İSTANBUL TECHNICAL UNIVERSITY ★ INSTITUTE OF SCIENCE AND TECHNOLOGY

ANALYZING VOICE AND VIDEO CALL SERVICE PERFORMANCE OVER A LOCAL AREA NETWORK

M.Sc. Thesis by
Zehra Çiğdem YEŞİL

Department : **Electronics and Communication Engineering**

Programme : **Telecommunication Engineering**

FEBRUARY 2010

İSTANBUL TECHNICAL UNIVERSITY ★ INSTITUTE OF SCIENCE AND TECHNOLOGY

ANALYZING VOICE AND VIDEO CALL SERVICE PERFORMANCE OVER A LOCAL AREA NETWORK

M.Sc. Thesis by
Zehra Çiğdem YEŞİL
(504061340)

Date of submission : **25 December 2009**
Date of defence examination: **02 February 2010**

Supervisor (Chairman) : **Assoc. Prof. Dr. Selcuk PAKER (ITU)**
Members of the Examining Committee : **Prof. Dr. Ahmet Hamdi KAYRAN (ITU)**
Assis. Prof. Dr. Berk ÜSTÜNDAĞ (ITU)

FEBRUARY 2010

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

YEREL ALAN AĞLARDA SES VE GÖRÜNTÜ ÇAĞRILARININ HİZMET PERFORMANSI ANALİZİ

YÜKSEK LİSANS TEZİ

Zehra Çiğdem YEŞİL
(504061340)

Tezin Enstitüye Verildiği Tarih : 25 December 2009
Tezin Savunulduğu Tarih : 02 February 2010

Tez Danışmanı : Doç. Dr. Selçuk PAKER (İTÜ)
Diğer Jüri Üyeleri : Prof. Dr. Ahmet Hamdi KAYRAN (İTÜ)
Yrd. Doç. Dr. Berk ÜSTÜNDAĞ (İTÜ)

ŞUBAT 2010

FOREWORD

I would like to express my deep appreciation and thanks for my advisor. This work is supported by ITU Institute of Science and Technology. I also would like to thank to my family due to their excellent patience and tolerence during my thesis.

December 2009 | Zehra Çiğdem YEŞİL
Telecommunication Engineering

TABLE OF CONTENTS

Page

ABBREVIATIONS

ACELP : Algebraic Code Excited Linear Prediction
ADPCM : Adaptive Differential Pulse Code Modulation
ARC : Admission Confirmation
ARJ : Admission Reject
ARQ : Admission Request
ASO : Arbitrary Slice Ordering
AVC : Advanced Video Codec
BCF : Bandwidth Confirmation
BRJ : Bandwidth Reject
BRQ : Bandwidth Request
CABAC : Context Adaptive Binary Arithmetic Coding
CAR : Committed Access Rate
CAVLC : Context Adaptive Variable Length Coding
CBP : Coded Block Pattern
CCITT : International Telegraph and Telephone Consultative Committee
CIF : Common Interchange Format
CNAME : Canonical Name
CODEC : COder/DECoder
COPS : Common Open Policy Service
COS : Class Of Service
CQ : Custom Queuing
CRLF : Carriage Return and Line Feed
CS-ACELP : Conjugate Structure- Algebraic Code Excited Linear Prediction
DCT : Discrete Cosine Transform
DiffServ : Differentiated Services
DTMF : Dual Tone Multi Frequency
DWRR : Deficit Weighted Round Robin
FIFO : First In First Out
FTP : File Transfer Protocol
FQ : Fair Queuing
GBSC : Group-of-Blocks Start Code
GUI : Graphical User Interface
GK : Gate Keeper
GN : Group Number
GOB : Group of Blocks
GQuant : Group-of-Blocks Quantizer
GW : Gateway
HTTP : Hypertext Transfer Protocol
IEC : International Electrotechnical Commission
IEEE : Institute of Electrical and Electronics Engineers
IETF : Internet Engineering Task Force
IP : Internet Protocol
IPDC : Internet Protocol Device Control

