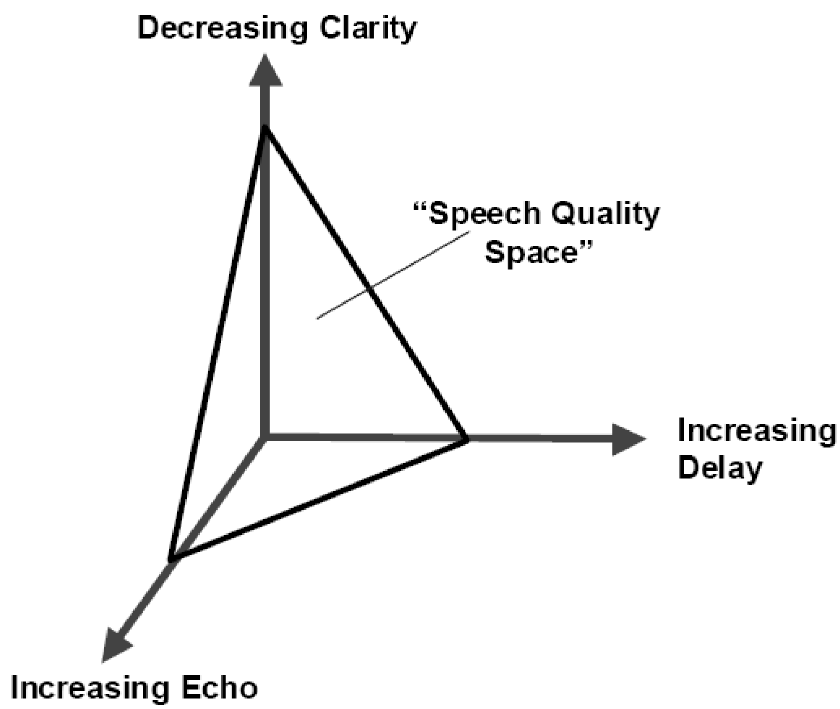


Figure 3.3: Relationship among speech quality, delay and Increasing Echo [29]

If echo increases in the VoIP, it results in a degraded quality and so does increased delay has adverse affect on voice quality. If voice quality is plotted as a single point in the graph, then the nearer it is to the intersection of the axis, the better the voice quality. As the distance between the intersection and the point increases the voice quality degrades.

3.2 Measurement of Speech Quality

Speech quality is the measurement of user experience when a VoIP call is established. The measurement of speech quality is divided into two broad categories: Objective measurement and Subjective measurement. Subjective tests are user listening tests where users are told to rate the speech quality. These tests are expensive to perform and the accuracy of speech quality rating relies on the user's mood. To measure the accuracy of these subjective tests, objective methods are used. These methods are the computational methods that usually compare a good quality signal to a degraded signal [29].

3.2.1 Mean Opinion Score (MOS) – Subjective Listening Test

Mean Opinion Score (MOS) is International Telecommunications Union Telecommunication Standardization sector (ITU-T) approved. It is a subjective listening test where the user rates the speech quality during the call. MOS tests can also be performed in two variations as: Absolute Category Rating (ACR) and Degradation Category Rating (DCR). The DCR test is used in some occasions to get the Degradation MOS (DMOS) scores [17].

A MOS test usually involves 12-24 participants who individually listen to an audio stream of several seconds and rate the audio quality on the scale of 1 to 5. The quality rate scale is summarized in table 3.1 below. The average score of the listeners gives the Mean Opinion Score (MOS). A MOS rating of 4.5 to 5 is considered excellent speech quality whereas a rating of 4 is also considered reasonably good and acceptable [17].

Score	LQ scale	LE scale
5	Excellent	Complete relaxation possible; no effort required
4	Good	Attention necessary; no appreciable effort required
3	Fair	Moderate effort required
2	Poor	Considerable effort required
1	Bad	No meaning understood with any feasible effort

Table 3. MOS rating based on Listening Quality (LQ) and Listening Effort (LE)

The MOS ratings are summarized in table 3.2 below:

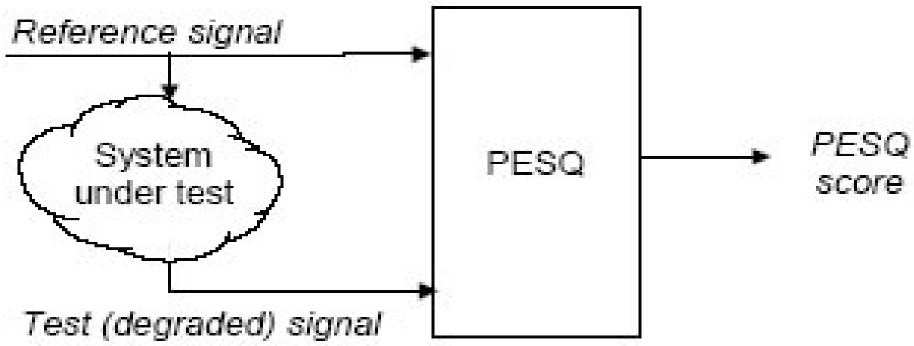
MOS Score Range	Voice Quality
4.0-5.0	Desirable speech quality
3.6 – 4.0	Generally acceptable
1.0 – 3.6	Unacceptable, not recommended

Table 4. MOS ratings

MOS test ratings can be used to compare various codec’s such as iLBC and G.711. Although, MOS tests are the most reliable method of measuring the speech quality they are cumbersome to perform. They are considered as expensive tests and are quite time consuming, so it’s difficult to perform them frequently.

3.2.2 Perceptual Evaluation of Speech Quality (PESQ) – Objective Method

Perceptual Evaluation of Speech Quality (PESQ) is an ITU-T standard for objective measurement. It was introduced as MOS subjective tests were expensive to conduct and required a lot of time. PESQ test setup automatically maps the PESQ score to the subjective MOS score. It takes into account two signals; one is the reference signal while the other one is the actual degraded signal. Both the signals are sent through the test that uses the PESQ algorithm and the result is a PESQ score as shown in figure 3.4 below [33].



\

Figure 3.4: PESQ Testing [33]

Major drawbacks of PESQ approach was that it did not take into account the various impairments such as acoustic echo, transmission delay etc. Also this approach cannot be used to monitor real time calls and compare codec's accurately.

3.2.3 E-Model – Objective Approach

It is a new objective model proposed by ITU-T and it takes into account all the drawbacks of PESQ. It is a non-intrusive method of predicting the voice quality. E-model takes into account various factors that affect the speech quality and calculates a Rating factor (R-factor) that ranges between 0 -100. The R-factor can also be converted into a MOS rating to give the MOS score. The R-factor is calculated as [15]:

$$R_{obj} = R_0 - I_s - I_d - I_e + A \qquad (3.1)$$

Where:

R_0: Signal to Noise Ratio (S/N) at 0 dBR point

Is: Various speech impairments (e.g. Quantization noise, side tone level)

Id: Impairments that occur due to delay (e.g. absolute delay, echo)

Ie: Impairments caused by the equipment (e.g. codec's, jitter, packet loss)

A: Advantage factor (A is 0 for wireline and A is 5 for wireless)

Figure 3.5 below summarizes the R-factor ratings and user satisfaction based on ratings between 0 -100.

Figure 3.5. R-factor rating with MOS score mapping and user satisfaction level [30]

ITU G.107 provides an equation to convert the R-factor value in MOS score:

$$
\begin{aligned}
&\text{For } R < 0: && \mathrm{MOS} = 1 \\
&\text{For } 0 < R < 100: && \mathrm{MOS} = 1 + 0.035R + R(R-60)(100-R)7 \cdot 10^{-6} \\
&\text{For } R > 100: && \mathrm{MOS} = 4.5
\end{aligned}
\tag{3.2}
$$

Based on ITU G.107 recommendation, the R- factor equation can be simplified as:

R-factor = 93.2 – Id – Ie – A (3.3)

Where A is the Advantage factor; 0 for wireline and 5 for wireless networks.

The value of Ie, which is codec dependent impairment, is calculated as:

Ie = a + b ln (1+ cP/100) (3.4)

Where, P is percentage packet loss and a, b and c are codec fitting parameters.

Codec fitting parameters for iLBC and G.711 are summarized in table 3.3 below [24]:

Parameters	**G.711**	**iLBC**
Bitrate(kb/s)/framesize(ms)	64/20	15.2/20
a	0	10
b	30	19.8
c	15	29.7

Table 5. Fitting Parameters for codec G.711 and iLBC [24]

The value of Id, which is impairment due to delay is calculated as:

$$I_d = 0.024d + 0.11\,(d - 177.3)\,H(d - 177.3) \qquad (3.5)$$

Where d is the total one way delay (includes serialization delay, processing delay and propagation delay) in milliseconds. H(x) is a step function defined as:

$H(x) = 0$, $x<0$ and $H(x) = 1$ otherwise.

The drawback of E-model is that the MOS scores calculated by E-model do not correlate very well with the subjective MOS scores. Also E-model does not calculate the packet loss and delay accurately during handovers, when a VoIP call moves from one network to another.

3.3 Summary

This chapter discussed various factors that affect the speech quality (delay, packet loss and jitter) and how speech quality is measured using the subjective and objective methods. Mean Opinion Score which is the subjective method of measuring speech quality was discussed as well as objective methods like PESQ and E-model was described.

E-model is a very popular objective model for measuring speech quality but fails to accurately measure the speech quality during call handovers. Chapter 4 introduces handovers in VoIP and performance of E-model during handovers. It also introduces a new approach based on E-model to accurately calculate the speech quality during handovers.

4. Impact of Handoff on VoIP

Various factors that affect the speech quality in VoIP have already been discussed in chapter 3, but how a speech quality of VoIP is affected by a call handover is studied in detail in chapter 4.

4.1 Handoff in VoIP

Handoff is the process of transferring a call connection from one base station to another base station in a different cell or network. Process of handoff usually takes place when a user moves around a geographical area. In a VoIP call established in a wireless network, handover happens when a user moves from one wireless network to another network. The process of VoIP handoff is summarized in figure 4.1 below.

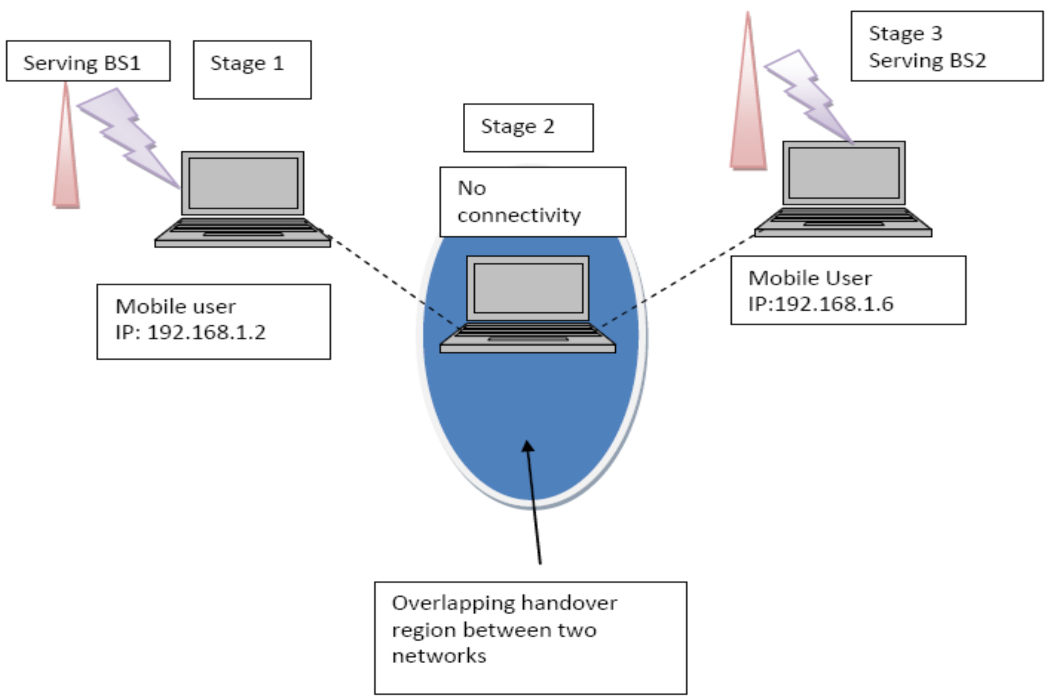


Figure 4.1. Stages of Handover during a VoIP call

So the handover stages can be summarized as:

Stage 1: The Mobile station (MS) communicates with the serving Base station 1.

