DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

VOIP VOICE QUALITY MEASUREMENT BY NETWORK TRAFFIC ANALYSIS

by
Farid BAHRAMI VAIGHAN

November, 2011
İZMİR

VOIP VOICE QUALITY MEASUREMENT BY NETWORK TRAFFIC ANALYSIS

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Master of Science
in Electrical and Electronics Engineering.

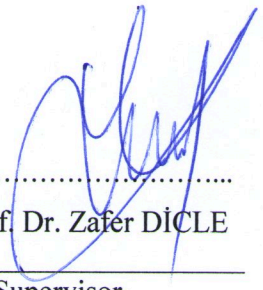
by

Farid BAHRAMI VAIGHAN

November, 2011
İZMİR

M.Sc THESIS EXAMINATION RESULT FORM

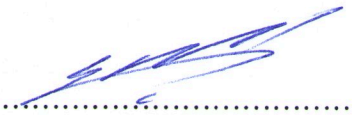
We have read the thesis entitled "**VOIP VOICE QUALITY MEASUREMENT BY NETWORK TRAFFIC ANALYSIS**" completed by **FARID BAHRAMI VAIGHAN** under supervision of **ASST. PROF. DR. ZAFER DICLE** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.



.....

Asst. Prof. Dr. Zafer DİCLE

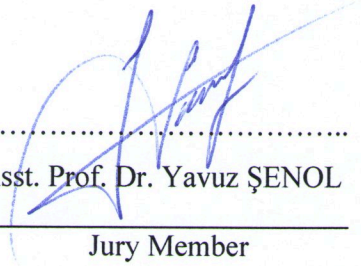
Supervisor



.....

Prof. Dr. Yalçın ÇEBİ

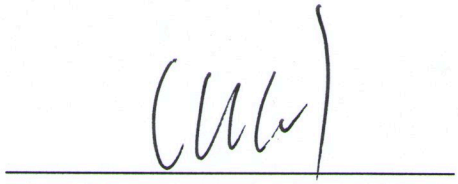
Jury Member



.....

Asst. Prof. Dr. Yavuz ŞENOL

Jury Member



Prof. Dr. Mustafa SABUNCU

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENTS

I would like to thank my supervisor, Asst. Prof. Dr. Zafer DİCLE, for his support, dedication and attention to detail. I would like to especially acknowledge his ability to identify the most important and challenging problems and provide guidance to solve those problems in the most elegant way possible.

Special thanks should be given to Asst. Prof. Dr. Yavuz ŞENOL for direction and motivation that helped to complete the work. I am indebted to him for his help in obtaining this degree.

I would like to thank my committee members, Prof. Dr. Yalçın ÇEBİ and Asst. Prof. Dr. Yavuz ŞENOL for their valuable suggestions and support.

Finally, I want to thank my family and my wife for their support and tolerance during this work and all my life.

Farid. BAHRAMI VAIGHAN

VOIP VOICE QUALITY MEASUREMENT BY NETWORK TRAFFIC ANALYSIS

ABSTRACT

In today's global world, the ability to communicate with people in distant locations with the low cost is increasingly important. In the absence of face-to-face interactions, voice communication is considered one of the most effective forms of communication. The increasing expectation levels for better audio performance has led to the need to understand the behavior of audio traffic as it affects end user perceived quality of the voice over the communication networks.

This research addresses Voice over Internet Protocol communication systems and focuses on the fundamental understanding and assessment of their voice quality.

After presenting basic concepts of VoIP, performance assessment methods are described and then the non-intrusive, objective voice quality assessment method that was developed in this thesis is demonstrated. In this method, real network's data flows of live VoIP calls were collected in packet-size and analyzed; E-model's R value and objective MOS value were then calculated.

Next, developed methods were tested by designing three different test beds and test scenarios. Voice calls were established by X-Lite and Blink SIP soft phones and monitored by Wireshark software in LAN, WLAN and WAN networks. To estimate the method accuracy, voice quality were investigated subjectively by different end users.

Finally, objective and subjective MOS correlation values indicate the assessment method is deemed to be an effective measurement of voice quality.

Keywords: VoIP, E-model, MOS, R value, Correlation value.

NETWORK TRAFİK ANALİZİYLE VOIP SES KALİTESİ ÖlÇÜMÜ

ÖZ

Günümüz global dünyasında, birbirlerinden uzakta bulunan insanların düşük maliyet ile haberleşebilmesinin önemi giderek artmaktadır. Yüz yüze görüşmenin mümkün olmadığı hallerde; ses iletişimi, etkili bir iletişim biçimi olarak göz önüne alınır. Daha iyi ses performansına olan beklenti seviyesinin artması, iletişim ağı üzerindeki ses trafiği son kullanıcının algıladığı ses kalitesini etkilediğinden dolayı, ses trafiği davranışının anlaşılması gerekliliğine neden olmuştur.

Bu araştırma, Internet Protokolü üzerinden sesli iletişim (VoIP) sistemlerinden söz etmekte ve bu sistemlerin ses kalitelerinin değerlendirilmesi ve temel anlayış üzerinde durmaktadır.

VOIP hakkında temel kavramların sunulmasından sonra, performans değerlendirme metotları tanımlanmıştır ve daha sonra araya girilmeden yapılan ve bu tezde geliştirilen objektif ses kalite değerlendirilmesi gösterilmiştir. Bu metotta gerçek bir ağ üzerinde VoIP aramaları sırasındaki veri akışları paketler şeklinde toplanmış ve analiz edilmiştir. Sonra ise E-Modelin R değeri ve objektif MOS değeri hesaplanmıştır.

Daha sonra, geliştirilen metotla üç farklı test yatağı ve test senaryosunun tasarlanmasıyla test edilmiştir. Sesli aramalar X-Lite ve Blink SIP yazılımları ile gerçekleştirilmiş, LAN,WLAN ve WAN ortamlarında Wireshark yazılımı ile gözlenmiştir. Bu metodun doğruluğunu tahmin etmek amacıyla, ses kalitesi farklı kullanıcılar tarafından subjektif olarak incelenmiştir.

Sonuç olarak, objektif ve subjektif MOS korelasyon değerleri, bu değerlendirme metodunun, ses kalitesi ölçümünde verimli olduğunu göstermektedir.

Anahtar sözcükler: VoIP, E-modeli, MOS, R değeri, Korelasyon değeri.

CONTENTS **Pages**

M.Sc THESIS EXAMINATION RESULT FORM ii

ACKNOWLEDGMENTS iii

ABSTRACT iv

ÖZ v

CHAPTER ONE - INTRODUCTION 1

1.1 Motivation & Goals 1

1.2 Outline 2

CHAPTER TWO - BACKGROUND 3

2.1 Communication Networks 3

2.1.1 Circuit-switched network 3

2.1.2 Packet-switched network 4

2.2 VoIP Concept 6

2.2.1 Architecture 7

2.2.2 VoIP versus PSTN 8

2.2.2.1 VoIP Advantages 8

2.2.2.2 Comparisons 10

2.3 VoIP Network 11

2.3.1 VoIP Protocols 11

2.3.1.1 Internet Protocol (IP) 11

2.3.1.2 Transitions Protocols (TCP & UDP) 12

2.3.1.3 Media Protocols (RTP & RTCP) 13

2.3.2 VoIP Call-Signaling Protocol Stacks 15

2.3.2.1 H.323 15

2.3.2.1.1 H.323 Elements. 15
2.3.2.1.2 H.323 protocol suite 16
2.3.2.1.3 H.323 Call Control Signaling. 18
2.3.2.2 SIP 20
2.3.2.2.1 SIP Systems Structure. 20
2.3.2.2.2 SIP Call Flow Sequence. 22
2.3.2.2.3 SIP protocol stack format. 23
2.4 VoIP Voice Codecs 24
2.4.1 Codec Description 24
2.4.2 Codec Selection Criteria 27
2.5 VoIP Security 28

CHAPTER THREE – VOIP VOICE QUALITY 30
3.1 General Telephony Impairments 30
3.2 Additive vs. Subtractive Distortion 32
3.3 Non-linearity and Time-variance 32
3.4 Human Perception's Role 33
3.5 Listening Quality vs. Conversational Quality 34
3.6 VoIP Conversation 34
3.7 Delay or Latency 36
3.8 Jitter 37
3.9 Packet Loss 38
3.10 Echo 40
3.11 VoIP Quality of Service (QoS) 41

CHAPTER FOUR – VOICE QUALITY MEASUREMENT IN VoIP 43

4.1 Overview of VoIP measurement methods 44

4.2 Subjective assessment of quality 45

4.2.3 Mean Opinion Score (MOS) 46

4.3 Objective voice quality measures 47

4.3.1 PSQM 48

4.3.2 PAMS 49

4.3.3 PESQ 49

4.3.4 E-Model 49

4.4 VoIP Voice Quality Measurement by Network Traffic Analysis 52

4.4.1 Call Setup, Call Completion, and Services Testing 54

4.4.2 Packet Performance Testing 54

CHAPTER FIVE – EXPERIMENTAL WORKS & RESUALTS 55

5.1 Test Bed Design 55

5.1.2 Scenario A 55

5.1.3 Scenario B 57

5.1.4 Scenario C 57

5.2 Test bed Experiments 58

5.2.3 Trace Files Information 58

5.2.2 Call Setup Analysis 59

5.2.3 Packet Performance Analysis 61

5.2.4 Voice Quality Analysis 62

5.3 The Measurement Method Accuracy Analysis 62

CHAPTER SIX - CONCLUSION 64

6.1 Conclusion 64

6.2 Future Work 65

REFERENCES 66

CHAPTER ONE

INTRODUCTION

In today's global world, the ability to reliably communicate with people in distant locations is increasingly important. In the absence of face-to-face interactions, voice communication is considered one of the most effective forms of communication.

Since the telephone was invented in the late 1800s, telephone communication has not changed substantially. Of course, new technologies have improved on this invention, but the basic functionality is still the same.

In the 1990s, a number of individuals in research environments, both in educational and corporate institutions, took a serious interest in carrying voice and video over IP networks. This technology is commonly referred to today as VoIP and is, in simple terms, the process of breaking up audio or video into small chunks, transmitting those chunks over an IP network, and reassembling those chunks at the far end so that two people can communicate using audio and video.

VoIP allows something else: the ability to use a single high-speed Internet connection for all voice, video, and data communications. This idea is commonly referred to as convergence and is one of the primary drivers for corporate interest in the technology. The benefit of convergence should be fairly obvious: by using a single data network for all communications, it is possible to reduce the overall maintenance and deployment costs. The benefit for both home and corporate customers is that they now have the opportunity to choose from a much larger selection of service providers to provide voice and video communication services. In short, VoIP enables people to communicate in more ways and with more choices.

1.1 Motivation & Goals

The advantages outlined above, have caused the number of VoIP users are increased day by day. Commensurate with the increased demand, the number of

VoIP service providers is on the rise. In this situation, the users need to have tools to make distinguish which provider is better to choose. Also in today's competitive environment, many of companies have been going to use Voice over Internet Protocol (VoIP) to get cut down expense, boost productivity and increase their competitive advantage. They should to verify (before deployment) that their infrastructure should be work on VoIP system; thoroughly examine all system elements during deployment; and govern their VoIP system proactively after deployment, including ongoing monitoring, troubleshooting, and planning for future growth.

In order to meet these needs and to measure voice quality, a method, focused on the network characteristics by using open source call establishment and net monitoring software, is aimed to develop.

1.2 Outline

The rest of this thesis is organized as follows: Chapter 2 provides the background information about VoIP network, popular VoIP protocols, VoIP voice codec, VoIP security and VoIP systems scenarios.

Chapter 3 outlines concepts related to understanding voice quality in VoIP and the factors that affect them.

In chapter 4, after describing some existing voice quality measurement methods; the method that developed in this study is described.

Experimental analysis and the results are described in chapter 5. Finally, Chapter 6 summarizes the thesis and outlines possible future work.

CHAPTER TWO

BACKGROUND

2.1 Communication Networks

A communication network is a collection of terminals, links, and nodes which connect together to enable communication between users via their terminals. The network sets up a connection between two or more terminals by making use of their source and destination addresses (Fiche & Hébuterne, 2004).

Switched networks are divided into circuit-switched and packet-switched networks. The packet-switched networks are further divided into connection oriented and connectionless packet networks (Kurose & Ross, 2003; Tanenbaum, 2003; Stallings, 1997). Figure 2.1 shows this classification.

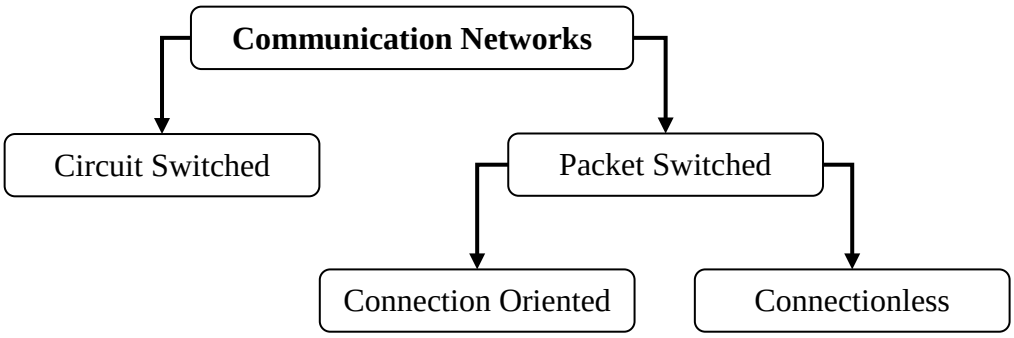


Figure 2.1 Communication network classifications.

2.1.1 Circuit-switched network

Besides voice transport, circuit-switched networks are regularly used to transport different traffic types, such as data and control signals between computers and terminals, respectively. However, no matter which traffic type is transported, the user equipment and the set of nodes are called terminal and network, respectively. The network establishes the communication path between the terminals. The path is a

connected sequence of links between nodes. The communication via circuit-switched networks implies that there is a dedicated communication path between two or more terminals all through the communication session. Therefore, the resources (links and nodes) are reserved exclusively for information exchanges between origin and destination terminals. (Kashihara, 2011)

This communication involves three phases: circuit establishment, data transfer, and circuit disconnect. Before communication can occur between the terminals, a circuit is established between them. Thus, link capacity must be reserved between each pair of nodes in the path, and each node must have available internal switching capacity to handle the requested connection. The nodes must have the intelligence to make these allocations and to devise a route through the network. In circuit-switched networks, the nodes do not examine the contents of the information transmitted; the decision on where to send the information received is made just once at the beginning of the connection and remains the same for the duration of the connection. Thus, the delay introduced by a node is almost negligible. After the circuit has been established, the transmission delay is small and it is kept constant through the duration of the connection. The circuit-switched networks can be rather inefficient. Once a circuit is established, the resources associated to it cannot be used for another connection until the circuit is disconnected. Therefore, even if at some point both terminals stop transmitting, the resources allocated to the connection remain in use. The most common examples of circuit-switched network are the PSTN and the Integrated Services Digital Network (ISDN). (Kashihara, 2011)

2.1.2 Packet-switched network

The data traffic is burst, and non-uniform. Terminals do not transmit continuously, i.e., are idle most of the time and very burst at certain time. Data rates are not kept constant through the duration of the connection but they vary dynamically. A particular data transmission has a peak and average data rates associated to it and these are usually not the same. Therefore, employing dedicated circuits to transmit traffic with these characteristics is a waste of resources. The

packet-switched network was first designed to fulfill the requirements of burst traffic presented by data transmission. In the packet-switched networks, the information is split up by the terminal into blocks of moderate size, called packets. These packets are autonomous, i.e., they are capable of moving on the network thanks to a header that contains the source and destination addresses. The packet is sent to the first node in this communication network. The nodes are referred to as routers. When the router receives the packet, it examines the header and forwards the packet to the next appropriate router. This technique of inspection and retransmission is called store-and-forward, and it is accomplished in all routers of the path until the packet reaches its destination, unless the packet is lost. After reaching the destination, the destination terminal strips off the header of the packet to obtain the actual data that was originated at the source. In this communication process, the terminal sends packets at its own rate, and the network multiplexes the packets from various origins in the same resources, to optimize their use. In this way several communications can share the same resources. The packet-switched network enables a better use of the transmission resource than circuit-switched network, in which the transmission resources are allocated without sharing. On the other hand, the multiplexing of different connections on the same resources causes delays and packet loss, which do not happen with circuit-switched network. (Kashihara, 2011)

Finally it must be noted that in packet-switched networks a distinction is made between two modes of operation: connection-oriented mode and connectionless mode.