ISDN	: Integrated Services Digital Network
ISO	: International Organization for Standardization
ISUP	: ISDN User Part
ITU	: International Telecommunication Union
JPEG	: Joint Photographic Experts Group
LAN	: Local Area Network
LD-CELP	: Low Delay-Code Excited Linear Predictio
LSB	: Least Significant Bit
MB	: Macro Block
MBA	: Macro-Block Address
Mbps	: Megabits per second
MC	: Multipoint Controller
MCU	: Multipoint Control Unit
MDRR	: Modified Deficit Round Robin
MEGACO	: MEdia GAteway COntrol Protocol
MG	: Media Gateway
MGC	: Media Gateway Controller
MGCP	: Media Gateway Control Protocol
MIME	: Multipurpose Internet Mail Extensions
MOS	: Mean Opinion Score
MP3	: MPEG-1 Audio Layer III
MPEG	: Moving Pictures Experts Group
MPMLQ	: Multi Pulse Maximum Likelihood Quantization
MPS	: Multipoint Processor
MQuant	: Macro-Block Quantizer
MTU	: Maximum Transmission Unit
MType	: Macro-Block Type
MVD	: Motion Vector Data
MWRR	: Modified Weighted Round Robin
NAL	: Network Abstraction Layer
NCS	: Network-based Call Signaling Protocol
OPNET	: Optimized Network Engineering Tool
OSI	: Open Systems Interconnection
PCM	: Pulse Code Modulation
PLC	: Packet Loss Concealment
PPS	: Packet Per Second
PQ	: Priority Queuing
PSTN	: Public Switched Telephone Network
PType	: Picture Type
QoS	: Quality of Service
QCIF	: Quarter Common Interchange Format
RAS	: Registration Admission and Status
RED	: Random Early Detection
RR	: Receiver Report
RSVP	: Resource Reservation Protocol
RTP	: Real-time Transport Protocol
RTCP	: Real-Time Control Protocol
RTD	: Round Trip Delay
RTSP	: Real-Time Streaming Protocol
SAP	: Session Announcement Protocol

SB-ADPCM : Sub Band - Adaptive Differential Pulse Code Modulation
SCCP : Skinny Client Control Protocol
SCTP : Stream Control Transmission Protocol
SDES : Source Description
SDP : Session Description Protocol
SG : Signaling Gateway
SGCP : Signal Gateway Control Protocol
SIP : Session Initiation Protocol
SIGTRAN : Signaling Transport Protocol
SLA : Service Level Agreement
SMTP : Simple Mail Transfer Protocol
SQCIF : Sub Quarter Common Interchange Format
SR : Sender Report
SS7/C7 : Signaling System 7/ CCITT 7
TCP : Transmission Control Protocol
TRIP : Telephony Routing over IP
TTL : Time To Live
UA : User Agent
UAC : User Agent Client
UAS : User Agent Server
UDP : User Datagram Protocol
URL : Uniform Resource Locator
UTF-8 : Unicode Transformation Format-8bit
VAD : Voice Activity Detection
VCL : Video Coding Layer
VLAN : Virtual Local Area Network
VLC : Variable Length Coding
VoIP : Voice over Internet Protocol
WAN : Wide Area Network
WFQ : Weighted Fair Queuing
WLAN : Wireless Local Area Network
WRR : Weighted Round Robin

LIST OF TABLES

Page

LIST OF FIGURES

Page

ANALYZING VOICE AND VIDEO CALL SERVICE PERFORMANCE OVER A LOCAL AREA NETWORK

SUMMARY

In last decades, VoIP is being a rapidly growing technology that enables to transport voice packets over data networks. VoIP became a viable alternative to the public switched telephone networks (PSTNs). However real-time applications such as voice communication are very susceptible to end-to-end delay and delay variation, thus requires a guaranteed quality of service. Voice packets sharing the same transmission medium with data packets need to be prioritized during the transmission. Hence some queuing mechanisms are used to give voice packets the highest priority. Also voice encoder schemes needs to be identified according to the system requirements.

In this thesis, we will analyze how the service quality requirements, such as end-to-end delay, delay variation and throughput, will be affected with the changing queuing mechanisms and voice encoder schemes over a local area network. We will then implement a wireless network including a set of wireless clients to the existing wireline network and analyze the simulation results measured by OPNET simulation tool.

In this thesis, we will simulate a real network which includes both voice, video and data communication simultaneously. Workstations are randomly assigned to different applications, such as voice, video and FTP. We will also implement a wireless network to our proposed system. However, this thesis does not include WIMAX (IEEE 802.16x) as a wireless medium, instead we will use IEEE 802.11 Wireless LAN technology.

This work is organized as follows: In Chapter 1, we present a brief introduction to VoIP and also the most common challenges of implementing voice communication into wireline or wireless networks are discussed. Common voice protocols, such as H.323, Session Initiation Protocol (SIP), Megaco and MGCP and video protocols such as H.261, H.263, H.264 will be described in Chapter 2 and Chapter 3, respectively. Chapter 4 covers the CODEC selection and factors affecting VoIP Quality of Service (QoS). The proposed scenarios and the simulation results with OPNET will be drawn in Chapter 5. Finally, conclusions will be discussed in Chapter 6.