When the MS enters the overlapping region of two networks then:

Stage 2: The MS is disconnected from the serving Base Station for a while, in this stage there is no connection to the network.

Stage 3: A new connection is established with the target Base station 2.

4.2 Quality of Service during Handover

The Quality of service of VoIP calls was subjectively calculated using the MOS tests and objectively it was calculated with E-model.

4.2.1 Subjective MOS Test

When MOS test was conducted during handover, the listeners experienced a gap (silence) for a while (that was during the handover phase) and after handover was complete they could hear the test sentences.

Some calls got dropped as the mobile user could not connect to the other wireless network. This happened due to the excessive delay during handover process; the mobile user moved out of the handover region and also did not get authenticated to the new network, leading to a dropped call.

4.2.2 E-model Calculations

The MOS score was also objectively calculated using the E-model. The comparison of the MOS scores for ilBC and G.711 are shown in figure 4.2.

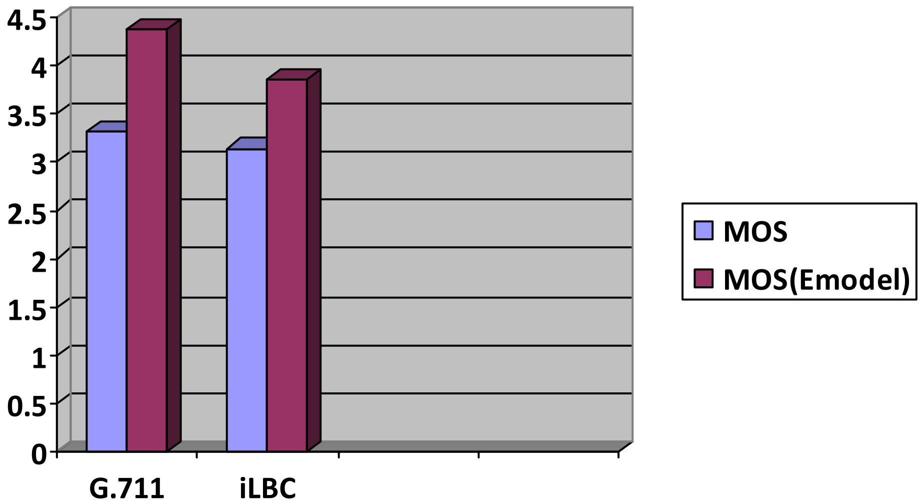


Figure 4.2. Comparison of MOS (objective) and MOS (subjective) during handover

The packets were captured using Ethereal during handover. During handover the voice call was temporarily disconnected as users did not hear anything for both G.711 and iLBC, but Ethereal showed a 0% - 0.02% packet loss for G.711 and 0.3-1.1% packet loss for iLBC codec.

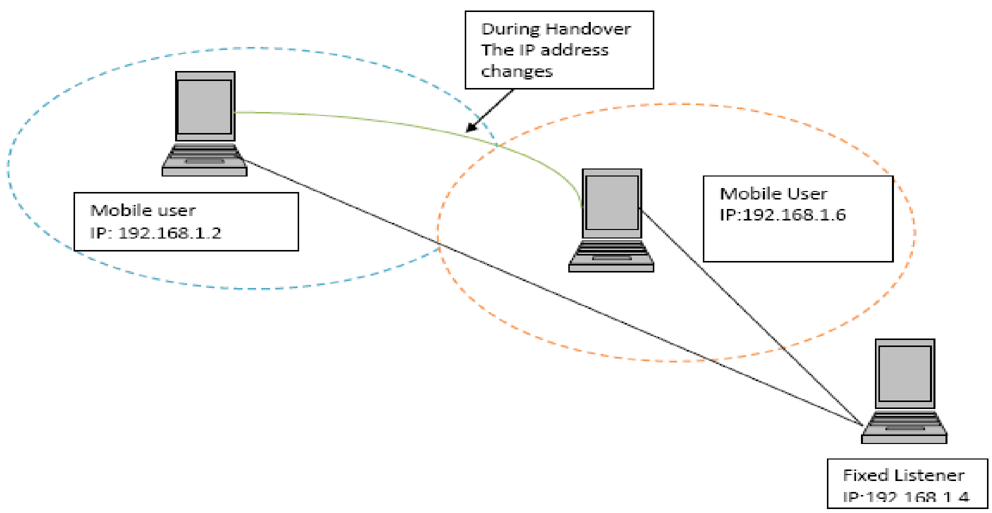


Figure 4.3. Scenario during handover

If Ethereal packets are studied, more analysis can be done as to what happens during handover:

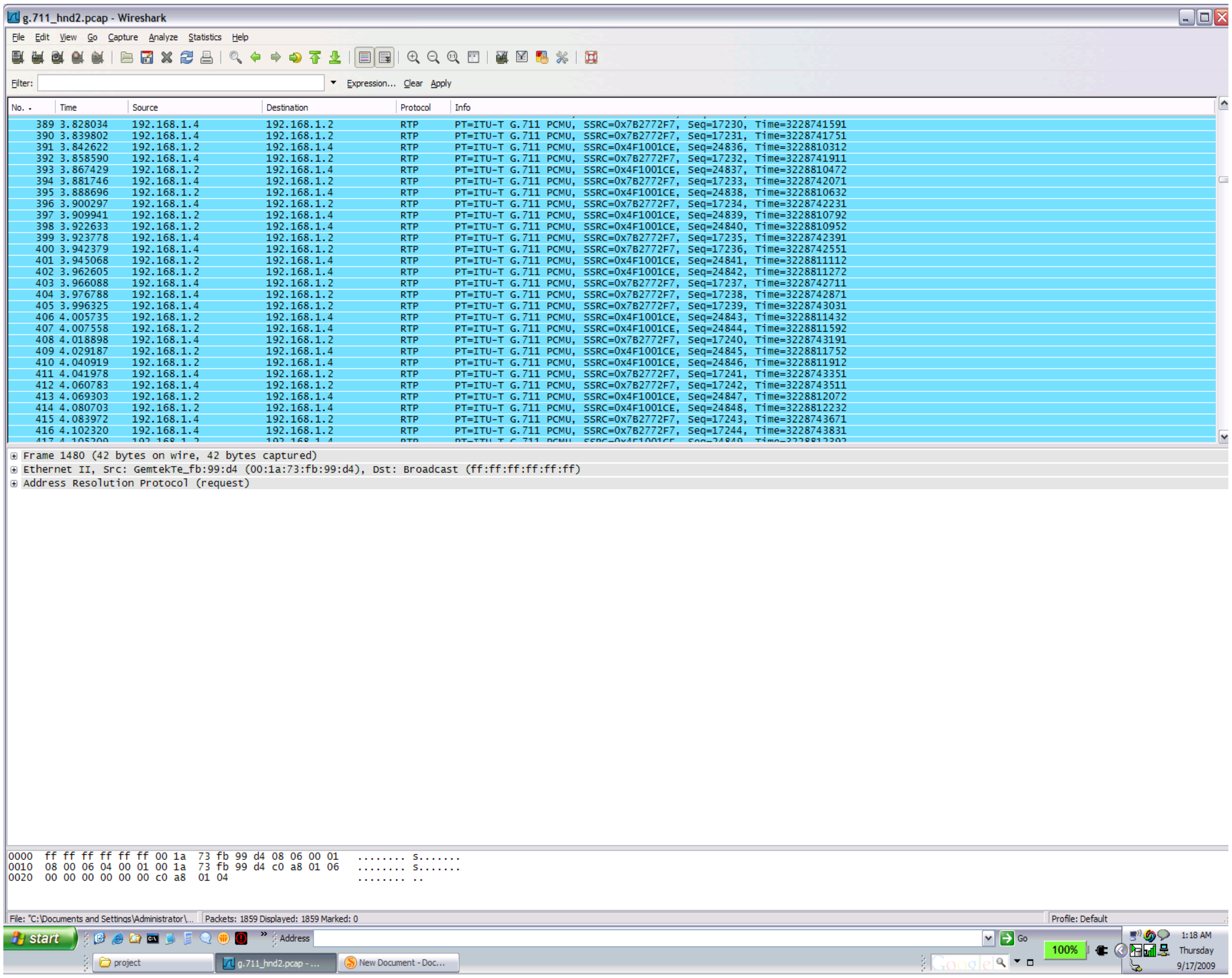


Figure 4.4. Screenshot of Wireshark before handoff

Before handover, the packets are sent between 192.168.1.4 and 192.168.1.2 as shown in figure 4.4. During handover process 192.168.1.4 keeps sending packets to 192.168.1.2 but does not receive any reply back as shown in figure 4.5. In this while ARP request is also broadcasted for the new IP address obtained 192.168.1.6. During this duration, user does not hear anything as packets are lost due to temporary disconnection of call.

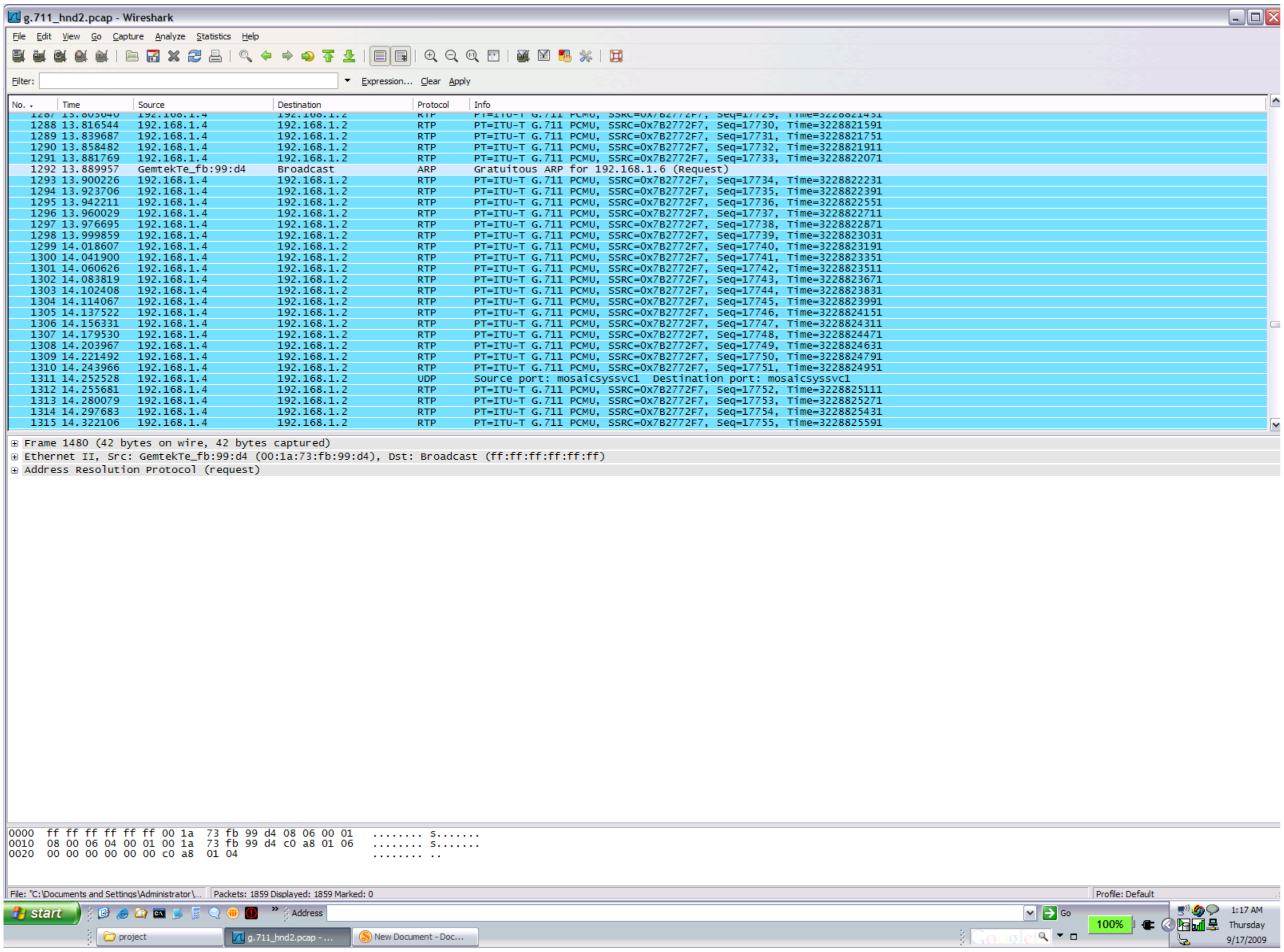


Figure 4.5: Screenshot of Wireshark during handover phase

When handover process is complete, 192.168.1.6 sends packets to 192.168.1.4 and then 192.168.1.4 replies back to 192.168.1.2 whose packets are routed to the new address as shown in figure 4.6 below.