- In connection-oriented mode, a path is established before any packets are sent; this path is called virtual circuit. There is a prior exchange of initial signaling packets to reserve resources and to establish the path. The connection-oriented mode is modeled after the telephone system. In order to talk to someone, one has to pick up the phone, dial the number, talk, and then hang up. Similarly, in connection-oriented mode, the user establishes a connection, uses the connection, and then releases the connection. The essential aspect of a connection is that it acts like a tube, the sender pushes

objects (packets) in at one end, and the receiver takes them out at the other end. In most cases the order is preserved so that the packets arrive in the order they were sent.

- In connectionless mode, each packet is treated independently, with no reference to packets that have gone before and the routing decisions are taken at each node. The connectionless mode is modeled after the postal system. Each message (packet) carries the full destination address, and each one is routed through the system independently of all the others. Normally, when two packets are sent to the same destination, the first one sent will be the first one to arrive. However, it is possible that the first one sent can be delayed so that the second one arrives first.

The connectionless mode has been popularized mainly by Internet protocol. The IP networks have progressed to the point that it is now possible to support voice and multimedia applications, but does not guarantee quality of service, because are based on "best effort" services. (Kashihara, 2011)

2.2 VoIP Concept

VoIP is the routing of voice traffic over the Internet or any other IP-based network. Using the Internet's packet-switching capabilities, VoIP technology has been implemented to provide telephone services. Figure 2.2 illustrates a typical VoIP function procedure.

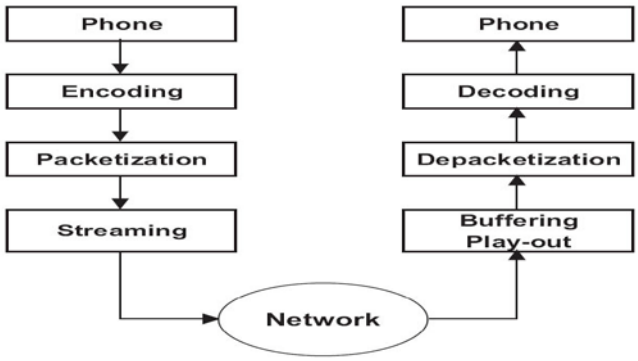


Figure 2.2 a typical VoIP function procedure.

At the sending end, the original voice signal is sampled and encoded to a constant bit rate digital stream. The digital stream can then be easily compressed. This digitized and compressed data is then encapsulated into packets of equal sizes for easy transmission over the Internet. Along with the compressed voice data, these packets contain information about the packet's origin, the intended destination, and a timestamp that allows the packet stream to be reconstructed in the correct order. These packets flow over a general-purpose packet-switched network, instead of traditional dedicated, circuit-switched voice transmission lines. At the receiving end, the continuous stream of packets are depacketized and converted back into the analog signal so that it can be detected by the human ear. In general, this means voice information is sent in digital form in discrete packets rather than using the traditional circuit-committed protocols of the Public Switched Telephone Network (PSTN). In addition to IP, VoIP uses the Real-Time Transport Protocol (RTP) to help ensure that packets get delivered in a timely way. (He, 2007)

2.2.1 Architecture

VoIP network technologies can be classified according to the IP network in which the speech data is transmitted.

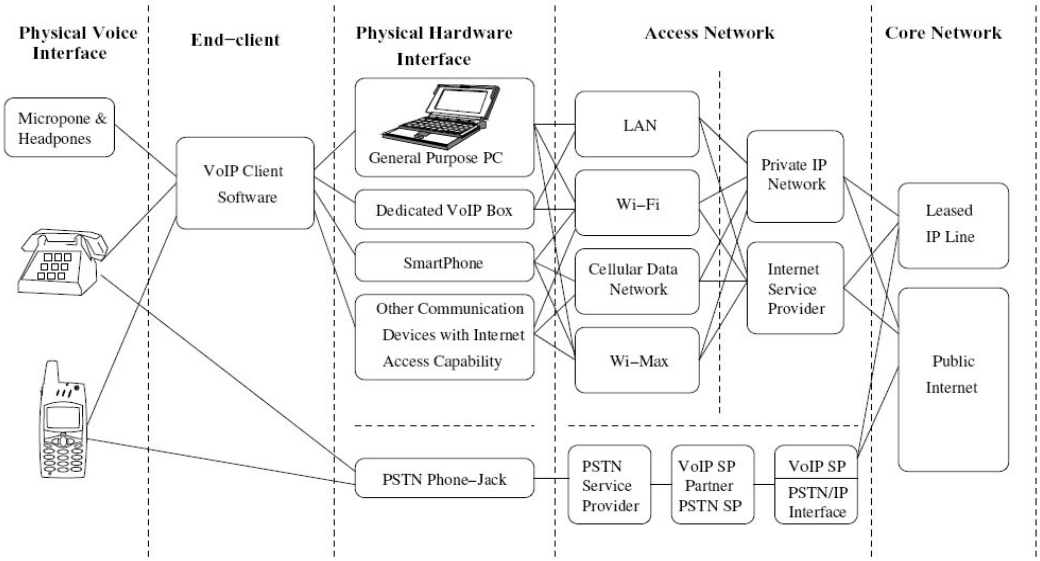


Figure 2.3 VoIP system infrastructure depicting different physical interfaces, hardware used and networks traversed. (Sat, 2010)

They can run on privately owned IP networks, such as enterprise networks and leased lines, or the public Internet, which can be accessed via Internet service providers (ISPs). The physical interfaces for VoIP clients include general purpose hardware (e.g. PCs), PDAs, smart-phones (e.g. iPhone), dedicated VoIP boxes that usually come with a subscription to a VoIP service, and any communication devices with access to the Internet. (Sat, 2010)

VoIP nodes can utilize a variety of hardware interfaces, such as laptop computers, PDAs, smart phones, and dedicated VoIP handsets (see Figure 2.3). Independent of the interface, there is a software client that communicates with a counterpart client over a network with a best-effort end-to-end service. The clients support an interactive conversation by processing speech packets and by shielding users from network imperfections. The popularity of these clients that run on general purpose computers has grown in recent years, even more than VoIP services that run on dedicated boxes.

Furthermore, third-party vendors have developed phone-like handsets (like Skype phone and Google phone) that can connect directly to Wi-Fi networks without a PC. These devices improve the ease of use and make the technology more transparent to users. (Sat, 2010)

2.2.2 VoIP versus PSTN

In this section public switched telephone network (PSTN) and VoIP communication systems are compared.

2.2.2.1 VoIP Advantages

A large number of benefits can be attributed to the substantial rise in growth of VoIP, most of which standard PSTNs do not offer. Some of these benefits can be summarized as follows:

- Use of Existing Data Infrastructures: When existing IP data infrastructures are available; these IP networks can support VoIP traffic as well as data traffic. Benefits include a cost savings, because only one wired network is needed versus two networks for standard PSTNs. Also, cost savings are increased due to the decrease of administration need for one IP network and system versus two separate networks and systems to administrate with standard voice. (Sasidar, 2004)

- Application Rich Environment: VoIP from its inception has offered the ability to carry not only voice, but multimedia and multi-service communications. A number of different applications include video presence, instant messaging, multiparty conferencing (multiple media types), interactive TV, gaming, video on demand, and text chat. Most of these services standard PSTNs cannot offer and the services they do, PSTNs cannot competitively compete with VoIP.

- Technological Malleability: VoIP, as a technology, is malleable into a countless numbers of different setup scenarios and cases where configurations are extremely different. Where VoIP becomes a benefit is that the configurations are software based and flexible to each individual setup configurations and can be changed simply. This makes administration of VoIP much easier and much less dependent on hard-wire changes.

- Market Divergence: VoIP has diverged into many different diverse markets and still continues to expand in different directions where this technology can present itself as a benefit. Such different markets include wired telephony, Wi-Fi telephony, Wi-Fi hand-held devises, Wi-Fi “dual mode” mobile phones, Mobile VoIP (MVoIP), voice over instant messaging (VoIM), text over Internet protocol (ToIP) and radio over Internet protocol (RoIP). (Broniecki, 2009)

2.2.2.2 Comparisons

Table 2.1 briefly presents a number of comparisons and facts about typical PSTN and VoIP systems.

Table 2.1 Comparisons of PSTN and VoIP systems (Anonymous, 2010)

	PSTN	**VoIP**
1	Very Reliable	Not as reliable as PSTN
2	Excellent audio quality	Quality depends on internet bandwidth availability
3	Dedicated lines required from Telco Provider	All calls can be transferred over the internet connection
4	Extra features available at extra cost from Telco	Extra features available for free or at a very low cost
5	Tele workers solutions are very costly and require additional dedicated lines	Remote extensions are standard features with VoIP PBX
6	To maintain a PSTN PBX can be very costly and require specific technical expertise	As long as you can use a computer, you can maintain a VoIP network
7	Upgrading and extending a PSTN PBX is costly and a complex process, and requires specific expertise	Upgrading and extending a VoIP PBX requires only an internet connection and software upgrades
8	Because of a limited number of telecoms in every country, least cost routing is very limited	Thousands of VoIP providers in different countries to choose from, thus providing a cheap solution
9	No free calls possible	Possible to call other VoIP users using the same gateway for free
10	Installation of PSTN PBX requires extra wiring, expensive proprietary hardware and telecom expertise	VoIP PBX uses same computer networking cables, thus do not require additional hardware or specific skills
11	Typically hard wired, remains on during power cuts	Both PBX and phones required electrical power, therefore will be unavailable with loss of power
12	Suitable for emergency calls and location can usually be traced	Not suitable for emergency calls.. Location can be difficult to be traced
13	Little or no integration at all with computer applications such as email clients	Full integration with a number of computer applications, such as email clients
14	Only audio can be transferred during a call	Audio, video and data (such as presentations) can be transferred during a call

2.3 VoIP Network

In VoIP systems implementation and analysis, network is a very important factor that must be considered.

2.3.1 VoIP Protocols

Network engineers and developers use the Open Systems Interconnection basic reference model (OSI model) as a standardized way of talking about the components required to support all network transmissions. Figure 2.4 shows how the model is divided into seven layers and how those layers identify the protocols used to transmit VoIP.

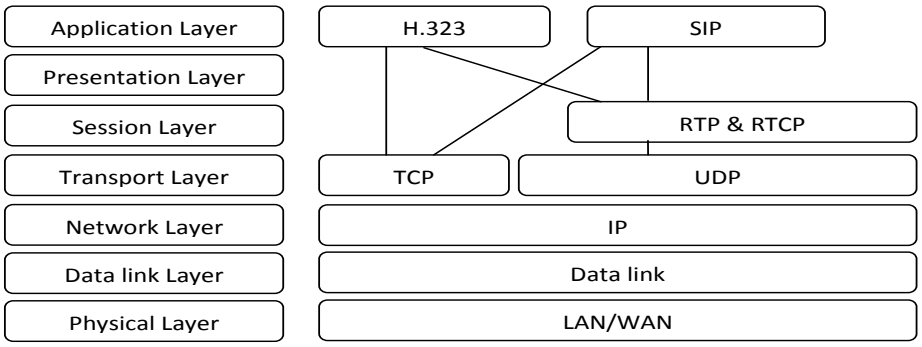


Figure 2.4 VoIP protocol stack

In the following sections, these VoIP specific protocols will be introduced in detail as well as the fundamental transport protocols on which they are built.

2.3.1.1 Internet Protocol (IP)

IP stands for Internet Protocol. It is responsible for the delivery of packets (or data grams) between host computers. IP operates at the network layer (see Figure 2.3). It is a connectionless protocol, that is, it does not establish a virtual connection through a network before starting transmission. IP makes no guarantees concerning reliability, flow control, error detection or error correction. The result is that data grams could arrive at the destination computer out of sequence, with errors, or not

even arrive at all. Nevertheless, IP succeeds in making the network transparent to the upper layers involved in voice transmission through an IP based network. Transport layer protocols use IP services to provide various levels of service guarantees. By definition, any Voice over IP transmission must use IP. As a real time application, voice transmission requires guaranteed connections with consistent delay characteristics. Many characteristics of IP, however, do not make it well-suited for voice transmission. Higher layer protocols address these issues. The focus of this thesis is on the higher layer protocols of VoIP; the study of lower level network protocols is not within the scope of this work. (Long, 2001)

In its most basic form, the IP header comprises 20 bytes. There are optional fields which can be appended to the basic header, but these offer additional capabilities which are not necessary for VoIP transmission. (Thomas, 2009)

2.3.1.2 Transitions Protocols (TCP & UDP)

Generally, there are two protocols available at the transport layer when transmitting information through an IP network. These are TCP (Transmission Control Protocol) and UDP (User Data- gram Protocol). Both protocols are associated with unique port numbers (for example, the HTTP application is usually associated with port 80).

TCP stands for Transmission Control Protocol. It is a connection-oriented protocol that is responsible for reliable communication between two end processes. TCP enables two hosts to establish a connection and exchange data streams and guarantees that packets will be delivered in the same order in which they were sent. TCP was designed to dynamically adapt to properties of the internetwork and to be robust in case of failures. The sending and receiving TCP entities exchange data in the form of segments. A segment consists of a fixed 20-byte header (plus an optional part) followed by zero or more data bytes. The source port and destination port specify the end points of the connection. The sequence number field identifies the first byte of data in this segment and the acknowledgement number contains the value of the next sequence number the sender of the segment is expecting to receive.

The TCP header length field tells the TCP header length in 32-bit words. There are six 1-bit flags. The window size field tells how many bytes may be sent starting at the byte acknowledged. The checksum field checks a sum of the bytes in the header. With this information in the header, TCP provides reliable transmission between the two end points. (He, 2007)

UDP stands for User Datagram Protocol which is connectionless and unreliable. It has minimal overhead. Each packet on the network is composed of a small header and user data, and is called a UDP datagram. A datagram can be sent at any time without prior advertising, negotiation or preparation. UDP routes data to its correct destination port, but does not try to perform any sequencing, or to ensure data reliability in common with IP. A UDP segment consists of an 8-byte header followed by the data. The two ports serve the same function as they do in TCP: to identify the end points within the source and destination machines. The UDP length field includes the 8-byte header and the data. With only this information in the header, UDP provides unreliable transmission between the two end points.

Voice is a real-time application, and mechanisms must be in place to ensure that information is received in the correct sequence, reliably and with predictable delay characteristics. Although TCP would address these requirements to a certain extent, there are some functions which are reserved for the layer above TCP. Therefore, some of TCP's functions must be reworked in some way to be more specific for VoIP as they aren't really used unmodified at the TCP layer. Also, the extra overhead of TCP and the possibility and high likelihood of increased latency make it unsuitable for real time applications. Therefore, for the transport layer, TCP is not used, and the alternative protocol, UDP, is commonly used. (He, 2007)

2.3.1.3 Media Protocols (RTP & RTCP)

RTP is an IETF protocol (Schilzrinne, 2003) designed to support the real-time transfer of data between two or more members of a multimedia session. Riding above the UDP transport layer, RTP focuses on providing timely media delivery rather than reliable services to session participants. VoIP calls in an H.323 system

pass packetized bit streams from the codec down the RTP-UDP-IP stack. A typical link level packet format is shown in Figure 2.5.

x bytes	**20 bytes**	**8 bytes**	**12 bytes**	**x bytes**
Link Header	IP Header	UDP Header	RTP Header	Voice Payload

Figure 2.5 A typical link level packet format

RTP header values include data source, timestamp, sequence, and payload identification fields to assist in the recovery of media packet data. Sequence and time information facilitate endpoint activities to defeat negative network effects to packet delivery. Buffers allow sequence and time data to assist during reconstruction of original packet order and a reduction in delay variation for final transmission.

RTP

V	P	X	CC	M	Payload Type	Sequence Number
Time Stamp						
Synchronization Control Source						

UDP

Source Port	Destination Port
Length	Checksum

IP

V	HL	TOS	Total Length	
Identification			Flags	Fragment Offset
TTL	Protocol		Header Checksum	
Source IP Address				
Destination IP Address				
Options				

Figure 2.6 Detailed views of the common VoIP header fields

RTP header values also facilitate network statistical analysis by tracking the distribution and rate of packet loss. RTP does not provide any form of error detection or control. Figure 2.6 provides a detailed view of the common VoIP header fields.

RTP Control Protocol (RTCP) is a companion protocol defined within RFC 3550. RTCP manages quality of service, identification, session scaling, and session control of the RTP stream. RTCP packets are issued periodically, using a separate port number, to session members in a multicast fashion. (Stallings, 2007)

2.3.2 VoIP Call-Signaling Protocol Stacks

Call-signaling protocols are central and integral to the process of discovering VoIP devices and the negotiation of communication flows between those devices. Media, in the VoIP form of audio and video, must flow between two devices. Before this flow can occur, sets of protocols must be utilized to locate each device, and to negotiate the transactions which those two devices will carry media flows between each other.

As previously stated, the most popular call-signaling protocols are H.323 and SIP. While H.323 and SIP are not the only call-signaling protocols, they certainly can be considered the most used non-proprietary call-signaling protocols.

2.3.2.1 H.323

H.323 is an International Telecommunication Union Telecommunication Standardization Sector (ITU-T) specification for transmitting audio, video, and data across an Internet Protocol (IP) network, including the Internet. (ITU-T, 1999)

2.3.2.1.1 H.323 Elements. Figure 2.7 illustrates the elements of an H.323 system. These elements include terminals, gateways, gatekeepers, and multipoint control units (MCU).

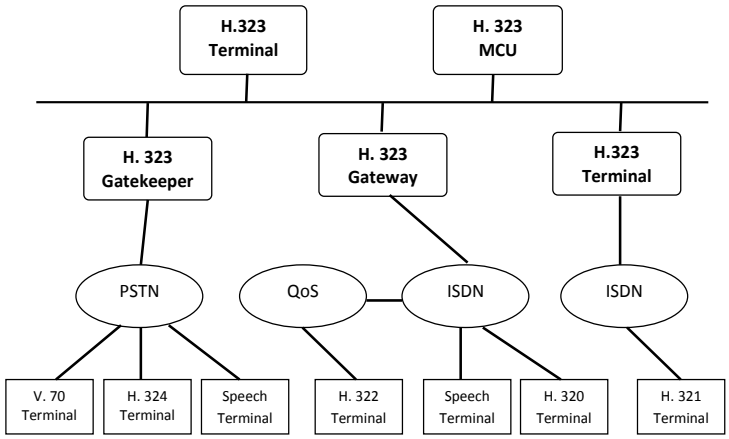


Figure 2.7 the elements of an H.323 system.

H.323 Terminal: H.323 terminal is an endpoint where H.323 data streams and signaling originate and terminate. It may be an IP telephone or a multimedia PC with an H.323 compliant stack that provides real-time two way communications.

Gateway (GW): A Gateway is an optional component in an H.323-enabled network. An H.323 Gateway is an H.323 endpoint that provides translation between terminals belonging to networks with different protocol stacks, enabling the endpoints to communicate.

Gatekeeper (GK): A Gatekeeper is a very useful but optional component of an H.323-enabled network. The gate-keeper provides several services such as address translation and network access control for the network resources to all endpoints in its zone. Also, it can provide other services such as band-width management, accounting and dial plans for scalability. (Peters, 2000)

Multipoint Control Unit (MCU): MCU is also an optional component of an H.323-enabled network and its basic function is to maintain all the audio, video data and control streams between all the participants in the conference. It is typically used for multiparty video conferences. The main components of an H.323 MCU are a mandatory Multipoint Controller (MC) and an optional Multipoint Processor (MP).

2.3.2.1.2 H.323 protocol suite. The protocols specified by H.323 are illustrated below in Figure 2.8.

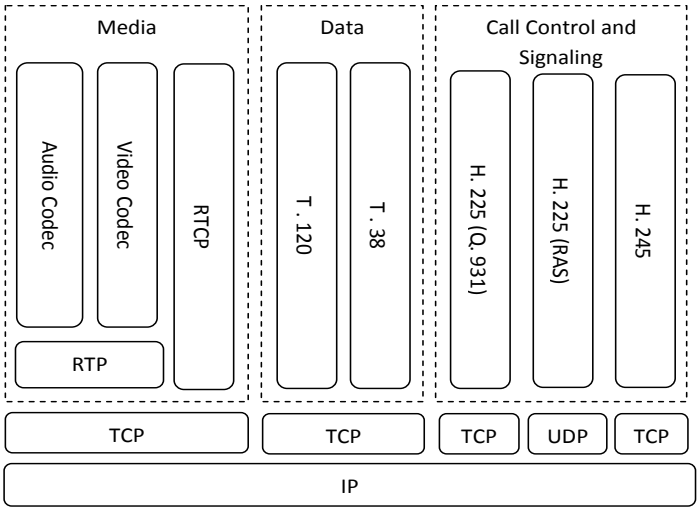


Figure 2.8 H.323 protocol suite.

Audio CODECs : An audio CODEC encodes the audio signal from the microphone for trans-mission on the transmitting H.323 terminal and decodes the received audio code that is sent to the speaker on the receiving H.323 terminal. Since audio is the basic service provided by the H.323 standard, all H.323 terminals must have at least one audio CODEC support such as ITU-T G.711, G.723.1 or G.729.

Video CODECs : A video CODEC encodes video from the camera for transmission and decodes the received video code that is sent to the video display. Since video is an optional service provided by H.323, the support of video CODECs is optional as well.

H.225 registration, admission, and status (RAS). Registration, admission, and status (RAS) is the protocol between endpoints (terminals and gateways) and gatekeepers. The RAS is used to perform registration, admission control, bandwidth changes, status, and disengage procedures between endpoints and gatekeepers.

Other Call Control Units: H.225 (Q.931) and H.245 provides call control, capability exchange, messaging, and signaling of commands for proper operation of the terminal.

Data Channel: Supports applications such as database access, file transfer, and audio graphics conferencing (the capability to modify a common image over multiple users' computers simultaneously), as specified in Recommendation T.120.

2.3.2.1.3 H.323 Call Control Signaling. In H.323 networks, call control procedures are based on International Telecommunication Union (ITU) Recommendation H.225, which specifies the use and support of Q.931 signaling messages. A reliable call control channel is created across an IP network on TCP port 1720. This port initiates the Q.931 call control messages between two endpoints for the purpose of connecting, maintaining, and disconnecting calls.

The following Q.931 and Q.932 messages are the most commonly used signaling messages in H.323 networks:

- Setup: A forward message sent by the calling H.323 entity in an attempt to establish connection to the called H.323 entity. This message is sent on the well-known H.225 TCP port 1720.

- Call Proceeding: A backward message sent from the called entity to the calling entity to advise that call establishment procedures were initiated.

- Alerting: A backward message sent from the called entity to advise that called party ringing was initiated.

- Connect: A backward message sent from the called entity to the calling entity indicating that the called party answered the call. The connect message can contain the transport UDP/IP address for H.245 control signaling.
- Release Complete**:** Sent by the endpoint initiating the disconnect, which indicates that the call is being released. You can send this message only if the call signaling channel is open or active.

- Facility: A Q.932 message used to request or acknowledge supplementary services. It also is used to indicate whether a call should be directed or should go through a gatekeeper.

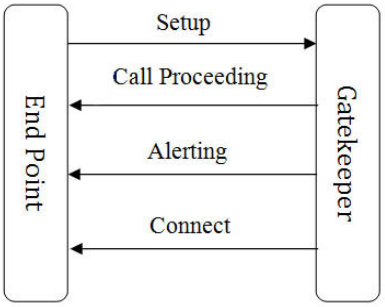


Figure 2.9 illustrates the signaling Messages for H.323 call setup.

Call signaling channel can rout in an H.323 network in two ways: through Direct Endpoint Call Signaling and Gatekeeper Routed Call Signaling. In the Direct Endpoint Call Signaling method, call signaling messages are sent directly between the two endpoints, as illustrated in Figure 2.10 (Peters, 2000)

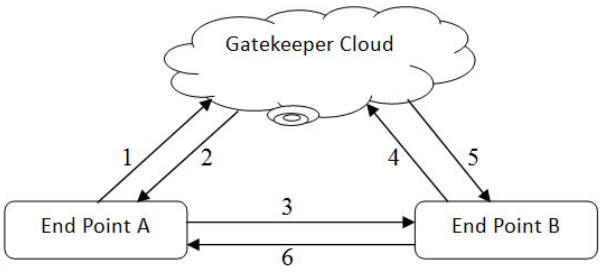


Figure 2.10 Direct Endpoint Call Signaling (1.ARQ, 2.ACF/ARJ, 3.Setup, 4.ARQ, 5.ACF/ARJ, 6.Connect)

Where ARQ is “An attempt by an endpoint to initiate a call”, ACF is “An authorization by the gatekeeper to admit the call” and ARJ is “Denies the endpoint's request to gain access to the network for this particular call”.

In the Gate Keeper Routed Call Signaling (GKRCS) method, call signaling messages between the endpoints are routed through the gatekeeper, as illustrated in Figure 2.11.

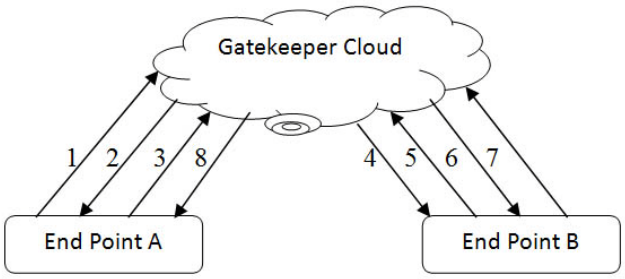


Figure 2.11 Gatekeeper Routed Call Signaling (1.ARQ, 2.ACF/ARJ, 3.Setup, 4.Setup, 5.ARQ, 6.ACF/ARJ, 7.Connect, 8.Connect)

In Figures 2.10 and 2.11, the Setup and Connect messages are call signaling channel messages, whereas the remaining messages are RAS channel messages.

2.3.2.2 SIP

Session Initiation Protocol (SIP) is a call signaling protocol, like H.323, and is used for creating, modifying and terminating VoIP sessions with VoIP devices or participants. SIP is backed by the IETF (Internet Engineering Task Force). SIP as a protocol was developed specifically as a application-layer control protocol, using a textual encoding schema for its data, which is based on HTTP and MIME. SIP was constructed, from its inception, as a packet based protocol and was developed for explicit use in IP packet based environments. SIP is transport-independent which makes it well suited for SIP message reading and system administrating, but not for large multimedia transport. (IETF, 2002)

Like Internet phone calls and multimedia conferences, SIP is used to create those two or more party and multicast sessions. Because SIP is transport independent, it can run on UDP, TCP and SCTP (Stream Control Transmission Protocol).

2.3.2.2.1 SIP Systems Structure. The composition of SIP is structured to define and use the following components:

- UAC (User Agent Client): A UAC is the client in the terminal which initiates the SIP signaling.

- UAS (User Agent Server): A UAS is a server in the terminal which responds to a SIP signaling from the UAC.

- UA (User Agent): A UA is a terminal in the SIP network which contains UACs and/or UASs. This is typically endpoints such as SIP telephones or a gateway to other networks. The endpoints can make and receive communication calls and be a device such as an IP phone or a soft phone.

- Proxy Server: These devices receive connection requests from the UA and transfer those requests on to an additional proxy server in the event that the additional device is not in the original server's administration.

- Redirect Servers: These devices receive connection requests and send those requests back to the requester, which includes the destination data instead of sending the requests to the calling party.

- Location Server: These devices receive registration requests from the UA and update the terminal database with those requests.

All SIP Server types (Proxy, Redirect and Location) are typically all available on the same physical device, which is again typically called a Proxy Server, that is responsible for client database maintenance, the establishment of connections, directing calls and maintenance and termination. The Proxy server also implements provider call routing policies and allows features to users. (Mastalir,2005)

SIP also provides a registration function which allows uploading of current locations by users for the use by a Proxy Server(s). A UAS that handles a "Register" function is granted a special qualification of Registrar.

There are two different configurations in which a Proxy Server can exist as, a Stateful Proxy or as a Stateless Proxy. The Stateless Proxy configurations are comparability simple and faster than Stateful Proxies. They are used as message translators, routers and load balancers for SIP messages. Stateful Proxy Servers do not handle retransmitted messages and do perform advanced routing administration.

In a typical configuration, the SIP Proxy Server is used by all user agents within the SIP domain. When user agents (devices) are registered through the Register, the Proxy server can determine user's location. Along with registering user agents, the Register requests the information and location from those agents and all information is stored in a database being the Location Server. The Proxy Server utilizes the database after a user agent has sent an invitation. The MIME type formats which SIP adheres to for addresses can be seen as follows: sip:user@ip:port. A Specific example can be seen as follows: sip:xyz@1.2.3.4:5060. After the Proxy has received an invitation, the Proxy Server will look up in the address database and send an invitation response to the address in the database. Again, because these functions are so closely intertwined, the Proxy Server, Registrar and Location Servers are usually incorporated into the same physical devise. (Sasidar, 2004)

2.3.2.2.2 SIP Call Flow Sequence. A sample call sequence is illustrated in Figure 2.12. Messages are divided into either request or response categories. Response messages also split into a numbered class system. Examples of the request and response message format are shown in Table 2.2 This fairly simple structure has made SIP an attractive alternative to the more complex H.323. (Maka, 2007)

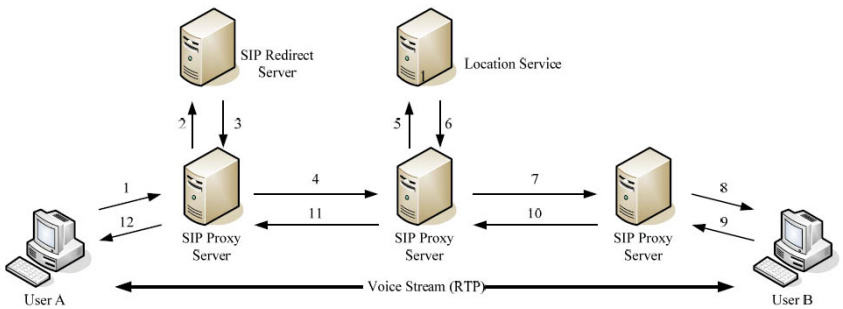


Figure 2.12 SIP Call Sequence: User A initiates a voice call to User B

Table 2.2 SIP Request and Response Formats

SIP Request	Purpose
INVITE	Invite a user to a call
ACK	Acknowledge
OPTIONS	Get server capabilities
BYE	Close or deny call
CANCEL	Terminate action
REGISTER	User Location Report
INFO	Mid session signal

Response Classes	Purpose
1XX	Informational
2XX	Successful
3XX	Redirect
4XX	Client Error
5XX	Server Error
6XX	Global Failures

2.3.2.2.3 SIP protocol stack format. With SIP, there can be three principal areas between the transport layer (IP) and the Application layer (User Interface and Apps) of the OSI model.