The drawback with E-model calculations using Wireshark tool is that it does not accurately calculate the packet loss during handover. Ethereal showed a 0% - 0.02% packet loss for G.711 and 0.3-1.1% packet loss for iLBC codec whereas the actual packet loss was much more. Thus the E-model calculations for handover scenario show a very high difference between MOS subjective and MOS objective scores as shown in figure 4.2.

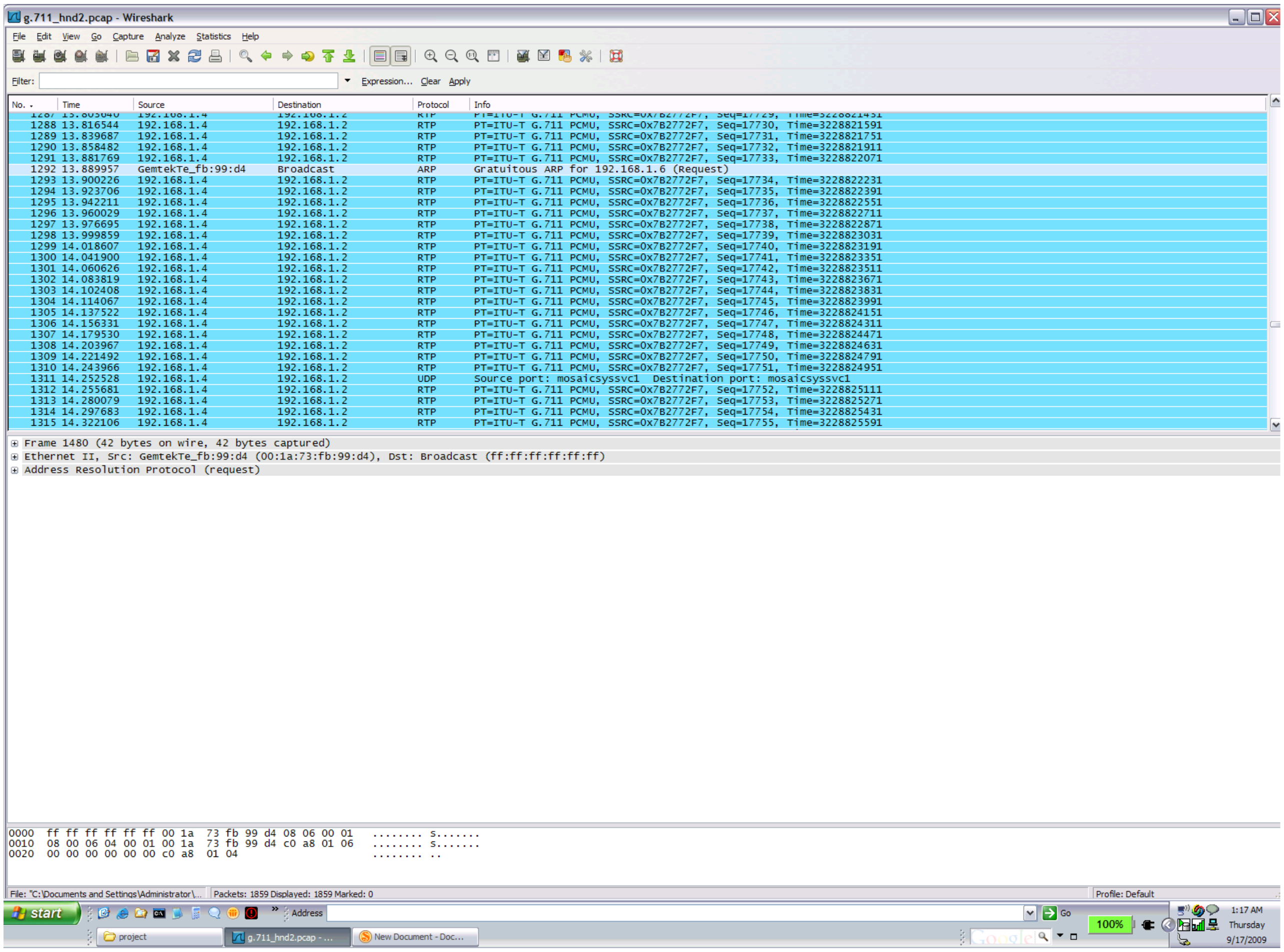


Figure 4.6: Screenshot of Wireshark after handover process is complete

4.3 New Objective Model

The new objective model that I propose is based on studying the Wireshark packets during handover. The handover delay with reference to the handover stages in figure 4.1 can be defined as:

The delay that occurs between the time of disconnection from BS 1 and the time of setting up the connection with the BS2 is the handover delay.

The handover process is summarized in figure 4.7.

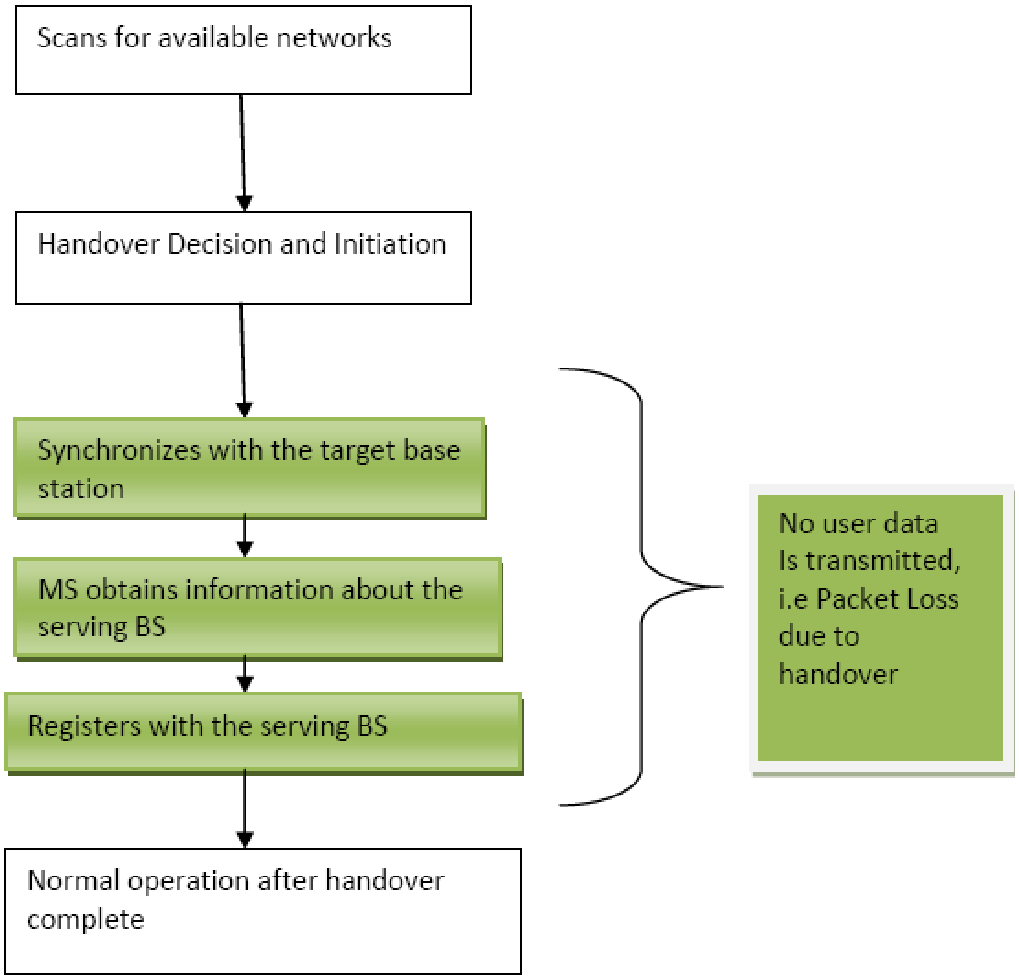


Figure 4.7. Flowchart of handover process

Therefore, from the above flowchart the handover delay can be calculated by measuring Synchronization delay, delay due to ranging information and Registration delay.

Delay (HO) = d(sync) + d(rang) + d(reg) -------(1)

Now the packet loss during handover is also measured by calculating the packets sent during the synchronization, ranging and registration phases. The Wireshark screenshot in figure 4.5 shows the packets that are being sent from 192.168.1.4 to 192.168.1.2 only but at this time user cannot listen to anything, i.e. they are the packets lost during handover.

Therefore handover packet loss will be:

$P(h)$ = No. of packets sent during handover phase / Total packets sent

Thus from the new approach the enhanced E-model equation becomes:

$R = 93.2 - I_{dh} - I_{eh}$

where

I_{dh} **$= 0.024*\Delta+0.11*(\Delta -177.3)*H(\Delta -177.3)$**

Where **Δ = d (sync) + d(rang) + d(reg) + d(net) + packetization delay + processing delay**

Therefore **Δ = D (ho) + d(net) + packetization delay + processing delay** ---- (2)

From E-model Ie was as:

$Ie = a + b \ln(1+cP/100)$

Therefore Ie for handover measurement will be:

$I_{eh} = a + b\ln(1 + c(p+Ph)/100)$

where, **Ph** is the packet loss during handover.

4.4 Summary

This chapter discussed the effect on voice quality during handover. A new objective approach was proposed based on the drawbacks on the E-model. The implementation results of the new approach would be discussed in chapter 6.

5. Analysis and Design

When a Voice call is established using a VoIP client, the voice packets start traversing through the VoIP application (application layer) first and travel all the way through the network to the receiver side. With the layered architecture of the TCP/IP network, as the packet travels through each layer a header and a trailer is inserted in front and back of the packet. At the receiver side, the reverse process takes place to remove the header and trailer from the packet.

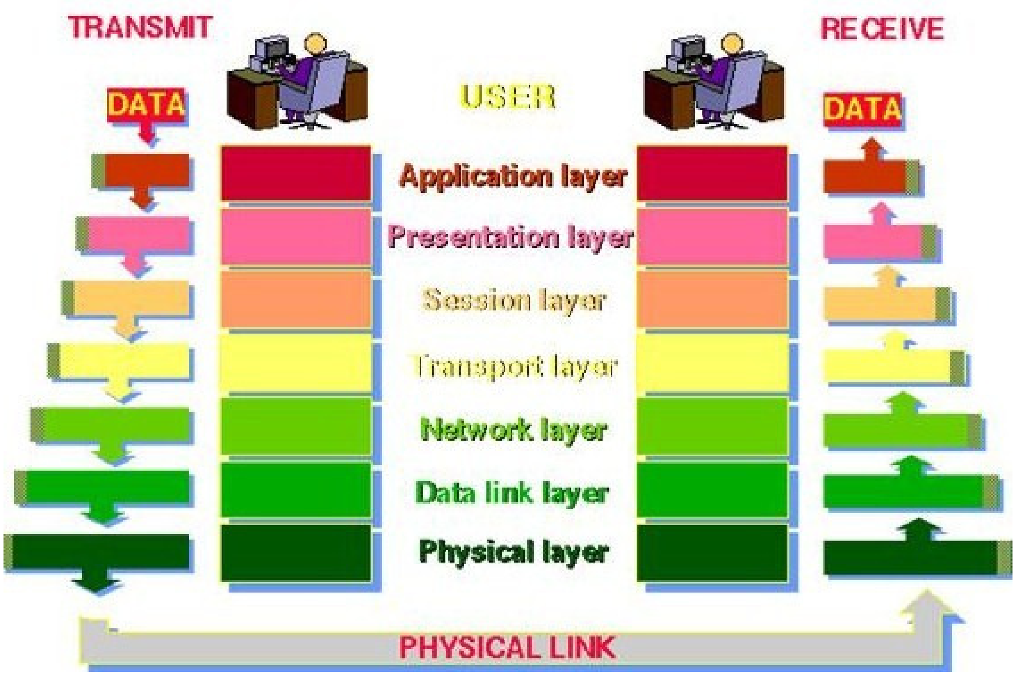


Figure 5.1. Seven layers of OSI model [6]

5.1 Test Setup

The test setup consisted of two windows XP machines that had a VoIP call established. The VoIP client was downloaded on both the laptops and was configured to use RTP ports 8500 for sending and receiving voice packets. One laptop was fixed while the user on the other laptop was mobile during handover. The MOS subjective tests were performed for VoIP call with and without handovers using both the codec's G.711 and iLBC. During each test, packets were captured using Wireshark tool.