Media Control/Transport: This area which encapsulates the protocols of RTP, RTCP, H.261/MPEG and RSVP, call signaling and channel usage and capabilities with the connections of endpoints and other elements. This area principally addresses RTP and RTCP and the underlying protocol CODECS for audio and video conferencing. (Broniecki, 2009)

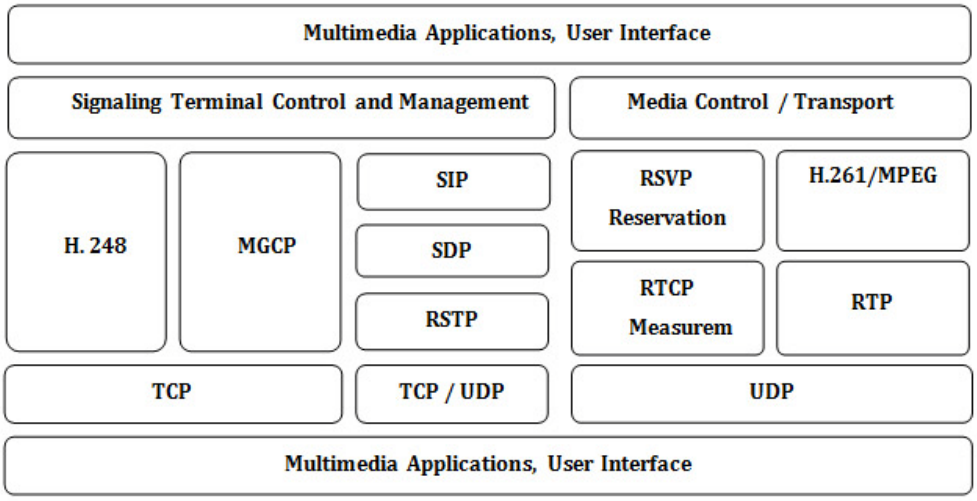


Figure 2.13 SIP Protocol Stack (Broniecki, 2009)

Signaling, Terminal Control and Management: Area that encompasses applications through the signaling function. This includes aspects such as H.248, MGCP/Megaco and principally SIPS, SDP and RSTP.

2.4 VoIP Voice Codecs

In order to be transmitted across computer networks voice (the same is true in principle for video) has to be converted from its analogue form into a digital format. This digitizing of data is done by discrediting the signal in time (sampling) and then discrediting the signal in amplitude.

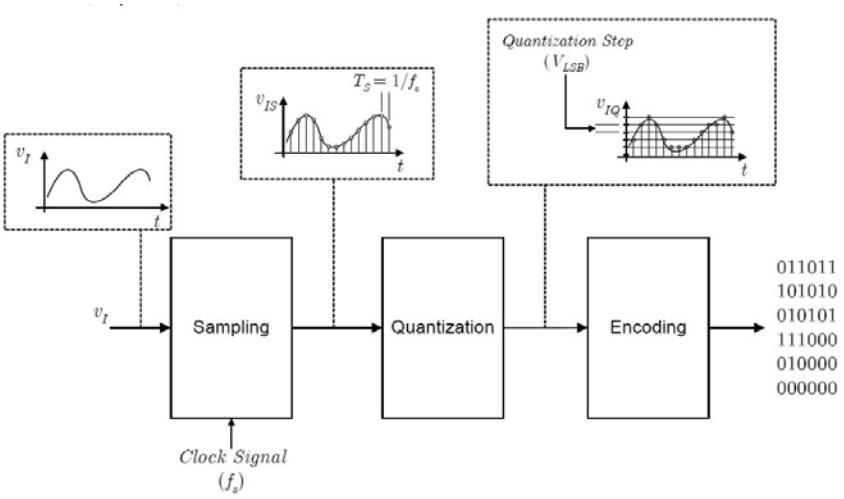


Figure 2.14 Analog to Digital conversation processing steps

The systems that perform this conversion process (in both directions) are called codecs. Typical examples of codes are G.711, G.723.1, G.729, G.726, G.723, and G.728.

Perhaps the one of most important factors with regard to voice signal quality in a VoIP environment is the voice codec (coder/decoder) implemented in VoIP gateways/routers, IP telephones, and other VoIP terminals.

2.4.1 Codec Description

Codecs digitize and packetize voice signals prior to their transmission across an IP network. Some codecs also compressing the voice signal to preserve network

bandwidth. Voice codecs are implemented in software and/or hardware and are often rated according to the following parameters (Madisetti & Williams, 1998):

- Bit rate is a measure of the compression achieved by the codec.
- Delay is a measure of the amount of time a Codec requires to process incoming speech signals. This processing delay is a portion of the overall end-to-end delay experienced by a voice packet.
- Complexity is an indication of a codecs cost and processing power.
- Quality is a measure of how speech ultimately sounds to a listener. MOS is a system of grading the voice quality of telephone connections. With MOS, a wide range of listeners judge the quality of a voice sample on a scale of one (bad) to five (excellent). The scores are averaged to provide the MOS for the codec.

Clearly, tradeoffs must be considered when deciding which codec(s) to use in a given VoIP network or device. For example, in situations where bandwidth is at a premium, low bit rate codecs may be preferred at the expense of some signal quality.

In other situations, voice quality must be preserved resulting in higher complexity, cost, and bandwidth requirements.
For telephony applications, there are three categories of codecs (Collins, 2001):

- **Waveform codecs** are the most common type and are used ubiquitously in most PSTNs. These codecs seek to reproduce the analog signal waveform at the receiving end of the call and generally introduce the least amount of distortion and noise. They also require the highest amount of bandwidth. ITU-T's G.711 is the most common waveform codec.

- **Vocoders** do not seek to reproduce the analog signal waveform, but instead seek to reproduce the subjective sound of the voice signal. Vocoders are targeted strictly at voice signals, use less bits to encode the voice signal (thus, requiring less bandwidth), and are generally believed marginally suitable for

telephony applications (although they have been and are used in some VoIP environments).

- **Hybrid codecs** are the most commonly used codecs in VoIP networks. Hybrid codecs meld the best characteristics of both waveform codecs and vocoders and also operate at very low bit rates.

Figure 2.15 below shows the three types of codecs with respect to bit-rate and Speech quality.

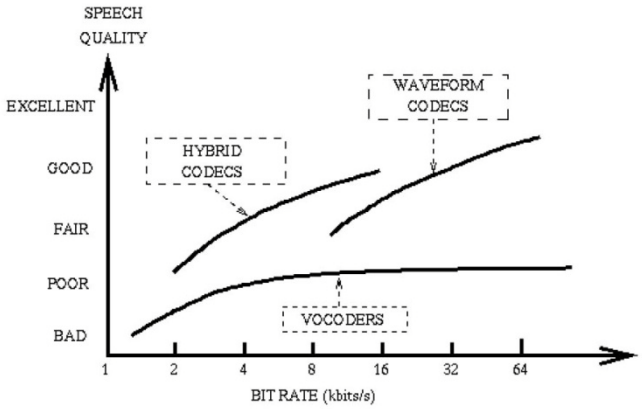


Figure 2.15 Codec types

Some techniques used to compress the digital voice data (Anonymous, 2011):

VAD – Voice Activity Detection: In IP Telephony, both the conversations as well as the silence in between the conversations are digitized. So, we have both – packets containing voice as well as packets containing silence. Using VAD, packets of silence can be discarded after their duration is appropriately marked. So, the total number of packets transmitted after compression is lesser (generally around 30% lesser).

CNG – Comfort Noise Generation: This is not a compression technique, but when voice is compressed using VAD, the awkward silence between the speech might be interpreted as lost connection and hence white noise is generated locally at both ends

using CNG. This makes the call appear connected to both the parties during silence, as some background noise is audible during that duration.

CELP – Code Excited Linear Prediction: In this method, various human sounds are mathematically modeled and a code book of all possible sounds is produced. So, instead of sending the actual sound packets across, only their codes are sent across. This is a very simplistic explanation, and a lot more techniques are involved.

In some compression techniques, the headers can be compressed separately like the payload compression which provides additional bandwidth efficiency while transmission. The following table lists the various codecs used in voice over IP:

Table 2.3 VoIP voice codecs list

Codec Standard	Bit Rate (Kbps)	Codec Sample Size (Bytes)	Payload Size (Bytes)	Packet Per Second	*MOS
G.711	64	80 Bytes	160	50	4.1
G.729	8	10	20	50	3.92
G.723.1	6.3/5.3	24/20	24/20	33.3	3.9/3.8
G.726	32/24	20/15	80/60	50	3.85
G.728	16	10	80	33.3	3.61
G.722	64	10	160	50	4.13
iLBC 20/30	15.2/13.33	38/50	38/50	50/33.3	NA
GSM-EFR	12.2	NA	NA	NA	4.4

* (Mean Opinion Score) is described in section 4.2.3

2.4.2 Codec Selection Criteria

When designing a VoIP network, designer has to take into account the following criteria for codec selection:

- Endpoint capabilities.
- VoIP equipment capabilities.
- Bandwidth allocated for VoIP traffic.
- Amount of VoIP traffic projected for future.
- Codec supported by potential customer and supplier VoIP carriers.

2.5 VoIP Security

The fast deployment of VoIP solutions not only offers new possibilities and opportunities but also introduces new risks. The fundamental technology changes for voice communication introduce new threats and new challenges for security specialists and network administrators. (Sme, 2010)

Eavesdropping on conversations in the network by intercepting a VoIP connection is only one example of the new security threats. However, security measures in general to not differ very much from networks without VoIP. Whether the goal of the attacker is to gain information, steal resources or to disrupt business processes, the used approaches and tools are pretty much the same.

But there are certainly security issues to be addressed that result from the specific VoIP technology implemented. The SIP standard for example does include functions to enhance media security (encryption), message exchange security and authentication. Also in H.323 there are functions and protocols that are designed to provide better levels of authentication, privacy and integrity.

In situations where higher levels of privacy and security are needed, technologies like firewalls, authentication systems and VPN technology can be implemented.

Some of the “classic” attack patterns or techniques like DDoS attacks, will probably also be directed at VoIP servers and gateways. Firewall systems need to be configured or even upgraded to be able to prevent damage to the network and its components.

Although not a security problem, special attention has to be given to communication issues and problems. In particular, those that have to do with techniques (mainly network address translation) used in many firewall solutions and IADs (Internet Access Devices) - such as access routers.

NAT (Network Address Translation) Techniques allow more than one system (for example a LAN) to be connect to the Internet using a single (often even dynamically

assigned) IP address. The address translation device converts the outgoing IP address of each LAN device into its single Internet address and vice versa. Because of serious limitations related to incoming connections that cannot be simply directed to the individual systems in the internal network, these network address translation devices need special software that work at an application/protocol level (for example SIP signaling proxies and H.323 proxies) to overcome these issues.

Other issues or problems that have to be taken into account when using systems like firewalls or VPN gateways in VoIP solutions have to do with the additional latency that these systems would cause. This means that a lot of attention has to be given to the overall network design and the selection of network and security devices that are in the communication path between VoIP systems to keep latency at a minimum.

CHAPTER THREE

VoIP VOICE QUALITY

As illustrated in previous sections of this thesis, in VoIP systems codecs are first used to change analog speech voice signals to digital voice packets; these are then sent via packet networks to listener hosts. At this processing step, there are some criteria that affect voice quality. These criteria are described in this section.

Before exploring VoIP-specific distortion issues and how they are dealt with, a few basic voice qualities concepts should be introduced.

3.1 General Telephony Impairments

As described in the previous section, VoIP networks almost always interface with some aspect of the public switched telephone network (PSTN). This means that most PSTN impairments can impact voice and conversation quality on interconnected VoIP networks. For example (ITU-T, 1993):

- Signal level is arguably the most important factor affecting perceived voice quality. Clearly, if signal levels are too low, users cannot understand what is said, and if levels are too high, clipping (distortion) can occur.

- Circuit noise and background noise have many sources from both the analog and digital portions of a telephony network. Since much of this noise is outside the voice band, it can cause some problems for VoIP vocoders if not eliminated via adaptive noise filters or other techniques. (Bellemy, 2000)

- Side tone is in fact a form of intentional echo that occurs at the telephone set. It is designed into telephone sets so that users can regulate their own voice levels and receive the necessary feedback that the circuit over which they are speaking is still “alive”. A similar phenomenon is addressed in VoIP

networks in which voice activity detectors (silence suppressors) are used. In this case, artificial background noise is actually injected into the voice circuit during silent periods between speech utterances to provide feedback that the circuit is still active.

- Attenuation and group delay distortion are impairments that are dependent on the frequency characteristics of a particular voice channel. Similar to analog circuit noise, attenuation and group delay distortion can cause unpredictable effects when coupled with low bit rate perceptual codecs used in VoIP. (Bellemy, 2000)

- Absolute delay is the time it takes for a voice signal to travel from talker to listener, and delay values typical of PSTNs (10s of milliseconds) have little effect on perceived voice quality if there is no echo or if echo is adequately controlled. However, due to signal processing, VoIP networks introduce unavoidable delays of 50 milliseconds and above which can expose echo (as described below) and affect conversational quality. (Hardman, 2003)

- Talker and listener echo can be problematic in traditional PSTNs and have been around for many years. In most situations, this echo is not perceptible because it returns to the talker/listener too quickly to be distinguished from regular speech. However, when larger end-to-end delays are introduced by VoIP processing, existing PSTN echo can become a real problem.

- Quantizing and non-linear distortion occurs in digital systems when an analog signal is encoded into a digital bit stream. The difference between the original analog signal and that which is recovered after quantizing is called quantizing distortion or quantizing noise. High quality PCM encoders used in PSTNs exhibit a predictable level of quantization noise and can, therefore, be dealt with in a relatively straightforward way. However, this assumption cannot be carried into the VoIP domain because voice-band codecs (vocoders) operate on a different premise and produce non-linear distortion. Thus, in VoIP

environments, quantization noise cannot always be measured or eliminated in the same way.

Because such PSTN impairments as described above can have an unpredictable effect on voice signals processed and transported across VoIP networks, aggressive noise reduction on circuits known to interface with VoIP networks should probably be employed.