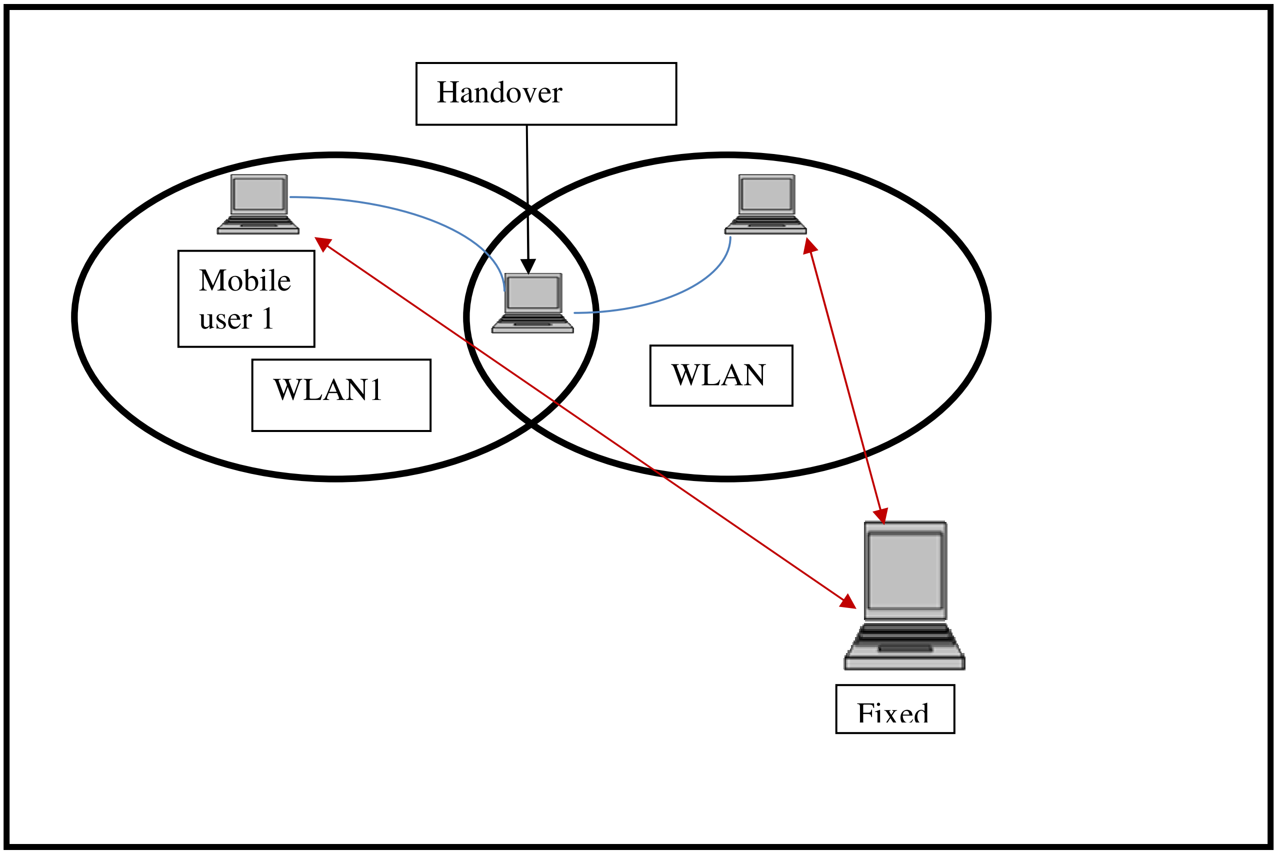


Figure 5.2. Test Setup Scenario

5.2 VoIP Client

The VoIP client used in the project is a freely available VoIP client developed by Global IP Solutions. The VoIP client is designed in Visual C++. The GUI of the VoIP client was modified and the snapshot is shown in figure 5.3 below.

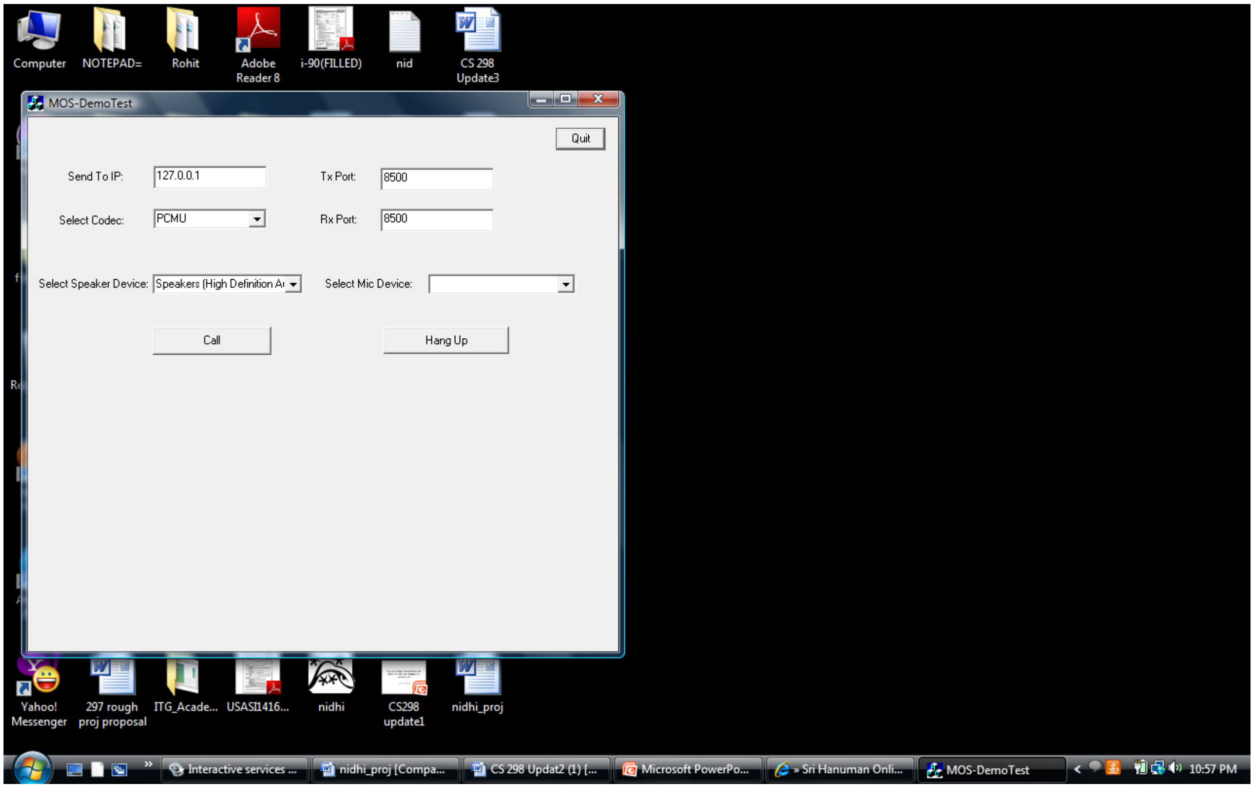


Figure 5.3. VoIP Client

A brief usage of VoIP client is as:

Send To IP: This VoIP client establishes a VoIP call based on IP address. The IP address of the receiver is entered here.

Select Codec: It consists of two codec options: PCMU (G.711) and iLBC. The codec can be selected using this drop down box.

Tx Port: The port is specified on which we want RTP packets to be transmitted. Tx port was set to 8500.

Rx Port: The port on which RTP packets are received. Rx port was set to 8500.

Select Mic Device: It selects the appropriate input device when call is established.

Call: The VoIP call is established by clicking this button. The VoIP packets are sent and received from the specified IP address and the ports.

Hang Up: It is used to disconnect a VoIP call.

5.3 Wireshark

Wireshark is a network analyzing tool; it captures the packets and displays its details. Wireshark helps to understand what is going on in the network in real time voice call. Some of the features of Wireshark are [10]:

1. Captures live packets
2. Displays a detailed protocol information for the packet captured
3. Can save the data packets captured
4. Can apply packet filters
5. Performs stream analysis
6. Captures information like delay, jitter, bandwidth, codec etc for packets

The screenshot of Wireshark is shown in figure 5.4 below.

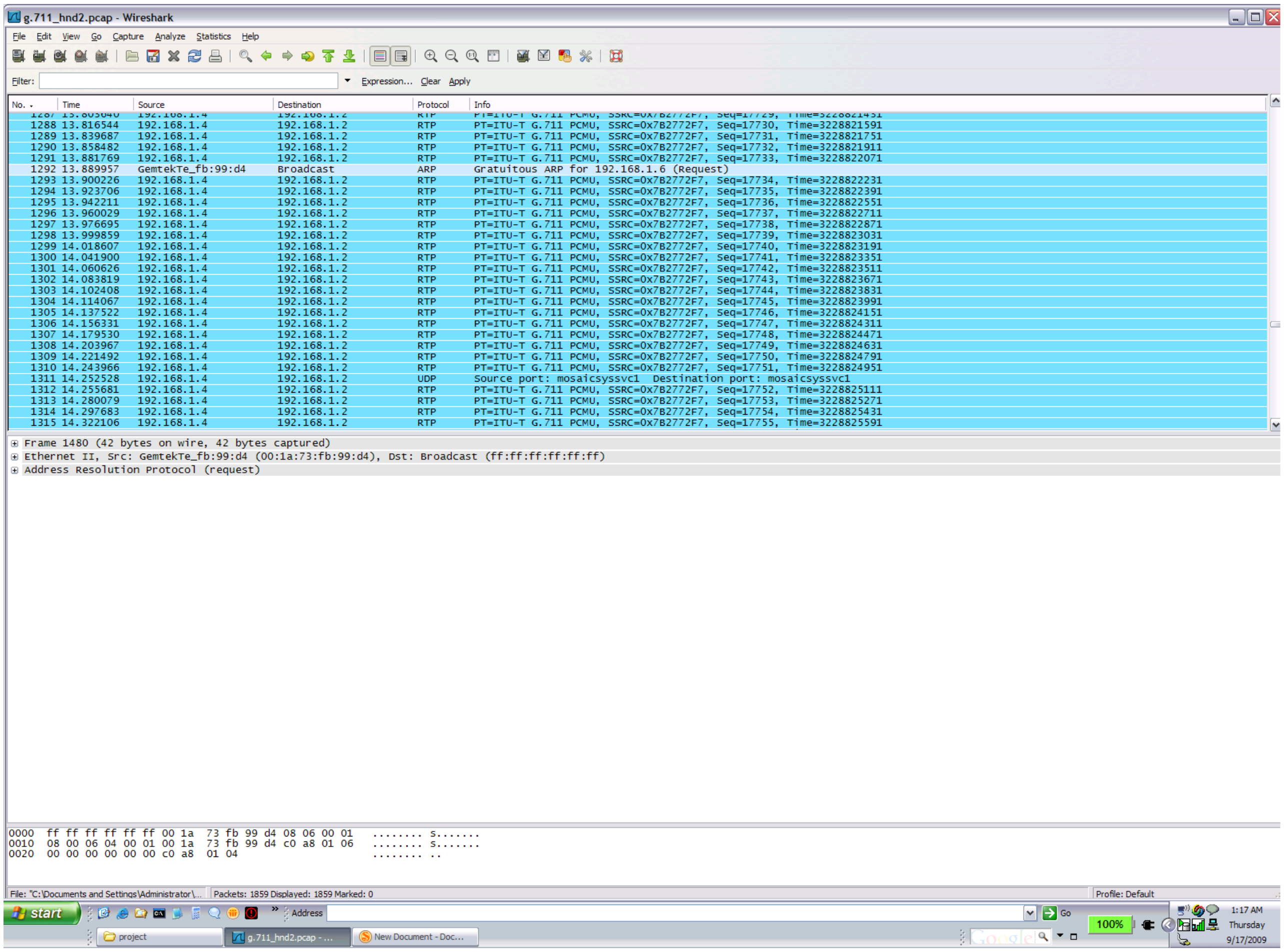


Figure 5.4. Wireshark screenshot

5.4 MOS Score Calculator

The MOS score calculator that I developed is based on the new proposed E-model for handovers. It calculates the R-factor and the corresponding MOS scores for speech samples during handovers using the new approach. This calculator reads data from the Wireshark captured files, thus the main purpose of this tool is to reduce the manual effort involved in reading the packets captured by Wireshark. It automatically calculates the handover packet loss, one way delay values from the captured data, thus making the tedious calculations using the new approach very simple and automated. The screenshot of the calculator is shown in figure 5.5 below.