3.2 Additive vs. Subtractive Distortion

All voice transmission systems are subject to the effects of both additive distortion (circuit noise, background noise, etc) and subtractive distortion (transient signal loss, severe attenuation, etc). For VoIP systems, however, these types of distortion are even more significant. Because perceptual codecs play such an important role in VoIP applications, noise added to the voice signal prior to encoding can have unpredictable effects depending on whether the noise has frequency components within the voice band or not and depending on the type of encoding used. In VoIP, traditional subtractive distortion such as excessive attenuation is now accompanied by the effects of packet loss where discrete portions of the encoded voice signal simply disappear. Again, due to the use of low-bit-rate codecs to preserve network bandwidth, this packet loss can be particularly disruptive. An equally interesting and related source of distortion is error concealment in which subtractive distortion such as packet loss is actually compensated for by *intentional* additive distortion in the form of predictive packet insertion. (Collins, 2001)

3.3 Non-linearity and Time-variance

Two of the primary differences between a PSTN or PSTN-like voice channel and a VoIP voice channel are the conditions of time variance and linearity. For the most part, a PSTN voice channel is LTI or Linear and Time-Invariant. (A voice channel is more or less linear if the voice waveform that enters the system is reproduced at the receiving end. A voice channel is time invariant if, once it is setup, its transmission characteristics normally do not change over time.) A VoIP voice channel, on the

other hand, is often non-linear and time-variant, a condition that makes noise reduction in a VoIP environment particularly challenging. For example, the end-to-end delay of the digital encoding/decoding scheme of a voice-over-IP channel can change during a single telephone call (time variance), resulting in changes in sound and conversational quality. Modern VoIP codecs encode and decode voice signals in non-linear ways because they strive primarily to preserve the subjective sound quality of a given voice signal rather than the objective audio waveform. (Hardman, 2003)

Depending on how these codecs are implemented (and depending on other network conditions such as packet loss), significant levels of distortion can be introduced to the voice signal.

3.4 Human Perception's Role

It is very difficult to separate the quantification of voice quality (that is, the evaluation or measurement of noise and distortion) from the subjective experience of the human talker and listener. Voice quality can really only be judged relative to the situation being assessed and the human experience of it (Moller, 2000). Voice circuit designers know that the physiology of the human ear and the psychology of human perception must be taken into account when designing voice processing and transmission systems, and therefore, when detecting and avoiding distortion. DSP (Digital Signal Processing) and voice processing design efforts increasingly concern themselves with only those parts of the voice signal likely to be perceived (Madisetti & Williams, 1998). This selective processing ultimately reduces transmission bandwidth requirements, benefiting those who must implement voice over IP systems in bandwidth limited situations. Therefore, noise reduction and avoidance in a VoIP environment often concerns itself only with the perceptually important aspects of noise and distortion.

Obviously, the human ear can detect only those auditory signals within a finite frequency and loudness range. However, cognitive aspects of human perception play

an important role in network design. For example, humans adapt to very brief auditory drop-outs without losing the meaning or content of a spoken phrase. Human listeners will perceive a particular voice sample as having worse quality if a burst of distortion occurs at the end of the sample as opposed to at the beginning of the sample. In addition, a listener's expectation and mood can also affect her/his assessment of voice quality. These and other aspects of human perception play a role in noise reduction in VoIP. (Hardman, 2003)

3.5 Listening Quality vs. Conversational Quality

As mentioned, two of the biggest challenges facing voice over IP systems are listening/sound quality and conversational quality. These two types of quality are related because end-users often do not make a conscious distinction between them. However, the distinction between the two should be preserved. Clearly, listening/sound quality is directly impacted by noise or other types of distortion. It is also clear that a distorted voice signal will negatively impact a telephone conversation. But several telephony phenomena, further exacerbated by VoIP processing, affect the character of voice conversations without really affecting sound quality at all. These phenomena include end-to-end and round-trip network delay, delay variance (jitter), and echo. Delay and echo will be covered, along with sound quality (clarity) in the next section. (Hardman, 2003)

3.6 VoIP Conversation

In a two-party conversation, each participant takes turns in speaking and listening, and both perceive silence duration (called *mutual silence* or *MS*) between turns when the current speaker ceases the floor and the listener takes over. A conversation, therefore, consists of alternating speech segments and silence periods. (Bosch, Oostdijk & Ruiter, 2004)

In a face-to-face setting, both participants have a common reality of the conversation: one speech segment is separated from another by a silence period that is identically perceived by both. However, when the same conversation is conducted

over the Internet, the participants' perception of the conversation is different due to delays, jitter, and losses incurred on the segments during their transmission. (Sat, B. & Wah, B.W., 2007)

During a VoIP session, a user does not have an absolute perception of MED (Mouth to Ear Delay) because the user does not know when the other person starts talking. However, by perceiving the indirect effects of MED, such as MS, the user can deduce the existence of MED. This asymmetry leads to a perception that each user is responding slowly to the other, and consequently results in degraded efficiency and perceptual quality. (Sat, 2010)

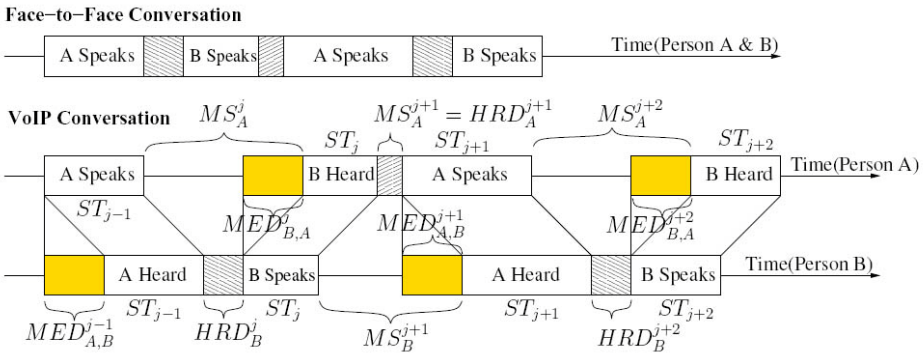


Figure 3.1 Conversational dynamics in a face-to-face and two-party VoIP setting (Sat, 2010)

Where ST is “Single-Talk speech segment”, MS is “Mutual Silence”, MED is “The Delay of The Mouth of a speaker to the ear of the Listener” and HRD is “Human Response Delay”

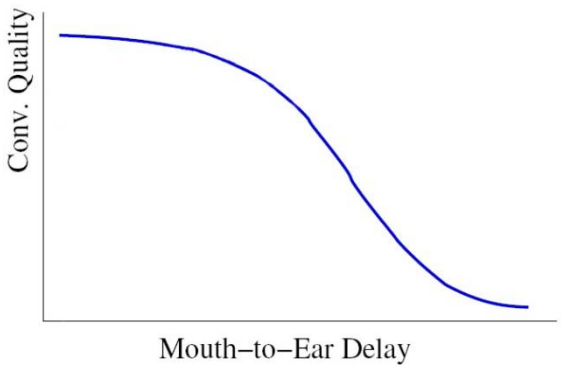


Figure 3.2 Affect of MED in VoIP Conversation Quality

3.7 Delay or Latency

As mentioned previously, End-to-end delay or MED is the time it takes a voice signal to travel from talker to listener. This voice signal delay is the additive result of VoIP/IP network processing and packet transport. Delay affects the quality of a conversation without affecting the actual sound of the voice signal – delay does not introduce noise or distortion into the voice channel.

Three types of delay are inherent in VoIP telephony networks: propagation delay, serialization delay, and handling delay. (Peters, 2000)

- **Propagation delay** is caused by the speed of light in fiber or copper-based networks. Light travels through a vacuum at a speed of 186,000 miles per second, and electrons travel through copper or fiber at approximately 125,000 miles per second. A fiber network stretching halfway around the world (13,000 miles) induces a one-way delay of about 70 milliseconds (70 ms). Although this delay is almost imperceptible to the human ear, propagation delays in conjunction with handling delays can cause noticeable speech degradation.

- **Handling delay**—also called processing delay—defines many different causes of delay (actual packetization, compression, and packet switching) and is caused by selected codec and devices that forward the frame through the network. For example, the Digital Signal Processor (DSP) generates a speech sample every 10 ms when using G.729. Two of these speech samples (both with 10 ms of delay) are then placed within one packet. The packet delay is, therefore, 20 ms. An initial look-ahead of 5 ms occurs when using G.729, giving an initial delay of 25 ms for the first speech frame. Vendors can decide how many speech samples they want to send in one packet. Because G.729 uses 10 ms speech samples, each increase in samples per frame raises the delay by 10 ms.

- **Serialization delay** is the amount of time it takes to actually place a bit or byte onto an interface.

The International Telecommunication Union Telecommunication Standardization Sector (ITU-T) G.114 recommendation specifies that for good voice quality, no more than 150 ms of one-way, end-to-end delay should occur.

When end-to-end delay reaches about 250 milliseconds, participants in a telephone conversation begin to notice its effects. For example, conversation seems "cold" and participants start to compensate. Between 300 to 500 milliseconds, normal conversation is difficult. End-to-end delay above 500 milliseconds can make normal conversations impossible. (Peters, 2000)

3.8 Jitter

Simply stated, jitter is the variations of packet inter arrival time. Jitter is one issue that exists only in packet-based networks. While in a packet voice environment, the sender is expected to reliably transmit voice packets at a regular interval (for example, send one frame every 20 ms). These voice packets can be delayed throughout the packet network and not arrive at that same regular interval at the receiving station (for example, they might not be received every 20 ms; see Figure 3.3). The difference between when the packet is expected and when it is actually received is jitter. (Anonymous, 2008)

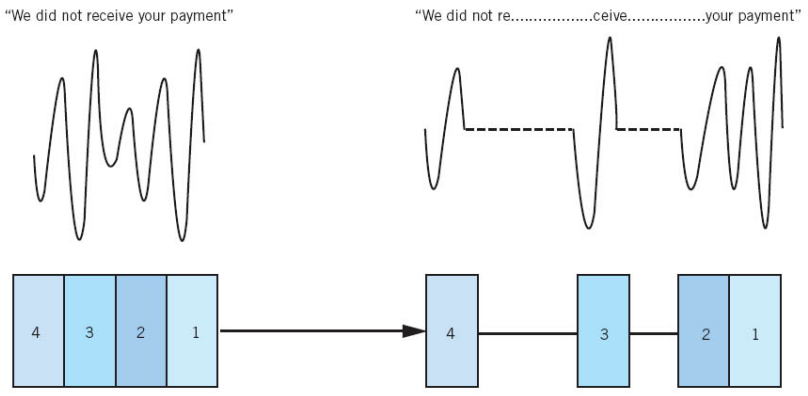


Figure 3.3 "Real-World" Packet movements with jitter

Packets leave their source in order, but they are very likely to use different paths as they travel over the network to their destination. This is especially true when packets must travel across a WAN. Because packets almost never follow the same route from point A to point B, each packet experiences different amounts of delay and may arrived out of order.

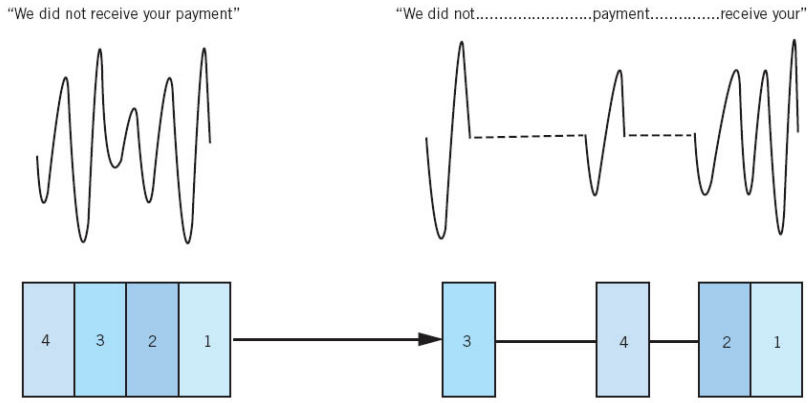


Figure 3.4 Delays cause jitter and out-of-order arrival these problems.

A jitter buffer is a shared data area where voice packets can be collected, stored, and sent to the voice processor in evenly spaced intervals. The jitter buffer, which is located at the receiving end of the voice connection, intentionally delays the arriving packets so that the end user experiences a clear connection with very little sound distortion.

There are two kinds of jitter buffers, static and dynamic. A static jitter buffer is hardware-based and is configured by the manufacturer. A dynamic jitter buffer is software-based and can be configured by the network administrator to adapt to changes in the network's delay.

3.9 Packet Loss

IP, by its very nature, is an unreliable networking protocol. In its most basic (and ubiquitous) form, IP makes no delivery, reliability, flow control, or error recovery guarantees and can, as a result, lose or duplicate packets.

IP assumes that higher layer protocols or applications will detect and handle any of these problems. Obviously, this kind of network behavior can be problematic for real-time VoIP.

When an IP packet carrying digitized voice is lost, the voice signal will be distorted. Before describing the kinds of distortion packet loss can create, it is useful to briefly describe the causes of packet loss (Hardman, 2003):

- Packet Damage - Many applications will discard incoming packets, when presented with one that has been damaged. An example of packet damage is bit errors due to circuit noise or equipment malfunction.
- Network Congestion, Buffer Overflow, and IP Routing - Perhaps the largest cause of packet loss is packet discard due to network congestion. When a particular network component receives too many packets at one time, its receive buffers overflow causing packets to be discarded. IP networks also deal with network congestion by rerouting traffic to less congested network paths, but this can increase delay and jitter.

Typically, when packets are intentionally discarded due to damage or congestion, networking applications will retransmit the data. This can cause duplicate packets to be sent, can result in packets arriving too late to be used, or can cause packets to be received in the incorrect order. Figure 3.5; illustrate the affect of packet loss.

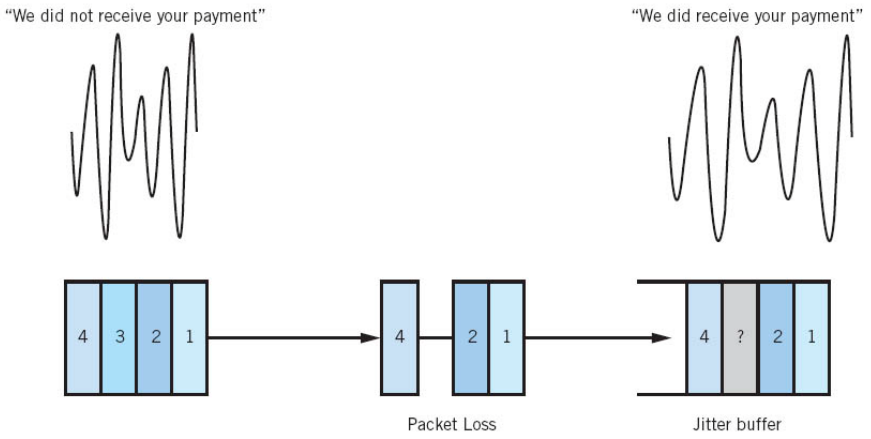


Figure 3.5 the affect of packet loss

3.10 Echo

The most important echo is **talker echo**, the perception by the talker of his own voice but delayed. It can be caused by electric (**hybrid**) echo or acoustic echo picked up at the listener side. (Hersent, Petit & Gurle, 2005)

If talker echo is reflected twice it can also affect the listener. In this unusual case the listener hears the talker's voice twice: a loud signal first, and then attenuated and much delayed. This is **listener echo**. These two types of echo are illustrated on Figure 3.6.

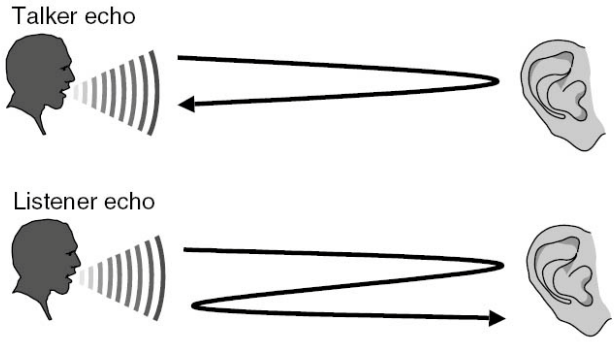


Figure 3.6 Two types of echo (Hersent, Petit & Gurle, 2005)

Hybrid echo (also known as "electrical echo") is caused by an impedance mismatch on the 4-wire to 2-wire conversion in wire line networks. It is the primary network-induced echo in today's networks. Acoustic echo is created as a result of insufficient acoustic isolation between the earpiece and the microphone, or when acoustic waves are reflected against a wall or enclosure, typically when using a hands-free unit. (Anonymous, 2011)

Echo, like delay, influences conversational quality more than it does sound quality. To solve echo problem Echo cancellation technologies uses digital signal processing (DSP) and echo cancellation algorithms.

3.11 VoIP Quality of Service (QoS)

For VoIP to be a realistic replacement for standard public switched telephone network (PSTN) telephony services, customers need to receive the same quality of voice transmission they receive with basic telephone services—meaning consistently high-quality voice transmissions. As mentioned, like other real-time applications, VoIP is extremely band-width and delay-sensitive. For VoIP transmissions to be intelligible to the receiver, voice packets should not be dropped, excessively delayed, or suffer varying delay (Jitter).

VoIP can guarantee high-quality voice transmission only if the voice packets, for both the signaling and audio channel, are given priority over other kinds of network traffic. For VoIP to be deployed so that users receive an acceptable level of voice quality, VoIP traffic must be guaranteed certain compensating bandwidth, latency, and jitter requirements. QoS ensures that VoIP voice packets receive the preferential treatment they require. (Kashihara, 2011)

Quality of Service allows control of data transmission quality in networks, and at the same time improves the organization of data traffic flows, which go through many different network technologies. Such a group of network technologies includes ATM (asynchronous transfer mode), Ethernet and 802.3 technologies, IP based units, etc.; and even several of the abovementioned technologies can be used together. An illustration of what can happen when excessive traffic appears during peak periods can be found in everyday life: an example of filling a bottle with a jet of water. The maximum flow of water into the bottle is limited with its narrowest part (throat). If the maximum possible amount of decantation (throughput) is exceeded, a spill occurs (loss of data). A funnel used for pouring water into a bottle, would in case of data transfer be in the waiting queues. They allow us to accelerate the flow, and at the same time prevent the loss of data. A problem remains in the worst-case scenario, where the waiting queues are overflowed, which again leads to loss of data (a too high water flow rate into the funnel would again result in water spills).

Priorities are the basic mechanisms of the QoS operating regime, which also affects the bandwidth allocation. QoS has an ability to control and influence the delays which can appear during data transmission. Higher priority data flows have granted preferential treatment and a sufficient portion of bandwidth (if the desired amount of bandwidth is available). QoS has a direct impact on the time variation of the sampling signals which are transmitted across the network. Such sampling time variation is also called jitter (T. & S.Subash IndiraGandhi, 2006). Both mentioned properties have a crucial impact on the quality of the data and information flow throughput, because such a flow must reach the destination in the strict real-time. A typical example is the interactive media market. QoS reflects their distinctive properties in the area of improving data-transfer characteristics in terms of smaller data losses for higher-priority data streams. The fact that QoS can provide priorities to one or more data streams simultaneously, and also ensure the existence of all remaining (lower-priority) data streams, is very important. Today, network equipment companies integrate QoS mechanisms into routers and switches, both representing fundamental parts of Wide Area Networks (WAN), Service Provider Networks (SPN), and finally, Local Area Networks. Based on the abovementioned points, the following conclusion can be given: QoS is a network mechanism, which successfully controls traffic flood scenarios, generated by a wide range of advanced network applications. This is possible through the priorities allocation for each type of data stream. (Kashihara, 2011)

QOS could classify too many factors in different level (At the edges of network, In the middle of network, and at the end of network) and different kinds of solution have been proposed that don’t have opportunity to express all.

CHAPTER FOUR

VOICE QUALITY MEASUREMENT IN VOIP

With the advent of Voice over Internet Protocol (VoIP) service, assessing its voice quality is an area of intense research interest.

Quality in non-managed IP networks such as the Internet is not guaranteed, therefore, it is important to monitor the speech quality in telecommunication systems and take appropriate actions when necessary. It is also important to measure the quality even in managed networks. In addition to its importance for legal, commercial and may be technical reasons, this also allows service providers to evaluate their own and their competitors' service using a standard scale. It is also a strong indicator of user's satisfaction of the service provided. (Takahashi, Yoshino & Kitawaki, 2004)

In the development of a VoIP systems solution, the developer needs to be fully assured that each of the components of the system that could affect voice quality is performing to their specification. For example, there are several components to a VoIP system which involve various signal processing algorithms. Some of these are packetization, echo cancellation, speech codec, delay, jitter processing, packet-loss and packet-loss recovery schemes, etc. All of these techniques impact the voice quality of transmitted speech to varying degrees. Therefore, it is important to measure voice quality in order to quantify the impact of each of these components.

Decisions such as these are crucial in the marketplace both for manufacturers /suppliers of telecom equipment in equipment selection, and for operators to ensure and monitor good/acceptable voice quality, and to fully optimize their networks. Hence, a valid measure of voice quality is an extremely valuable and much needed metric. Thus an end-to-end or systems' measure of voice quality is an essential feature characterizing any successful VoIP system. It is one of the features that

determine the market success of a product used in telecommunications. (Anonymous, 2002)

4.1 Overview of VoIP measurement methods

VoIP quality assessment methods can be categorized into either subjective methods or objective methods. Objective methods can be either intrusive or non-intrusive. Non-intrusive methods can be either signal-based or parametric-based. Figure 4.1 depicts different classifications.

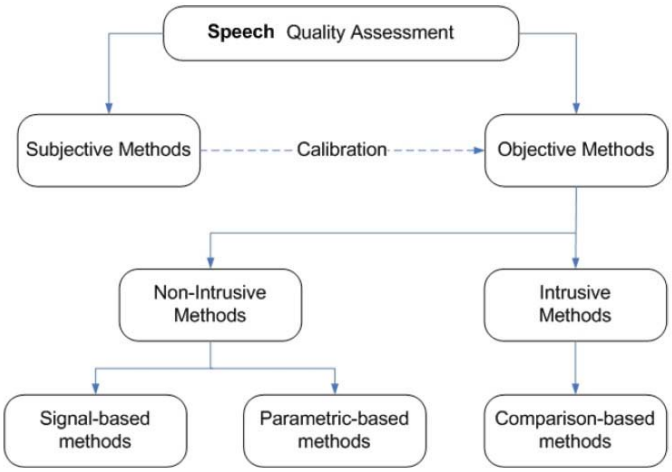


Figure 4.1 Overview of VoIP measurement methods (Sun, 2004)

The primary criterion for voice and video communication is subjective quality, the user's perceptions of service quality. A subjective quality assessment method is used to measure the quality. Subjective quality factors affect the quality of service of VoIP, among those factors are: packet loss, delay, jitter, loudness, echo, and codec distortion. To measure the subjective quality, a subjective quality assessment method is used; the most widely accepted metric is the Mean Opinion Score (MOS) as defined by ITU-T Recommendation P.800 (ITU-T, 1996b).

However, although subjective quality assessment is the most reliable method, it is also time-consuming and expensive as any other subjective test. Thus other methods to automatically estimate quality objectively should be considered. This can be done intrusively by comparing the reference signal with the degraded signal or non-

intrusively utilizing physical quality parameters or the received signal without using the reference signal.

The applicability of any solution for measuring the speech quality in VoIP networks should take into consideration the nature of IP networks and the characteristics of voice traffic. Among the desired features for a VoIP speech quality assessment solutions are (Kashihara, 2011):

- *Automatic*: It should provide measurement of speech quality online while the network is running.
- *Non-intrusive*: It should be able to provide measurement of the speech quality depending on the received speech signal or network parameters without the need for the original signal.
- *Accurate*: It should provide accurate measurement of speech quality to reflect how the quality is perceived by the end-user.

With the changing world, it should be applicable to new and emerging applications and networking conditions. As such it should avoid the subjectivity in estimating parameters. The E-model (section 4.3.4) for example depends on subjective tests to estimate packet loss parameters which hinder its applicability for new networking conditions.

4.2 Subjective assessment of quality

The most widely used subjective quality assessment methodology is opinion rating defined in ITU-T Recommendation P.800 in which a panel of users (test subjects) perform the subjective tests of voice quality and give their opinions on the quality (ITU-T, 1996b).

Subjective tests could be conversational or listening-only tests. In conversational test, two subject share a conversation via the transmission system under test, where they are placed in separated and isolated rooms to report their opinion on the opinion

scale recommended by ITU-T and the arithmetic mean of these opinions is calculated. In listening tests, one subject is listening to pre-recorded sentences (ITU-T, 1996b).

In opinion rating methodology the performance of the system is rated either directly (Absolute Category Rating, ACR) or relative to the subjective quality of a reference system as in (Degradation Category Rating, DCR), or Comparison Category Rating (CCR) (ITU-T, 1996b; Takahashi, 2004).

4.2.3 Mean Opinion Score (MOS)

The most common metric in opinion rating is Mean Opinion Score (MOS) which is an ACR metric with five-point scale: (5) Excellent, (4) Good, (3) Fair, (2) Poor, (1) Bad (ITU-T, 1996b).

Figure 4.2 The ITU-T MOS listening quality scales.

MOS is internationally accepted metric as it provides direct link to the quality as perceived by the user. A MOS value is obtained as an arithmetic mean for a collection of MOS scores (opinions) for a set of subjects.

In May 2003 ITU-T approved recommendation P800.1 (ITU-T, 2006) that provides a terminology to be used in conjunction with voice quality expressions in terms of MOS. As shown in Table 4.1, this new terminology is motivated by the intention to avoid misinterpretation as to whether specific values of MOS are related to listening quality or conversational quality, and whether they originate from

subjective tests, from objective models or from network planning models. The following identifiers are recommended to be used together with the abbreviation MOS in order to distinguish the area of application: LQ to refer to listening quality, CQ to refer to conversational quality, S to refer to subjective testing, O to refer to objective testing using an objective model, and E to refer to estimated using a network planning model. (Mahdi & Picovici, 2007)

Table 4.1 MOS Terminology

Measurement	**Listening-Only**	**Conversational**
Subjective	MOS-LQS	MOS-CQS
Objective	MOS-LQO	MOS-CQO
Estimated	MOS-LQE	MOS-CQE

While subjective quality assessment techniques are used extensively, they do have their limitations. For example, they require specialized methodologies, personnel with specific expertise, controlled acoustic environments, well defined and controlled speech inputs, and are very time consuming and expensive to implement. Moreover, they are not practical as field tests or for testing a very large number of parameters.

These limitations, and the need to obtain a measure of voice quality in field applications in a timely manner, have motivated the development of objective tools that measure voice quality.

4.3 Objective voice quality measures

Objective testing methods are based on measurements of physical quantities of the system such as delay, jitter and packet loss. Typically, this can be achieved either by injecting a test signal into the system or by monitoring live traffic.

Contrary to subjective tests, objectives tests can be repeatedly carried out to evaluate the performance of a system under different set of parameters. (Karapantazis & Pavlidou, 2008)

Objective quality assessment methodologies can be categorized into two groups: Intrusive speech-layer models and Non-Intrusive models (Signal-based and parametric-based). Figure 4.3 shows the three main types of objective measurement.

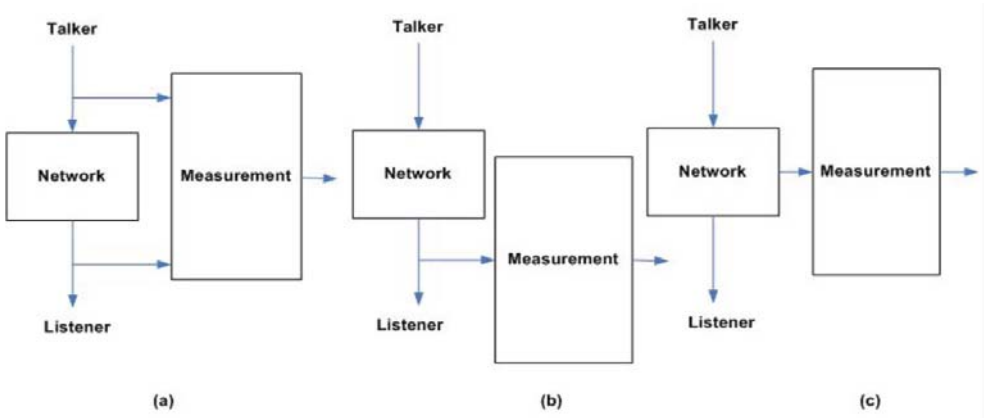


Figure 4.3 Three main categories of objective quality measurement: (a) Comparison-based Intrusive method, (b) Signal-based non-intrusive method, (c) Parametric-based non-intrusive Method (Sun, 2004)

The accuracy, effectiveness and performance evaluation of an objective measure is, therefore, determined by the correlation of its scores with the subjective MOS scores. If an objective method has a high correlation, typically greater than 0.8, it is deemed to be effective measure of perceived voice quality, at least for the speech data and transmission systems with the same characteristics as those in the test experiment. (Kashihara, 2011) A description of the most well-known objective testing methods follows.

4.3.1 PSQM

The Perceptual Speech Quality Measure (PSQM) represents the first objective testing method developed by John G.Beerends and J.A. Stemerdink at the KPN Research in 1996. It is defined in ITU-T Recommendation P.861 (ITU-T, 1998). At the time that this method was standardized, the scope was to devise a method for evaluating voice codecs used primarily in cellular systems. The measure of quality predicted by PSQM is given on a scale from 0 to 6.5. PSQM is appropriate for evaluating speech quality in environments that are not subject to bit or frame errors. Thus, this standard

is not suited for the assessment of networks, but rather for codecs, and it was withdrawn when PESQ was adopted in February 2001.

4.3.2 PAMS

The Perceptual Analysis Measurement System (PAMS) was developed by British Telecom in order to evaluate the perceived voice quality (Rix & Holier, 2000). PAMS was the first method to provide robust assessment for VoIP. This repeatable testing method derives a set of scores by comparing one or more high-quality reference speech samples to the processed audio signal. The resulting score is on a MOS-like scale from 1 to 5. Nevertheless, PAMS does not always show 100% correlation with live-listener MOS test scores conducted on the same speech samples.

4.3.3 PESQ

The Perceptual Evaluation of Speech Quality (PESQ) method is an intrusive method developed jointly by British Telecom and KPN. The details of this method are spelt out in ITU-T Recommendation P.862 (ITU-T, 2001). PESQ combines the excellent psycho-acoustic and cognitive model of PSQM with a time alignment algorithm adopted from PAMS that is able to handle varying delays perfectly. In this method, the received signal is compared to the one initially transmitted using a perceptive hearing model that is a replica of the human hearing apparatus. The result of this comparison is given on a scale from 1 to 4.5.