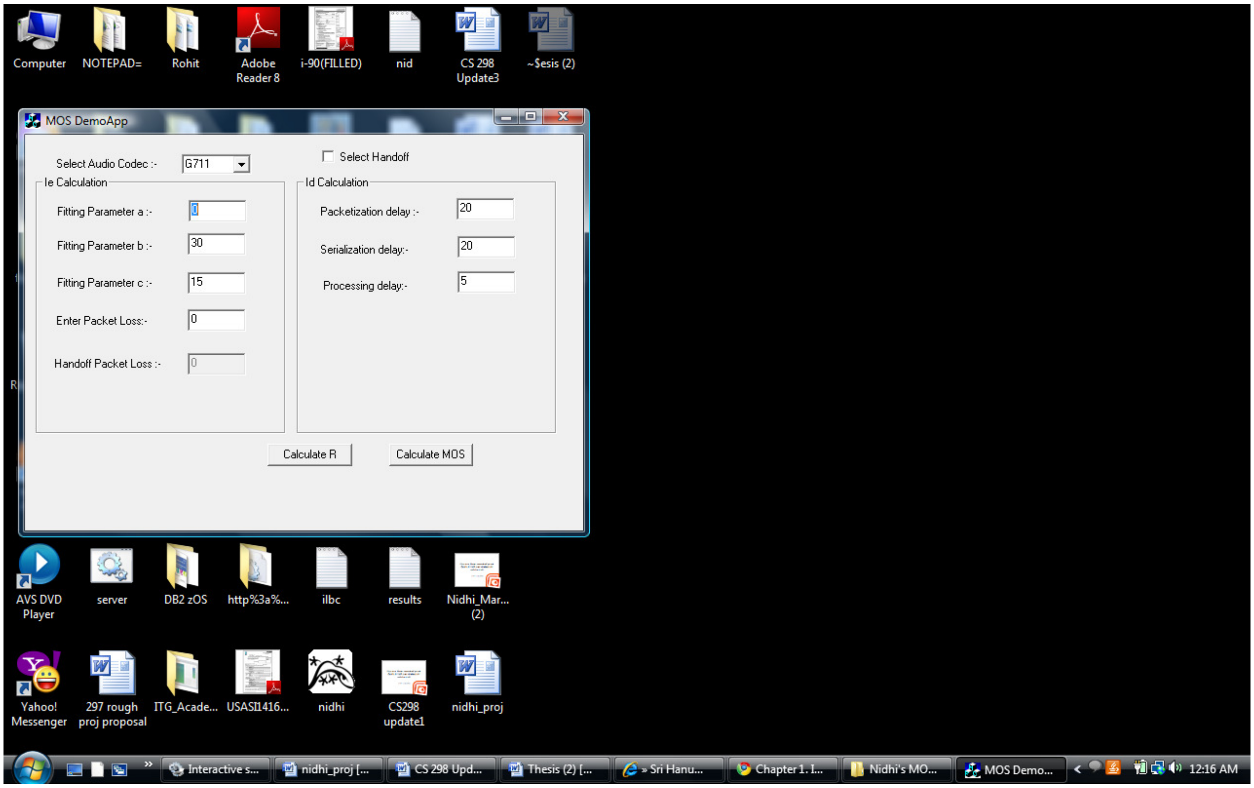


Figure 5.5. MOS Score Calculator

The GUI of the calculator has the following options:

1. Select Audio Codec: It is a drop down box that contains two codec options, iLBC and G.711. On selecting the specified codec, the values of fitting parameters (a, b and c), packetization delay and processing delay are automatically filled as they are codec dependent parameters. For example, when codec G.711 is selected, following values are filled in fields described below:

Fitting parameter a: 0

Fitting parameter b: 30

Fitting parameter c: 15

Processing delay: 5

Packetization delay: 20

2. Select Handoff: It is a checkbox, when selected it uses the new approach to calculate the R-factor during call handover. The screenshot is shown in figure 5.6 below.

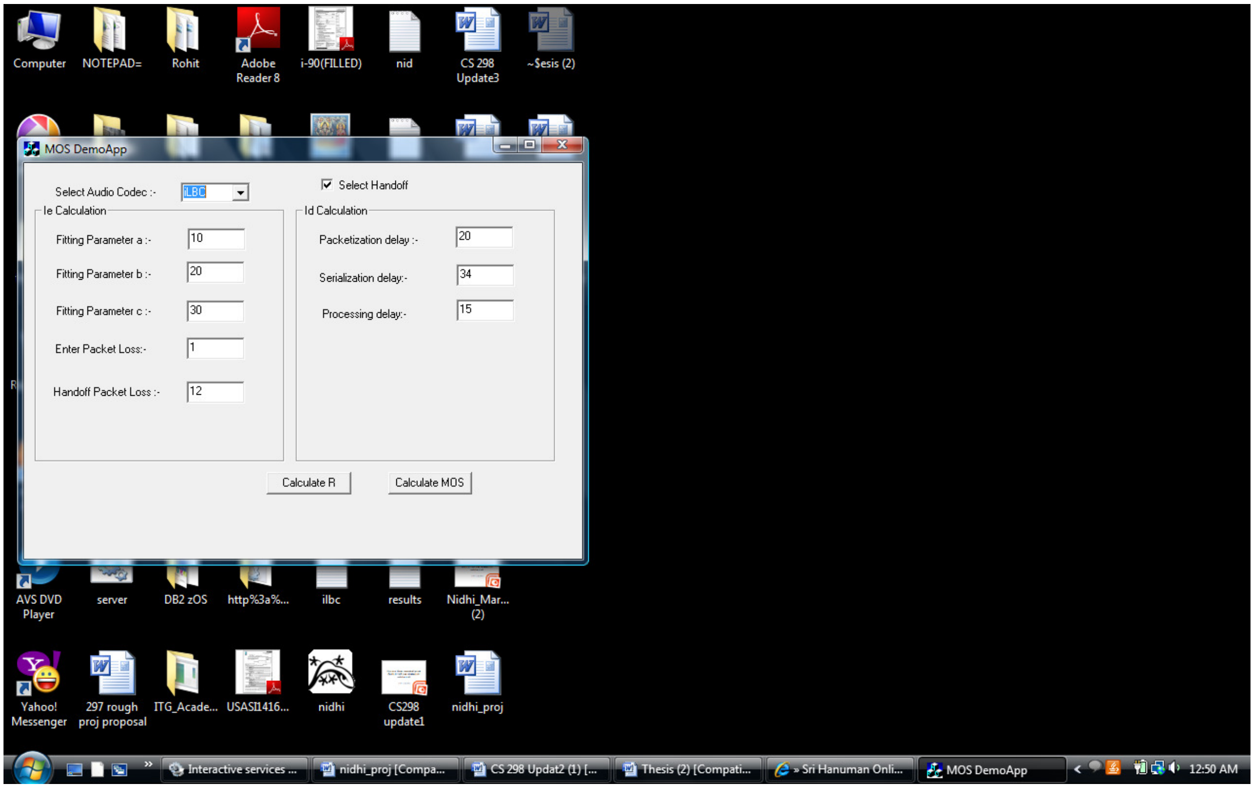


Figure 5.6. Screenshot of MOS calculator when handoff is selected

When Select handoff field is not checked then Handoff Packet Loss field is grayed out, such that no values can be entered. Also when Select Handoff field is checked, the handoff packet loss value is calculated and filled in automatically according to the codec selected.

Handoff Packet Loss: The value of handoff packet loss is automatically entered when select handoff field is checked. It calculates the value based on the codec selected. It reads the Wireshark captured file which is saved as ilbc.txt for iLBC codec and G711.txt for G.711 codec and calculates the handoff packet loss percentage as discussed in the new approach in chapter 4.

Enter Packet Loss: This is the only field where user has to enter value. The value is obtained from the Wireshark file.

Serialization Delay: The value of serialization delay, that is the one way delay is automatically calculated from the Wireshark captured file when the codec is selected. It is calculated by getting the average of the one way delay (delta) from the Wireshark file.

Fitting Parameters (a, b and c), Processing Delay and Packetization delay: These values are automatically filled when a codec is chosen as they are codec dependent parameters.

Calculate R: On clicking this button, the R-factor value is calculated. If handoff is checked then it uses the new approach to calculate the R-factor. Screenshot is shown in figure 5.7.

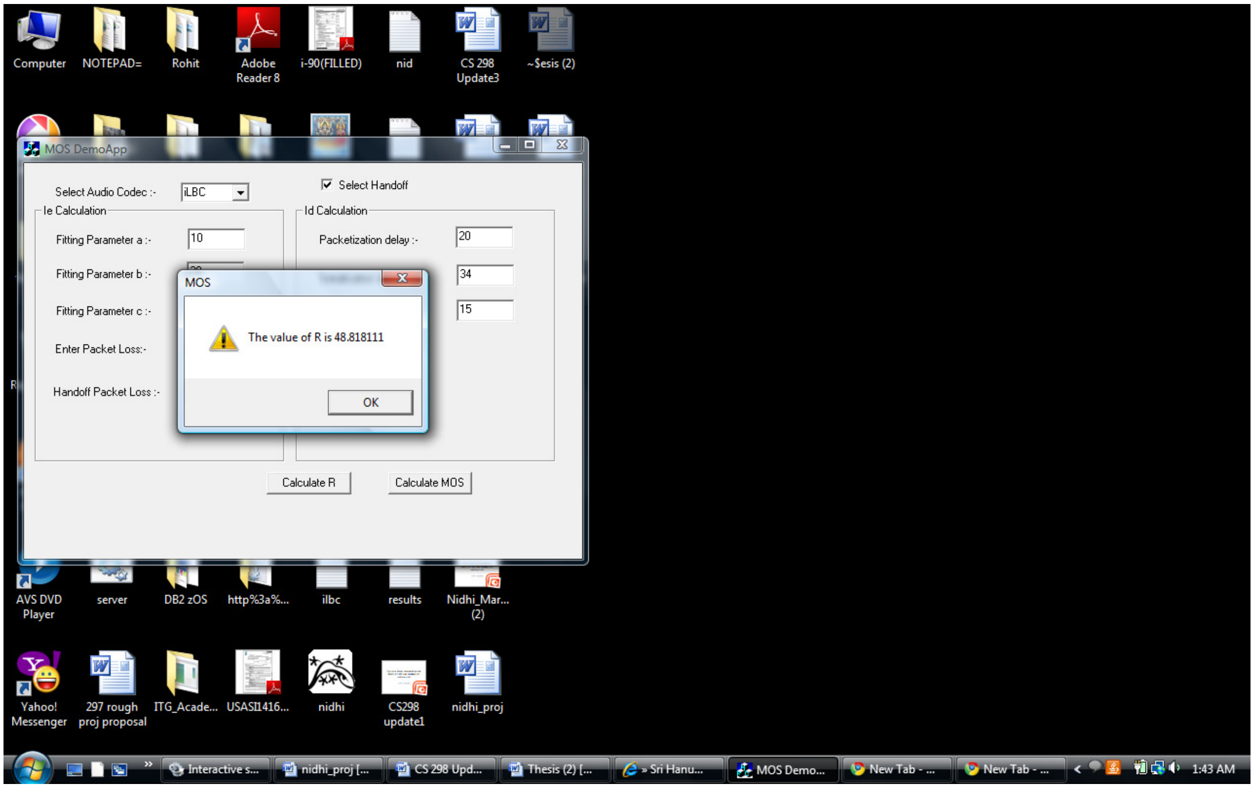


Figure 5.7. Screenshot for calculation of R-factor

Calculate MOS: On clicking this button the R-factor value is mapped to the corresponding MOS score. Screenshot is shown in figure 5.8.