4.3.4 E-Model

A useful tool for assessing the relative impact of transmission planning decisions on speech performance is the E-Model. The E-Model has been included in various ITU-T Recommendations on transmission planning. In particular, the E-Model is the subject of Recommendation G.107. (ITU-T, 1998)

Notwithstanding, it has also been adopted by ETSI (European Telecommunications Standards Institute) and TIA (Telecommunications and Industry Association) and has become the most widely used tool for objective assessment of speech quality. This model is predicated upon the assumption that the impairments caused by transmission parameters have an additive effect on speech quality. According to this model, speech quality is determined by the following equation:

$$R = Ro - Is - Id - Ie + A \quad (4.1)$$

where Ro accounts for noise effects, Is represents impairments such as too loud speech level, non-optimum side tone and quantization noise, Id is the sum of impairments due to delay and echo effects, Ie represents impairments due to low bit rate voice codecs, whereas A represents an ‘‘advantage of access” that some systems have in comparison with PSTN (for instance, A for mobile systems is 10). It should be noted that the first two terms in Eq. (4.1) are intrinsic to voice signal itself and do not depend on the transmission of voice over the Internet. The value of this function has a nominal range from 0 for terrible up to 100 for perfect voice. However, there is a direct relation between R and the MOS score.

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

For R<0: MOS = 1
For 0<R<100: MOS = 1+0.035R+R(R-60)(100-R)7*10^-6 (4.2)
For R>100: MOS = 4.5

Figure 4.4 depicts the relation among PSQM, PAMS, PESQ, MOS and E-Model scales.

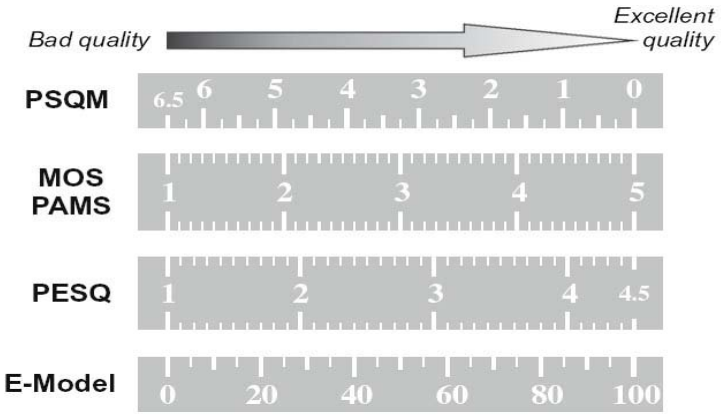


Figure 4.4 PSQM, PAMS, PESQ and MOS scales in comparison.

Even though the E-Model is widely used for the assessment of voice quality, it can also be used for selecting some network parameters such as the voice codec and the maximum allowable link utilization. (Gardner, Frost & Petr, 2003)

Id and Ie are the two parameters among all the others that are important for the VoIP systems. Equation (4.1) can be modified after substituting the default values of the other parameters (Cole & Rosenbluth, 2001). Equation (4.3) is the modified and final equation to determine the R-factor that determines voice quality in a VoIP application using best-effort networks to transmit information.

$$R = 93.2 - Ie - Id + A \tag{4.3}$$

Where A is the Advantage factor; 0 for wire line and 5 for wireless networks.
The value of Ie, which is codec dependent impairment, is calculated as:

$$Ie = a + b \ln (1 + cP/100) \tag{4.4}$$

Where, P is percentage packet loss and a, b and c are codec fitting parameters. (ITU-T, 2001)

Table 4.2 Fitting parameters for different codecs (Mehta &Udani, 2001).

Parameters	G711	G729(10ms)	G729(20ms)	G723.1	iLBC
Bit rate(Kb/s)/Frame size(ms)	64/20	8/10	8/20	6.3/20	15.2/20
a	0	10	10	15	10
b	30	25.21	25.21	36.59	19.8
c	15	15	20.2	6	29.7

The value of Id, which is impairment due to delay is calculated as (Gambhir, 2009):

$$Id = 0.024d + 0.11 (d – 177.3) H (d – 177.3) \quad (4.5)$$

Where d is the total one way delay in milliseconds is calculate as:

$$d = \text{One way delay} + \text{Packetization delay} + \text{Processing Delay} \quad (4.6)$$

And H(x) is a step function defined as: $H(x) = 0$, $x<0$ and $H(x) = 1$ otherwise.

4.4 VoIP Voice Quality Measurement by Network Traffic Analysis

Understanding network traffic characteristics in terms of packet-size distributions is important because it has implications for the end-to-end performance achieved by the traffic streams. (Calyam & Lee, 2005) In addition, voice quality measurement by network traffic analysis can help better manage, optimize and troubleshooting networks.

Non-Intrusive or Passive Monitoring examines a stream of voice traffic and produces a transmission quality metric that can be used to estimate a MOS score. This has the advantage that all calls in a network can be monitored without any additional network overhead but the disadvantage that the effects of some impairment are not incorporated.

As mentioned in chapter 3, Voice quality composes of three main components: clarity, delay, and echo. Clarity and echo are independent while echo relies on delay threshold. The proportional contribution that each factor affects voice quality is pretty fuzzy since the subjective test can be interpreted in different ways. To easily manage the evaluation, only some most significant parameters should be strictly used on evaluation. However, the tested parameters must encompass all voice quality characteristics.

To practically measure voice quality, it is possible to discard some unnecessary variables. Four primary parameters sufficiently representing voice performance factors are delay, jitter, loss rate, and codec.

The first and most recognizable component, clarity, is measured by loss rate, jitter, and codec. The next component, delay, is measured by the propagation time between hosts. In addition, the compression and packetization times are included in the overall delay.

The last component, echo, should be measured by TELR (Talker Echo Loudness Rating) and the end-to-end transmission time. According to the current VoIP application design, the echo canceller on the tail-end host performs effectively and diminishes the echo amplitude to lower than -25 dB, which is unrecognizable by human. Moreover, echo presents a negative impact only when the end-to-end transmission time is beyond a certain threshold. So TELR is ignored and only the transmission time is measured in this study.

Therefore, the tests of this study are designed to measure delay, jitter, and loss rate. These objective parameters are also used in E-model and many VoIP performance measurement applications.

The measurement method that is offered in this thesis provides passive monitoring through observation of the RTP stream and incorporates effects. This produces an R-Factor which can be used to estimate a MOS score.

To get accurate assessment results of an individual VoIP system performance was tried to process measurement in two phases:

- Call Setup, Call Completion, and Services Testing
- Packet Performance Testing

4.4.1 Call Setup, Call Completion, and Services Testing

An important area of VoIP operations and performance that must be tested involves the signaling that occurs to establish, maintain, and disconnect VoIP telephone calls. Metrics include percentages of call success/completion, call services validation, call setup times, and so on. This aspect of VoIP operations has little direct effect on voice signal quality. However, “negotiations” occur during some call setup processes between VoIP entities which can result in a noise or distortion baseline. For example, SIP signaling protocols negotiate codecs and other channel characteristics. Protocols analyzers that can deliver data stream decode are often used for this type of testing. (Hardman, 2003)

4.4.2 Packet Performance Testing

Given the impact packet loss and jitter have on a voice signal carried across a VoIP network, it is clear that packet delivery performance must be tested. Test methods can range from monitoring actual IP traffic to find evidence of packet loss and jitter, to injecting into the network under test specific packet streams with specific transmission and payload characteristics. Data communications test solutions that provide VoIP decodes, RTP and RTCP monitoring, and general IP traffic analysis capabilities represent perhaps the best ways to measure packet performance in a voice over IP environment. (Hardman, 2003)

In the next chapter these phases of measurement are described with experiments.

CHAPTER FIVE

EXPERIMENTAL WORKS AND RESULTS

This chapter describes and tests the VoIP systems performance measurement method which is offered in this thesis. First test environments are described in section 5.1; then three different measurement phases are implemented on each designed test bed. Lastly, to assess the method accuracy, the subjective satisfaction scores that are also collected simultaneously are correlated.

5.1 Test Bed Design

In this study, experiments in three scenarios were performed (A, B and C). In the first scenario (A), two computers were connected together with crossover cable and VoIP LAN was implemented by using a soft phone that was installed on each computer. One of the computers was used as an SIP server by installing virtual SIP server software. Scenario B was designed like scenario A, except it was tested on a WLAN network. In the last scenario (C), in order to analyze VoIP voice quality on WAN, instead of a virtual SIP server, an actual free SIP service provider was used.

5.1.2 Scenario A

In this experiment, VoIP traffic was analyzed using a soft phone named X-Lite. X-Lite is a proprietary freeware VoIP soft phone that uses the Session Initiation Protocol. It is a suite of voice calls, video calls and Instant Messaging, internet telephone software.

A simple test bed (shown in Figure 5.1) was setup to collect the traffic on an X-Lite system. X-Lite was installed on two computers connected via copper cross-over cable. Because X-Lite does not include a service MiniSIPServer software was installed to provide SIP service. MiniSipServer is a professional cross-platform VOIP server (SIP Software BPX) which can run on Windows and Linux/Ubuntu system.

Figure 5.1 Scenario (A) test bed setup

Wireshark was the primary data capture and protocol analysis software used in this experiment to capture and analyze the VoIP traffic in packet-size. Wire shark is a popular network analyzer widely used by network professionals for troubleshooting, analysis software and protocol development, and teaching. It reads packets from either the network or a trace file, decodes them, and presents them in an easy to understand format. Wireshark was chosen as the tool to use because it is an actively-maintained open source program and its graphical user interface is very configurable and easy to use. The following are the primary features of Wireshark:

- It can capture data from the network or read from a captured file.
- It supports Tcpdump format capture filters.
- It runs on more than 20 OS platforms, both UNIX-based and Windows.
- It supports over 480 protocols, and because it is open source, new ones are contributed very frequently.

Finally, after gathering the all information about experimental calls, R value and MOS were calculated with equation 4.3 and 4.2 (shown in section 4.3.4).

5.1.3 Scenario B

By changing the network used in scenario A to a wireless network, VoIP voice quality was measured using a wireless network in scenario B. In this scenario, a wireless router was used to build a wireless network between two end users.

Figure 5.2 Scenario (B) test bed Setup

5.1.4 Scenario C

In this scenario, the voice quality measurement method that was used in this thesis was tested in an internet network environment. The soft phone was Blink software and SIP2SIP free SIP service provider was used as an SIP server.

Figure 5.3 Scenario C Test bed Setu

5.2 Test bed Experiments

As mentioned in section 4.4 of this thesis, in this measuring method three phases of analysis were processed in each test bed scenario. Call setup and packet performance were analyzed and voice quality was calculated for each scenario. In all experiments, a trace of five calls had been collected with approximately five minutes of voice data for each call. Subjective voice quality scores (MOS – CQS) were also collected simultaneously.

5.2.3 Trace Files Information

The information of the trace files in experiments is summarized separately in Tables 5.1, 5.2 and 5.3.

Table 5.1 Scenario A traces file information

Attribute	**Sample test value**	**Total calls average value**
Codec Type	G711 alaw	
Network Potential BW	100 Mb/s	
Total Packets	30406	32796
Average Packets / sec	90.634	102.071
Average Packet Size	213.577	202.957
Total Call Data Size (Bytes)	6494028	6655823.6
Average bytes / sec	19357.363	20716.1
Average MBit / sec	0.155	0.166
Total Time	5.4 min	5 min

Table 5.2 Scenario B traces file information

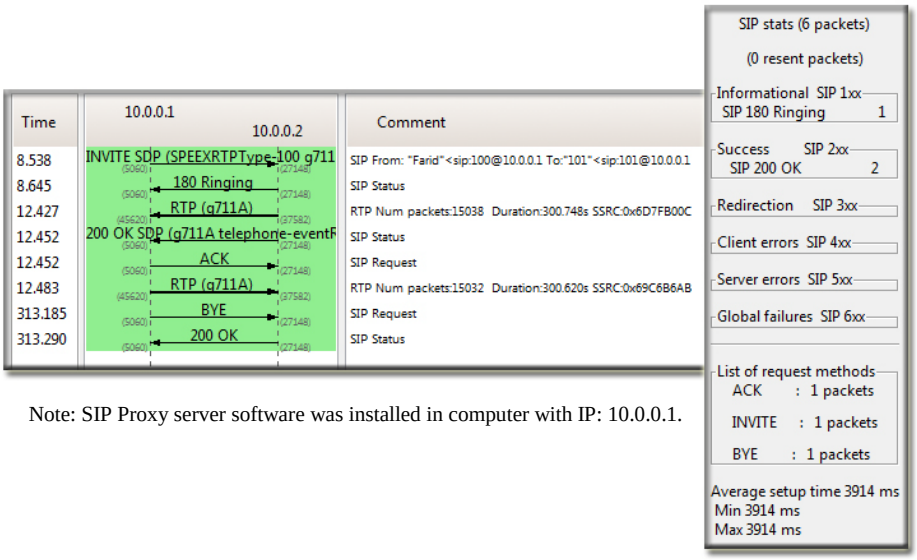
Attribute	**Sample test value**	**Total calls average value**
Codec Type	G711 alaw	
Network Potential BW	54 Mb/s	
Total Packets	34883	32198
Average Packets / sec	106.209	101.35
Average Packet Size	205.611	216.57
Total Call Data Size (Bytes)	7172325	6973125
Average bytes / sec	21837.805	21949.34
Average MBit / sec	0.175	0.176
Total Time	5.2 min	5 min

Table 5.3 Scenario C traces file information

Attribute	Sample test value	Total calls average value
Codec Type	G711 alaw	
Network Potential BW	Variable	
Total Packets	30703	32097
Average Packets / sec	98.552	95.416
Average Packet Size	206.176	211.634
Total Call Data Size (Bytes)	6330230	6792924
Average bytes / sec	20319.037	20193.338
Average MBit / sec	0.163	0.162
Total Time	4.53 min	5 min

5.2.2 Call Setup Analysis

The packet format for different call signaling protocols is quite different. The following figures display sample SIP call setup processes for the three test bed scenarios; the tables summarize sample tests and average call set up times score and percentage of call success for each test bed scenario.



Note: SIP Proxy server software was installed in computer with IP: 10.0.0.1.

Figure 5.4 Scenario A test bed sample test Call setup information with SIP packet flow in the horizontal tab and analysis shown in the vertical tab.

As results show that 6 packets are rout successfully and average call setup time for all 5 tests in this scenario is 1821.4 ms. The percentage off success call has been 100 percent in all of tests.

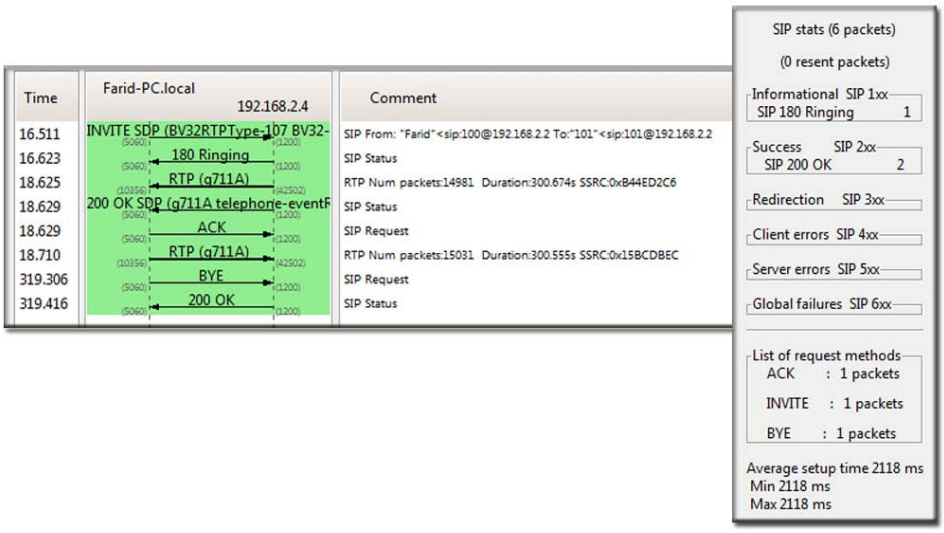


Figure 5.5 Scenario B test bed sample test call setup information with SIP packet flow in the horizontal tab and analysis shown in the vertical tab.

As results show, all call setup packets are rout without any error and average call setup time for all 5 tests in this scenario tests is 4520 ms. As same as LAN testing outcome, call success percentage is 100% in all tests.

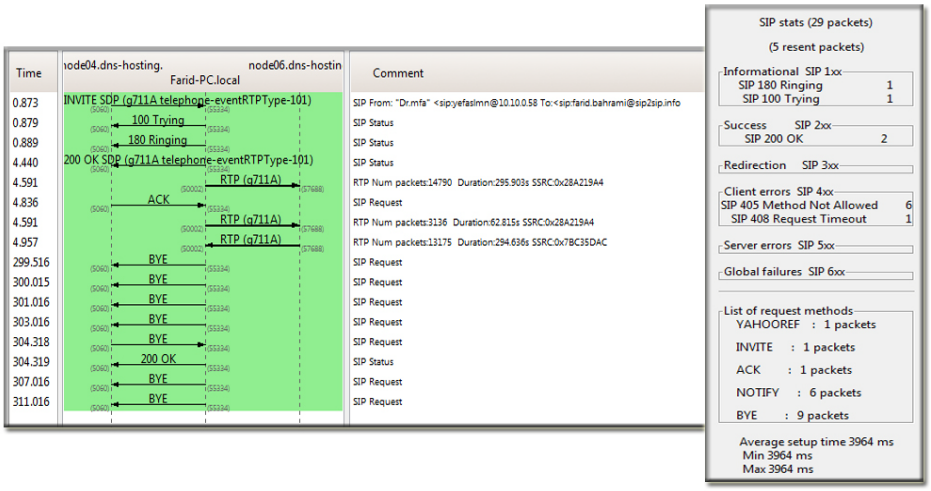


Figure 5.6 scenario C test bed sample test Call setup information with SIP packet flow in the horizontal tab and analysis shown in the vertical tab.

Figure 5.6 (sample test of scenario C) shows that SIP route 29 packets at all. In ‘client error SIP’ section, 4xx type error showed up. They came from other program that was active at test time simultaneously. The average call setup time

for all 5 tests in this scenario was 4427 ms and call success percentage came out 100%.

As mentioned in section 4.4.1 however this aspect of VoIP operations has little direct affect on voice signal quality but problems of call could be monitored and solved by analyzing call setup signaling.

5.2.3 Packet Performance Analysis

After the tests were completed, collected packets were analyzed. The most significant parameters that affect voice quality for each test bed scenarios are shown in tables 5.4.

Table 5.4 the packet performance statistics of test scenario A, B and C's calls

	Sample Test			**Total tests average**		
Scenario	**A**	**B**	**C**	**A**	**B**	**C**
Average Delay (ms)	100.4	98.6	165	97.72	101.94	167.5
Average Jitter (ms)	2.75	45	20.9	3.15	7.15	21.133
Average Packet loose (Jitter buffer)	1.7%	7%	4.4%	2.34%	4.7%	4.37 %

According to the results, in the LAN tests (Scenario A), total tests average transmission delay and jitter of RTP packet is too low. For the wireless LAN tests (Scenario B), the average delay and jitter are a little bit longer than LAN tests. The delay and jitter of the WAN tests are more than too others.

Every single test was first evaluated based on the assumption of symmetric delay. Only the WAN test, asymmetric delays were also considered. Because of inappropriate jitter buffer, unexpected packet loose, was reported in the LAN and WLAN test.

Note that all RTP packets that are dropped cause of the jitter buffer are reported ("Drop by Jitter Buff") as well as the packets that are out of sequence (Out of Seq).

5.2.4 Voice Quality Analysis

The E-Model's R value can be processed by data derived from the previous section and by using equation 4.3 in section 4.3.4. The MOS-CQO values were calculated from R values by using equation 4.2 for each scenario. Subjective voice quality scores (MOS-CQS) were also collected simultaneously. Table 5.5 show these values for each test scenario.

Table 5.5 Voice Quality Measurement Results for test Scenarios

	Sample Test			**Total Calls Average**		
Test Scenarios	**A**	**B**	**C**	**A**	**B**	**C**
R value	83.37	68.7	73.435	82.32	74.30	72.614
MOS - CQO	4.14	3.54	3.75	4.09	3.79	3.71
MOS - CQS	3.8	3.5	3.5	3.9	3.74	3.65

As seen in table 5.5, the quality of voice in test scenarios are decreased due to the increase delay and packet loose average value that mentioned in table 5.4. While the most R value belong to scenario A, least R value calculated for Scenario C's tests related to the highest delay average delay and packet loose.

5.3 The Measurement Method Accuracy Analysis

Correlation of calculated MOS-CQO and collected MOS-CQS scores could help to estimate the developed measurement method. Table 5.6 shows these values.

Table 5.6 Correlation of total tests average MOS-CQO and MOS-CQS scores

Test Scenario	**MOS-CQO average score**	**MOS-CQS average score**	**Pearson Correlation**
A	4.09	3.9	0.79
B	3.79	3.74	0.99
C	3.71	3.65	0.93

As mentioned in section 4.3 of chapter 4, the objective method with a correlation greater than 0.8 is deemed to be effective.

The correlation values of the test scenarios B and C are more than 0.8 but the calculated correlation value of scenario A is less than 0.8. Since of the human's ear sensitivity limitation, little voice quality changes that affect on R value can't sense by user's ear.

However, subjective MOS scores are affected by a lot of factors like test environment situation, headphone quality etc. Therefore, the correlation values which were calculated in this thesis may change in different test situations with different users.

CHAPTER SIX

CONCLUSION

6.1 Conclusion

The increasing expectation levels for better audio and video performance has led to the need to understand the behavior of audio and video traffic as it affects end user perceived quality of the application and voice over the network.

Understanding network traffic characteristics in terms of packet-size distributions is important because it has implications for the end-to-end performance achieved by the traffic streams. In addition, voice quality measurement by network traffic analysis can help better manage, optimize and troubleshooting networks.

The non-intrusive, parametric objective measurement method that is offered in this thesis provides passive monitoring through observation of the RTP stream and incorporates effects. This produces an R- Factor which can be used to estimate a MOS score.

The measurement method was applied on three different test bed scenarios (A, B and C). In the first scenario (A), two computers were connected together with crossover cable and VoIP LAN was implemented by using a soft phone that was installed on each computer. One of the computers was used as an SIP server by installing virtual SIP server software. Scenario B was designed like scenario A, except it was tested on a WLAN network. In the last scenario (C), in order to analyze VoIP voice quality on WAN, instead of a virtual SIP server, an actual free SIP service provider was used.

According to the results from LAN test, the transmission delay, jitter and loss of RTP packet is very low. For the wireless LAN test, the average delay was a little bit longer. The WAN test produced the largest delays. Every test was first evaluated

based on the assumption of symmetric delay. Only for the WAN test, asymmetric delays were also considered.

Because of inappropriate jitter buffer, unexpected packet loss, in LAN and wireless LAN tests, was reported.

Calculated objective and subjective MOS correlation values indicate the assessment method is an effective measure of perceived voice quality, at least for the speech data and transmission system with the same characteristics as those in the test experiments.

6.2 Future Work

This thesis focused on VoIP voice quality assessments, and measuring voice quality in different types of network between two users. It will be interesting to use the method developed in this thesis to test the VoIP voice quality between groups of user conferencing.

In this thesis SIP protocol with G711 codec was used in all tests; it will be changed to H.323 with another codec to test performance differences by the offered method.

Finally, it will be interesting to test the performance of video phone applications. The integration of voice and video media may further test the reliability of this thesis method since the media frame size is much larger and more bandwith is required.

REFERENCES

Anonymous, (2011). *Echo basics tutorial*. Retrieved December 15, 2010, from http://www.ditechnetworks.com/learningCenter/echoBasics.html

Anonymous, (2011). *Licenced & open source voice codecs in IP telephony*. Retrieved May 10, 2010, from http://www.excitingip.com/1088/licenced-open-source-voice-codecs-used-in-ip-telephony-voip/

Anonymous, (2010). *PSTN vs VoIP*. Retrieved October 12, 2010, from http://www.voipproducts.org/pstn-vs-voip/

Anonymous, (2010). *Voice over IP solution*. Retrieved September 24, 2010, from http://www.thesmenetwork.com/

Anonymous, (2008). *Diagnosing voice quality impairments and designing solutions for Voice over IP Systems*. Retrieved May 9, 2010, from http://www.dialogic.com.

Anonymous, (2002). *Measuring voice quality*. Retrieved June 20, 2010, from http://www.mindspeed.com.

Bellemy, J. C. (2000). *Digital Telephony* (3rd ed.). New York:John Wiley and Sons Wiley Press.

Bosch, L. T., Oostdijk, N. & Ruiter, J. P. (2004). Durational aspects of turn-taking in spontaneous face-to-face and telephone dialogues, *in Proceedings 7th International Conference on Text, Speech, and Dialogue*, 563-570.

Broniecki, T. (2009). *A detailed analysis: VoIP comparisons of H.323 standard and SIP protocol in an 802.31g environment*. Master's thesis, University of Nebraska at Omaha.

Calyam, P. & Lee, C. (2005). Characterizing voice and video traffic behavior over the internet. *In Proc. of 20th International Symposium on Computer and Information Sciences (ISCIS)*, Istanbul, Turkey.

Collins, D. (2001). *Carrier grade voice over IP* (2th ed.). San Francisco ; McGraw-Hill Press.

Fiche, G. & Hébuterne, G. (2004). *Communicating systems & networks: traffic & performance.* London and Sterling, VA: Kogan Page Science.

Gambhir, N. M. (2009). *Objective measurement of speech quality in VoIP over wireless LAN during handoff*. Master's Thesis, San Jose State University.

Gardner, M.T., Frost, V.S. & Peter, D.W. (2003). Using optimization to achieve efficient quality of service in voice over IP networks. *IEEE International Conference on Performance, Computing and Communications Conference,* 475–480, USA.

Hardman, D. (2003). *Noise and voice quality in VoIP environments*. Retrieved September 23, 2010, from http://www.agilent.com/comms/XPI.

He, Q. (2007). *Analysing the characteristics of VoIP traffic*. Master's thesis, University of Saskatchewan at Saskatoon

Hersent, O., Petit. J. P. & Gurle, D. (2005). *Beyond VoIP protocols: Understanding voice technology and networking techniques for IP telephony.* John Wiley & Sons, Ltd. ISBN: 0-470-02362-7

IETF, (2002). *SIP: Session initiation protocol*. Retrieved May 5, 2010, from http://www.ietf.org/rfc/rfc3261.txt

ITU-T, (2006). *Recommendation P.800.1, Mean Opinion Score (MOS) Terminology*. Retrieved 5 May, 2010, from http://www.itu.int/rec/T-REC-P.800.1-200607-I.

ITU-T, (2001). *Recommendation P.862, Perceptual evaluation of speech quality (PESQ).* Retrieved 10 February, 2010, from http://www.itu.int/net/itu-t/sigdb/genaudio/Pseries.htm.

ITU-T, (2001). *Recommendation P.833, Methodology for derivation of equipment impairment factors from subjective listening-only tests*. Retrieved June 13, 2010, from http://www.itu.int/rec/T-REC-H.323/e.

ITU-T, (1999). *Recommendation H.323. Infrastructure of audiovisual services-systems and terminal equipment for audiovisual services*. Retrieved 18 March, 2010, from http://www.itu.int/rec/T-REC-H.323/e.

ITU-T, (1998). *Objective quality measurement of telephone band (300-3400 Hz) Speech Codecs*. Retrieved March 20, 2010, from http://www.itu.int/rec/T-REC-P.800-199608-I.

ITU-T, (1998). *The E-Model.* Retrieved September 19, 2010, from http://www.itu.int/rec/T-REC-G.107-199812-S.

ITU-T, (1996b). *Methods for subjective determination of transmission quality.* Retrieved April 26, 2010, from http://www.itu.int/rec/T-REC-P.800-199608-I.

ITU-T, (1993). *Effect of transmission impairments*. Retrieved August 12, 2010, from http://www.itu.int/rec/T-REC-P.11-199303-I.

Karapantazis, S. & Pavlidou. F. N. (2008). VoIP: A comprehensive survey on a promising technology. *Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Panepistimioupoli,* 54124 Thessaloniki, Greece, Elsevier

Kashihara, S. (2011). *VoIP Technologies*. Retrieved 14 November, 2010, from http://www.intechopen.com/books/show/title/voip-technologies.

Kurose, J., & Ross, K. (2003). Computer networks: *A top-down approach featuring the internet.* USA; Pearson Education Press.

Long. C. (2001), *IP network design*. California; McGraw-Hill Press.

Madisetti, V. K. & Williams, D. B. (1998). *The digital signal processing handbook*. Boca Raton; CRC Press.

Mahdi, A. E. & Picovici, D. (2007). *Advances in voice quality measurement in modern telecommunications*. Masters thesis, University of Limerick, from www.elsevier.com/locate/dsp.

Manka, D. (2007). *Voice over internet protocol testbed design for non-intrusive, objective voice quality assessment*. Master's thesis, United States Naval Academy.

Mastalir, J. (2005). *Understanding SIP-Based VoIP*, Retrieved 15 June, 2010, from http://www.packetizer.com/ipmc/sip/papers/understanding_sip_voip/.

Mehta, P. C. & Udani, S. (2001). *Overview of voice over IP*, Retrieved 1 January, 2010, from http://www.cis.upenn.edu/~udani/papers/OverView VoIP.pdf

Moller, S. (2000). *Assessment and prediction of speech quality in telecommunications*, 116-117, Boston: Kluwer Academic Press.

Peters, J. (2000). *Voice over IP fundamentals*. USA: Cisco Press.

Rix, A.W., & Holier, M.P. (2000). The perceptual analysis measurement system for robust end-to-end speech quality assessment, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 3, 1515–1518, Istanbul, Turkey.

Sasidar, A. (2004). *Cross platform implementation of VoIP in 802.31b environment.* Master's thesis, University of Nebraska at Omaha.

Sat, B. (2010). *Design and evaluation of real-time voice-over-IP (VOIP) systems with high perceptual conversational quality*. Master's thesis, University of Illinois at Urbana-Champaign.

Sat, B. & Wah, B.W. (2007). Evaluation of conversational voice quality of the Skype, Google-Talk, Windows Live, and Yahoo Messenger VoIP systems. *IEEE Int'l Workshop on Multimedia Signal Processing*, 9, 135-138, Retrieved Oct. 2010.

Schilzrinne, H. (2003). *RTP: A transport protocol for real-time applications*. Retrieved 6 September 2010, from www.ietf.org/rfc/rfc3550.txt

Stallings, W. (2007). *Data and computer communications.* Pearson Prentice Hall, Upper Saddle River.

Sun, L. (2004). *Speech quality prediction for voice over internet protocol networks*. PhD thesis, University of Plymouth,U.K.

Takahashi, A. (2004). Opinion model for estimating conversational quality of VoIP. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 3, iii,1072–5.

Takahashi, A., Yoshino, H. & Kitawaki, N. (2004). Perceptual QoS assessment technolo- gies for VoIP, *IEEE Communications Magazine 42* , 7, 28–34.

Tanenbaum, A. S. (2003). *Computer Networks*. Upper Saddle River, NJ: Pearson Education, Inc.