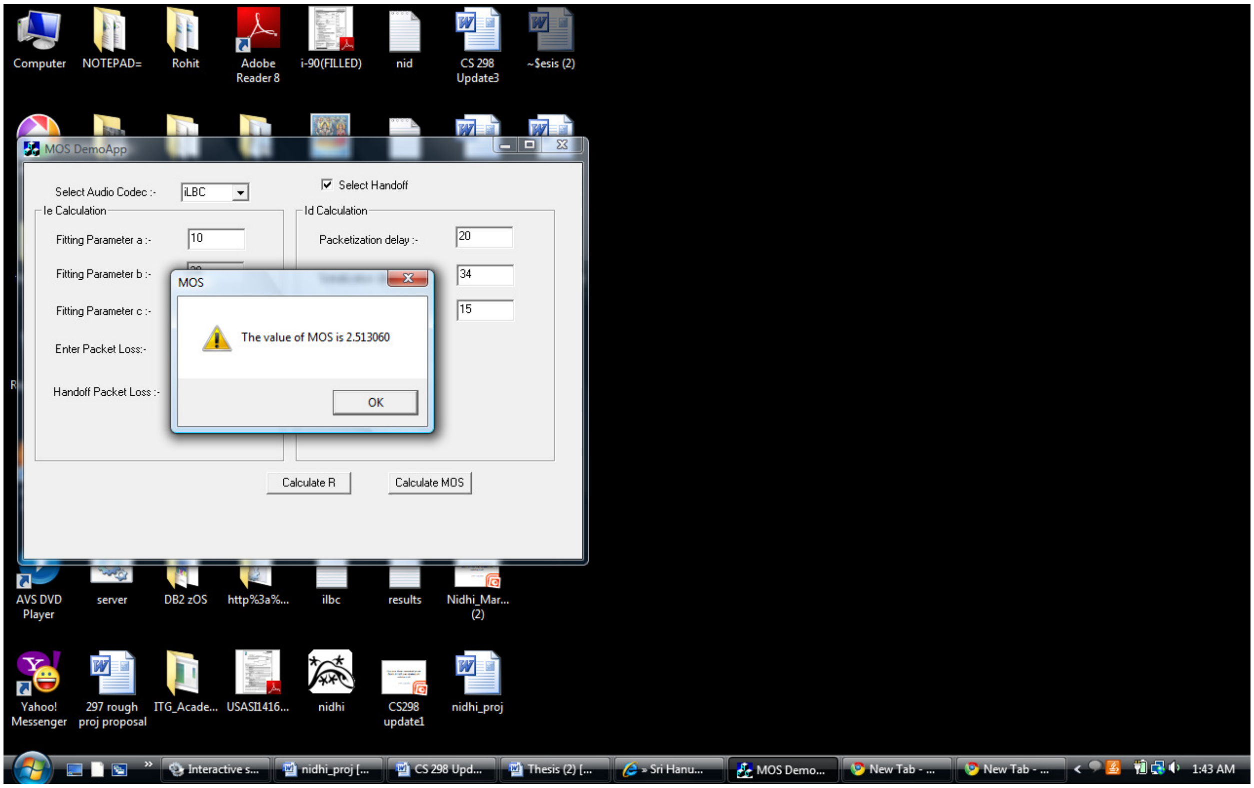


Figure 5.8. Screenshot for calculation of MOS score

The implementation results of the new approach are discussed in chapter 6.

6. Implementation Results

In order to calculate the speech quality of VoIP during handover, firstly MOS test was conducted for 12 participants. This subjective MOS test was performed with and without handover. Ten sample Hindi test sentences were played and participants rated each test sentence based on the quality.

6.1 Scenario without Handover

6.1.1 MOS Score (objective) for G.711

E-model calculations for G.711 without handover:

Delay delta=14ms

Total delay d= delta + packetization delay + processing delay; = 14 + 20 + 5 = 39ms

Id=0.024 *d + 0.11(d-177.3) H (d-177.3) = 0.936

Ie is 0 for G.711

R=Ro-Id-Ie = 92.264

Therefore MOS = 4.39

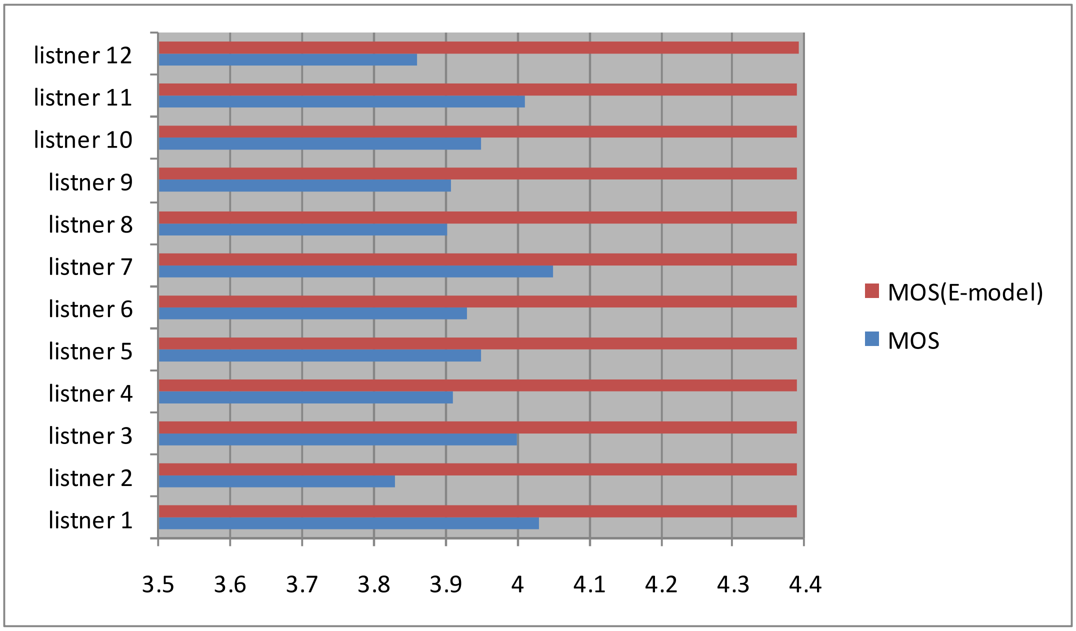


Figure 6.1. Individual Comparison of MOS- subjective and MOS (E-model) for G.711

6.1.2 MOS Score (objective) for iLBC

E-model calculations for iLBC without handover:

Delay delta=14ms

Total delay d= delta + packetization delay + processing delay; = 14 + 20 + 15 = 49ms

Id=0.024 *d + 0.11(d-177.3) H (d-177.3) = 1.176

Ie is 10 for iLBC

R=Ro-Id-Ie = 82.024

Therefore MOS = 4.09

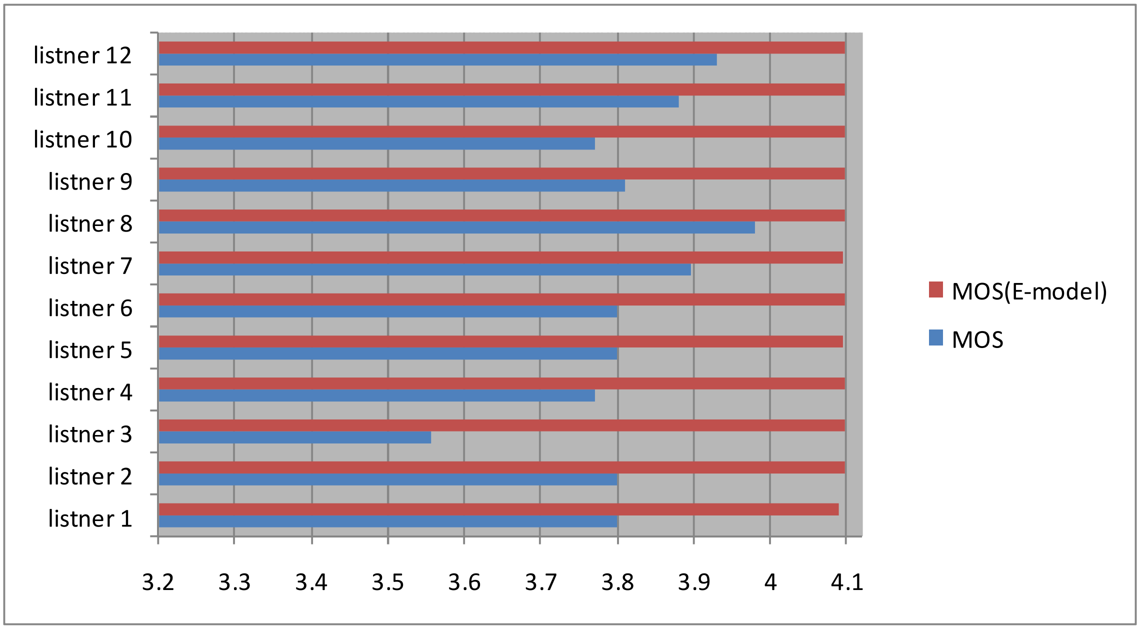


Figure 6.2. Individual Comparison of MOS- subjective and MOS (E-model) for iLBC

The result for G.711 and iLBC without handover can be summarized in figure 6.3 below.

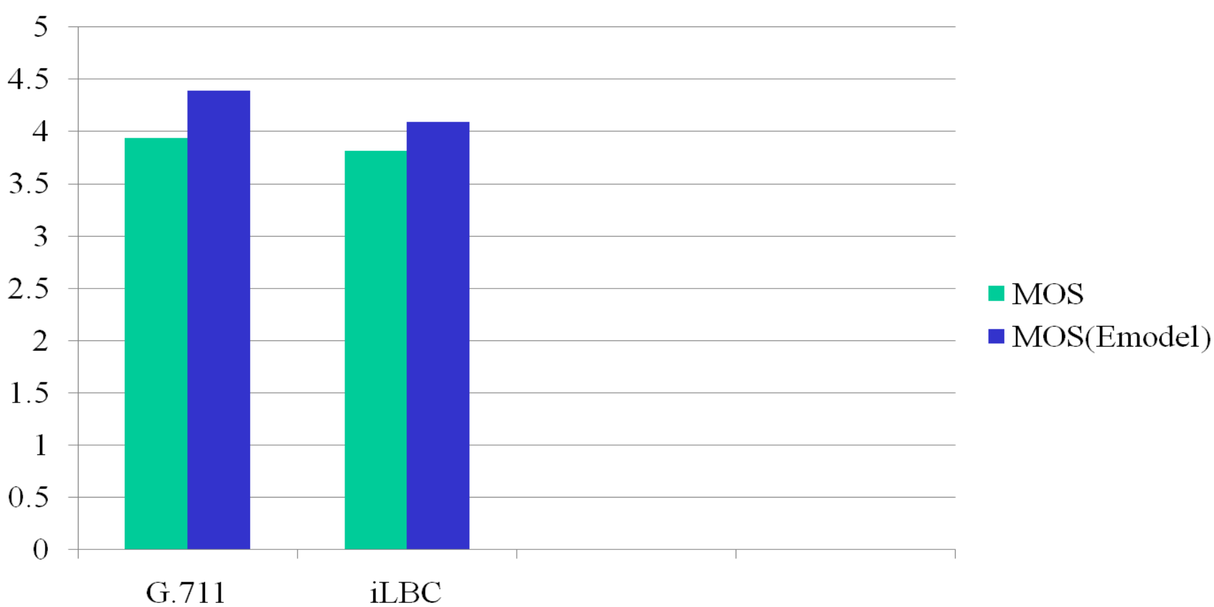


Figure 6.3. Comparison of MOS for G.711 and iLBC codec without handover

6.2.1 Subjective MOS Score

When MOS test was conducted during handover, the listeners experienced a gap (a silence) for a while (that was during the handover phase) and after handover was complete they could hear the test sentences. The MOS test was performed with 12 participants.

Some calls got dropped as the mobile user could not connect to the other wireless network. This happened due to the excessive delay during handover process; the mobile user moved out of the handover region and also did not get authenticated to the new network, leading to a dropped call.

Codec	Average MOS Score
iLBC	3.143
G.711	3.311

Table 6. Average MOS scores for G.711 and iLBC during handover

6.2.2 E-model Calculation for G.711

Avg one way delay =22.3ms

d=22.3+5+20 = 47.3ms

Therefore Id = 0.024*d = 1.135

Packet loss (P) for G.711 was 0.02%

Therefore, Ie = a + bln(1+cP/100)

= 0+30ln(1+.02*15/100) =.0898

Rfactor = 93.2-1.135 - .0898 = 91.97

MOS (emodel) = 1 + 0.035*R + R(R-60)(100-R)7*10^-6

MOS (emodel) = 4.383

6.2.3 E-model calculation for iLBC

Packet loss = 1.1%

Delay = 34 ms

Id = 1.536

Ie = a + bln(1+cP/100)

= 10 + 19.8ln(1 + 1.1*29.7/100)

Ie = 15.597

Rfactor = 93.2-1.536-15.597 = 76.067

Therefore, **MOS (emodel) = 3.864**

The results of MOS (subjective) and MOS (E-model) are summarized in figure 6.4.

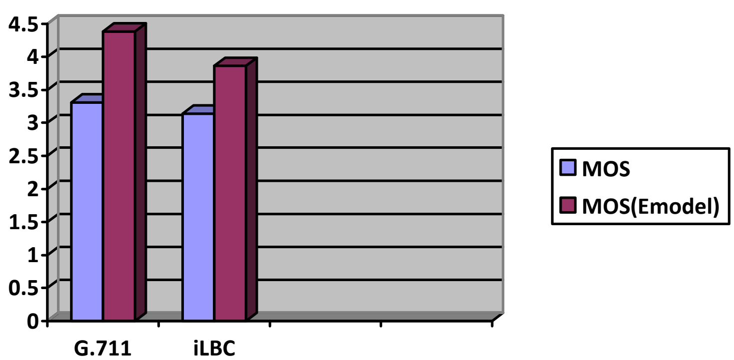


Figure 6.4. Comparison of iLBC and G.711 for handover

6.3 Calculations using New Approach

6.3.1 New E-model calculation for iLBC

Avg one way delay =34ms

d=34+10+20 = 64ms

Therefore Idh = 0.024*d = 1.536

Packet loss (P) for iLBC is 1.1%

Handover packet loss (Ph) = 16%

Therefore, I_{eh} = a + bln(1+c(P+Ph/100))

= 10+19.8ln(1+17.1*29.7/100) = 45.7

Rfactor = 93.2- 45.7 = 47.5

MOS (New-Emodel) = 1 + 0.035*R + R(R-60)(100-R)7*10^-6

MOS (New - Emodel) = 2.469

6.3.2 New E-model calculation for G.711

Avg one way delay =22.3ms

d=22.3+5+20 = 47.3ms

Therefore I_{dh} = 0.024*d = 1.135

Packet loss (P) for G.711 was 0.02%

Handover packet loss (Ph) = 11%

Therefore, I_{eh} = a + bln(1+c(P+Ph/100))

= 0+30ln(1+11.02*15/100) = 29.27

Rfactor = 93.2- 29.27 = 63.93

MOS (New-Emodel) = 1 + 0.035*R + R(R-60)(100-R)7*10^-6

MOS (New - Emodel) = 3.29

6.4 Comparisons

The new approach calculated the MOS score for G.711 very close to the subjective MOS score as compared to the MOS score calculated by E-model. For iLBC, the new approach shows that the voice quality was tremendously degraded during handover. One of the reasons for this is the extremely high jitter for iLBC codec.

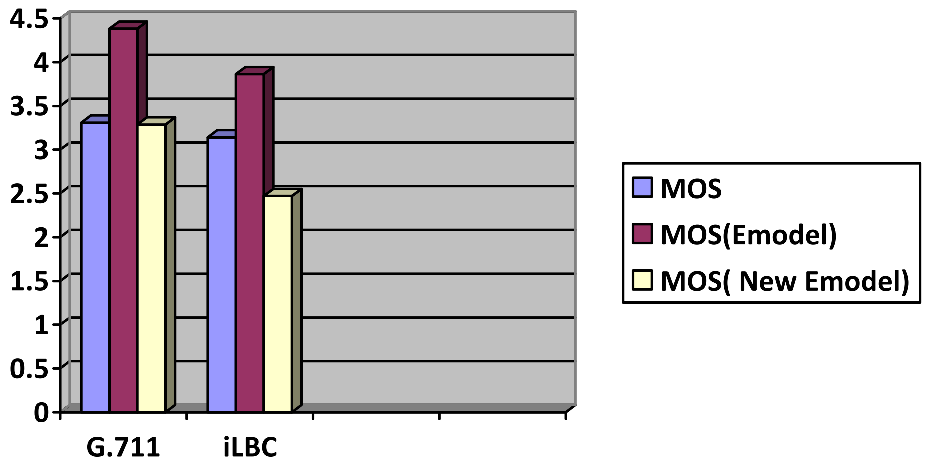


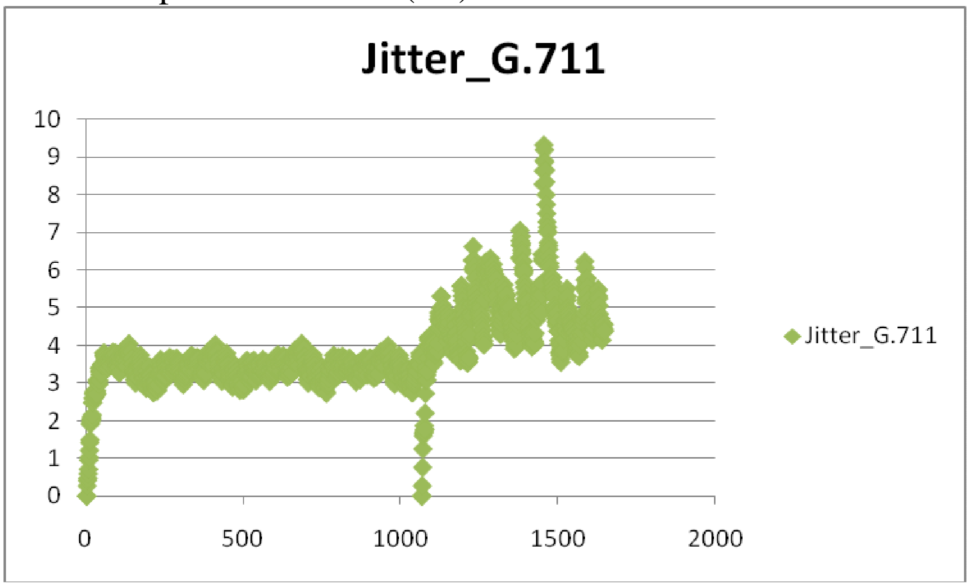
Figure 6.5. Comparison of New E-model with E-model and Subjective MOS

6.5 Correlation between G.711 and iLBC

Various scatter plots were made to correlate between G.711 and iLBC codec during handoff.

They are shown below:

6.5.1 Comparison of Jitter (ms)



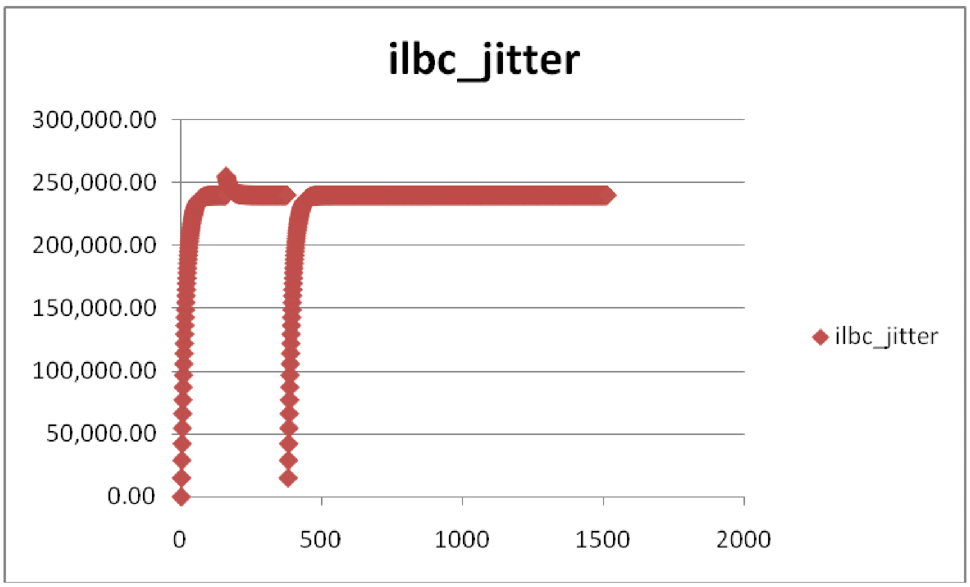
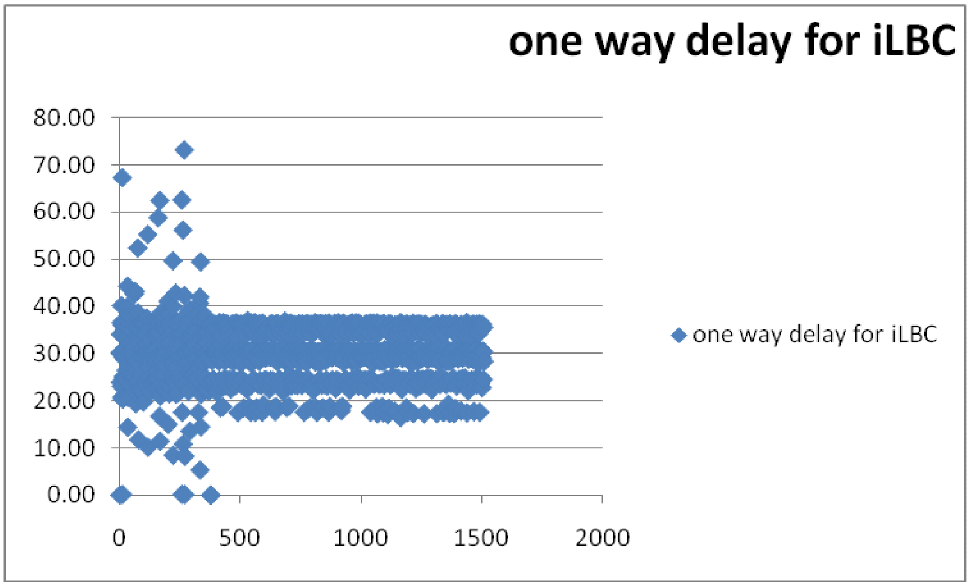


Figure 6.6. Correlation between jitter for iLBC and G.711 during handover

Jitter in (ms) for ilbc was extremely high as compared to G.711. This was one of the reason that voice quality in iLBC is adversely affected during handovers.

6.5.2 One way Delay (ms)



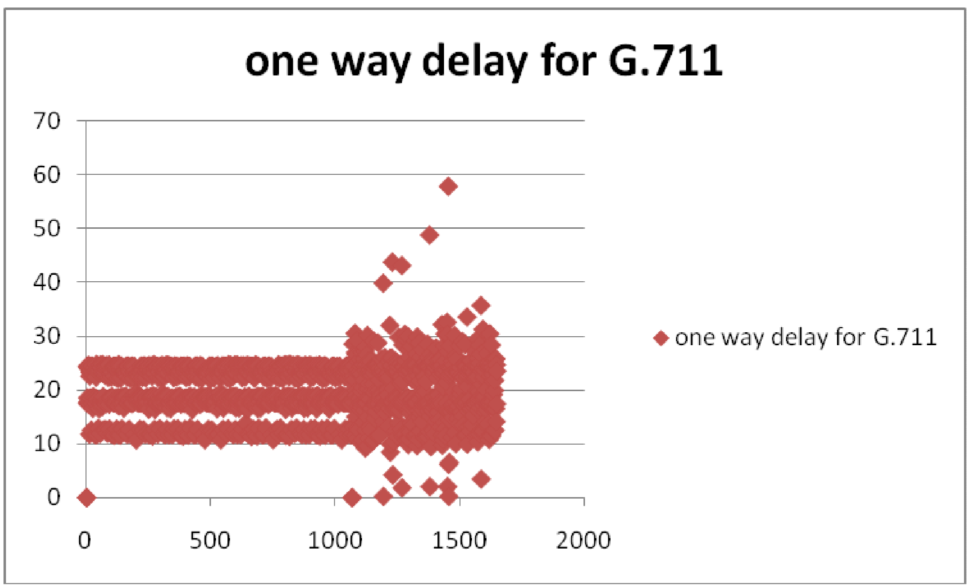
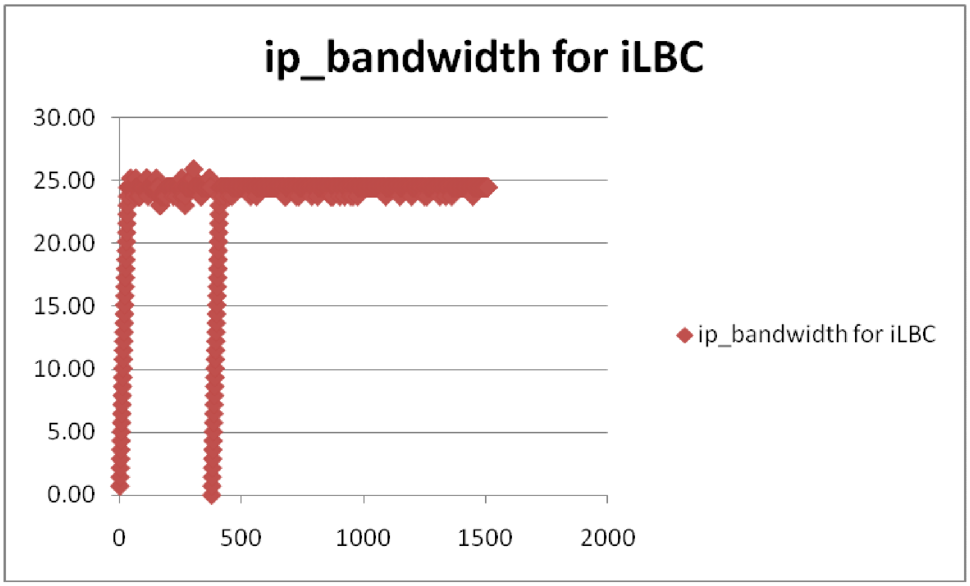


Figure 6.7. Correlation between delay for iLBC and G.711 during handover

For G.711 the delay was less as compared to iLBC. G.711 was mostly between 10-25ms and that for iLBC is between 20-40 ms.

6.5.3 Bandwidth



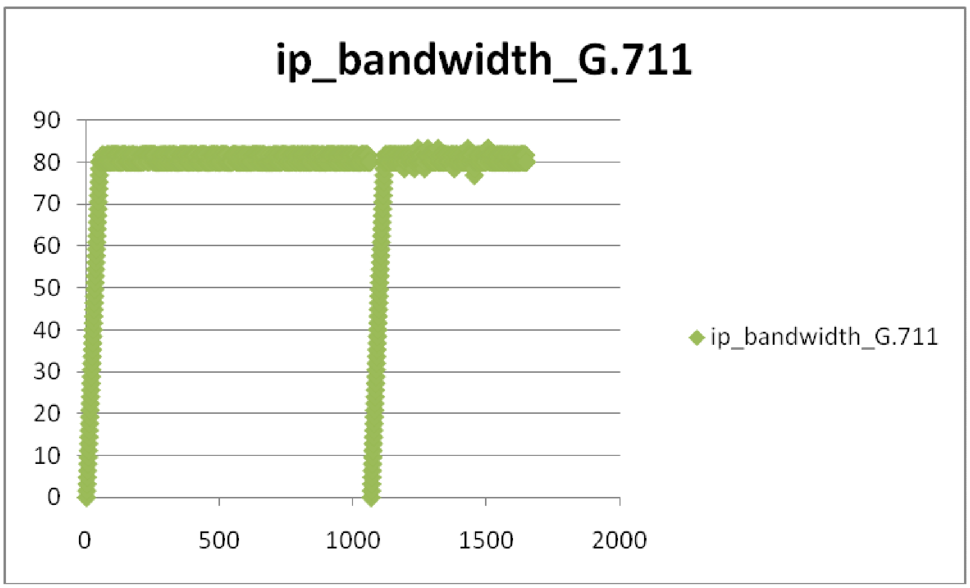


Figure 6.8. Correlation between bandwidth for iLBC and G.711 during handover

The bandwidth required for G.711 is higher as compared to iLBC because of the reason that G.711 is a telephony codec and sends speech in uncompressed format where as iLBC is internet low bit rate code that compresses the speech before sending it over the internet thus requiring less bandwidth.

7. Conclusion and Future Work

7.1 Conclusion

The new approach maps very close to the subjective MOS scores as compared to E-model and helps to calculate the speech quality during handoff much accurately. G.711 codec is a better speech codec than iLBC codec. The speech quality for G.711 is extremely good without handoff, but during a call handoff, the speech quality does degrade but not as much as iLBC. The speech quality for iLBC is tremendously degraded during a call handover and leads to user dissatisfaction.

7.2 Future Work

The work showed in this report builds a MOS score calculator based on the new objective model. This calculator reads the packet files captured by the Wireshark tool to calculate the delay and packet loss during handover. Possible future work could be integrating the new objective approach with a packet capture tool like Wireshark to calculate the speech quality during real time packet capture. This would help to efficiently monitor the speech quality of a live call in real time. This can result in designing an effective algorithm to reduce the handover latency in VoIP.

The new approach is based on E-model which has its own set of advantages and disadvantages. Combining the new approach with the PESQ model could be a possible future work. It could combine the advantages of the PESQ model and calculate the speech quality of VoIP during handovers more efficiently.

8. References

[1] A. Markopoulou, F. Tobagi and M. Karam: *Assessment of VoIP quality over Internet backbones*, in *Proc. IEEE INFOCOM*, New York, NY, June 2002, pp. 150 - 159.

[2] Agora Labs, Speech Codecs , retrieved at http://www.agoralabs.com/speech-codec/g711-codec.htm

[3] Application Note: Tactical Communications Using Cisco IPICS 2.0 across High Latency Networks, Cisco Networks, Whitepaper, 1992- 2007

[4] *Combining Voice over IP with Policy Based Quality of Service.*Technical White Paper. http://www.csd.uch.gr/~hy5366/VoIP.pdf

[5] *Cisco – Voice Over IP – Per Call Bandwidth Consumption*. Technical White paper 2000.

[6] Dagunduro Oluwabunmi, OSI Model The seven dynamic layers, retrieved from http://homepages.uel.ac.uk/u0415051/OSI%20MODEL/

[7] Dong Hoi Kim and Kyungkoo Jun (October 5, 2005). The Effect of the QoS Satisfaction on the Handoff for Real-Time Traffic in Cellular Network. In High Performance Computing and Communications (Volume 3726). Retrieved from http://www.springerlink.com/content/361y8ax45kbg4xvc/

[8] Ditech Networks whitepaper, Limitations of PESQ for Measuring Voice Quality in Mobile and VoIP Networks, 2007.

[9] ePipe VPN and Security Family: Key Networking Concepts, retrieved from http://www.ml-ip.com/html/documentation/vpn-ug-key-concepts-1.html

[10] Ethereal download website, 2007. http://www.ethereal.com

[11] Gabriel Torres, How TCP/IP works – Part 1, retrieved from http://www.hardwaresecrets.com/article/433/3

[12] GlobalIPSound developer community website, 2007.Retrieved from http://developer.gipscorp.com

[13] GIPs iLBC , Retrieved from http://www.softfront.co.jp/products/lib/codecs_iLBC.pdf

[14] H. Schulzrinne, S. Cassner, R. Federick and V. Jacobson, RTP : A Transport Protocol for Real-Time Applications, July 2003

[15] International Telecommunication Union: *The E-model, A Computational Model for Use in Transmission Planning, ITU-T Recommendation G.107*, July 2000.

[16] ITU-T P.833, Methodology for derivation of equipment impairment factors from subjective listening-only tests, 2001.

[17] J Light & A. Bhuvaneshwari: *Performance Analysis of Audio Codecs over Real-Time Transmission Protocol (RTP) for Voice Services over Internet Protocol,* in Proceedings of the Second Annual Conference on Communication Networks and Services Research (CNSR'04), Fredericton, N.B., Canada, May 19 - 21, 2004.

[18] Jane Dudman & Gaynor Blackhouse: *Voice over IP: What it is, why people want it, and where it is going,* JISC technology and Standards Watch, September 2006

[19] Jean-Yves Le Boudec,Andrez Duda and Patrick Thiran, Real Time Protocol (RTP), Retrieved from http://icapeople.epfl.ch/thiran/CoursED/RTP.pdf

[20] Jon Crowcroft, Mark Handley, and Ian Wakeman. Internetworking multimedia. Taylor & Francis, London, 1999.

[21] John Q. Walker and Jeffrey T. Hicks: *The Essential Guide to VoIP Implementation and Management,* NETIQ, 2002.

[22] Jonathan Davidson and James Peters: Voice over IP Fundamentals", Cisco Press, Indianapolis, 2000.

[23] Lijing (Eric) Ding," Intro to VoIP Quality Measurements", Article by Texas Instrument, July 2008.

[24] Pincy C. Mehta, Sanjay Udani: *Overview of Voice over IP*, Technical Report MS-CIS-01-31, February 2001. Retrieved from http://www.cis.upenn.edu/~udani/papers/OverView VoIP.pdf

[25] R.G Cole and J,Rosenbluth, "Voice over IP performance monitoring" , Journal on Computer Communications Review, vol. 31, April 2001.

[26] RFC Sourcebook, Volume 10 Number 4, Dec 2008

[27] S.Andersen and A.Duric, "Internet Low Bit-Rate Codec (iLBC),IETF Draft," Feb 2002.

[28] Shengquan Wang, Zhibin Mai, Walt Magnussen, Dong Xuan and Wei Zhao: *Implementation of QoS- Provisioning System for Voice over IP*, In the Proceedings of the Eight IEEE Real time and Embedded Technology and Applications Symposium, 2002.

[29] S. Pracht, D. Hardman:Voice Quality in Converging Telephony and IP Networks, Agilent Technologies, White Paper, http://literature.agilent.com /litweb/pdf/5980-0989E.pdf

[30] Splicenet, Article on VoIP Integrity retrieved from http://www.splicenet.co.uk/voip.htm

[31]Timothy A. Hall, "Objective Speech Quality Measures for Internet Telephony".

[32] The Primer, Article on Jitter, November 2007 retrieved from http://www.hometheaterhifi.com/the-primers/131-jitter-defined.html

[33] QoE Systems. Retrieved from http://www.qoesystems.com/audio