YEREL ALAN AĞLARDA SES VE GÖRÜNTÜ ÇAĞRILARININ HİZMET PERFORMANSI ANALİZİ

ÖZET

Son yıllarda, VoIP, ses paketlerinin veri şebekeleri üzerinden iletimine izin veren ve hızla gelişmekte olan bir teknoloji halini almıştır. VoIP, analog PSTN şebekesine geçerli bir alternatif teşkil etmektedir. Buna rağmen, ses iletişimi gibi gerçek-zamanlı uygulamalar, uçtan uca gecikme ve gecikmedeki varyasyona oldukça duyarlıdır ve garanti edilmiş bir servis kalitesi gerektirir. Bu servis kalitesini sağlamak için, veri paketleriyle aynı iletim ortamını kullanan ses paketleri, iletim esnasında önceliklendirilmelidir. Bu yüzden, ses paketlerine en yüksek önceliği verebilmek için, bazı kuyruklama mekanizmaları kullanılmaktadır. Aynı zamanda ses kodekleri sistem ihtiyaçlarına göre belirlenmelidir.

Bu tezde, yerel alan ağlarda, uçtan uca gecikme, gecikme varyasyonu ve işlem hacmi gibi servis kalitesini etkileyen faktörlerin, değişen kuyruklama mekanizmaları ve değişen ses kodeklerinden nasıl etkilendiği analiz edilecektir. Daha sonra, mevcut kablolu sisteme, bir dizi kablosuz kullanıcı içeren kablosuz bir şebeke ilave edilecek ve OPNET ile ölçülen simülasyon sonuçları analiz edilecektir.

Bu tezde, ses, görüntü ve veri iletişimini aynı anda bünyesinde barındıran gerçek şebekeler simüle edilecektir. Kullanıcılara rastlantısal olarak ses, görüntü ve FTP gibi birtakım uygulamalar atanmıştır. Ayrıca önerilen kablolu şebekeye, kablosuz bir şebeke ilave edilerek sonuçlar incelenecektir. Ancak bu tezde bahsi geçen kablosuz teknoloji WIMAX' i (IEEE 802.16x) kapsamamaktadır. Bunun yerine simülasyon IEEE 802.11 teknolojisi kullanılarak gerçeklenecektir.

Bu tez şu şekilde düzenlenmiştir: Bölüm 1'de, kısaca VoIP teknolojisine giriş yapılacak ve bu teknolojiyi kablolu ve kablosuz ortamda gerçeklemenin en önemli darboğazları anlatılacaktır. H.323, SIP (Session Initiation Protocol), Megaco ve MGCP gibi yaygın olarak kullanılan ses iletim protokolleri ve H.261, H.263 ve H.264 gibi görüntü iletim protokolleri sırasıyla Bölüm 2 ve Bölüm 3'te incelenecektir. Bölüm 4, kodek seçimi ve VoIP servis kalitesine etki eden faktörleri kapsamaktadır. Önerilen senaryolar ve bu senaryoların OPNET ile ölçülmüş simülasyon sonuçları Bölüm 5'te incelenecektir. Sonuçlar ise Bölüm 6'da tartışılacaktır.

1. INTRODUCTION

Voice over Internet Protocol (VoIP) is a rapidly growing technology that enables the transport of voice communication over the public Internet. VoIP became an eligible alternative to the public switched telephone networks (PSTNs) in terms of cost, efficiency, quality, versatility and reliability. Concurrently, the deployment of Wireless Local Area Networks (WLAN) has become more and more popular in buildings and corporate campuses. Therefore, Voice over Internet Protocol (VoIP) over a Wireless Local Area Network (WLAN) is becoming one of the most growing and important internet application technologies in recent years.

1.1 Purpose of the Thesis

The purpose of this thesis is to evaluate the performance of a wireline network deployed with both voice, video and data communication by using different queuing scehemes and different CODEC capabilities. Besides the effects of deploying a wireless network with a set of wireless clients to an existing wireline network will be discussed in terms of end-to-end delay, delay variation (jitter) and thus quality of service.

With fast deployment of wireless local area networks (WLANs), the ability of WLAN to support real time services with acceptable quality of service (QoS) requirements is becoming an increasing research interest. However, real-time communication over wireless networks has many limitations and challenges, since the network conditions are not stable in wireless networks compared to wireline systems. VoIP quality may vary over a wide range and is impacted by a lot of factors, such as voice codecs, network protocols, system transport capabilities, and so on. The transport performance of packet-switched systems can be evaluated in terms of packet loss, end-to-end delay, and jitter.

Impairness and dropping in packet transport are the main reasons for packet loss which affects the quality of voice. This degrades the voice quality; for example, clicks, muting, or unintelligible speech. Packet loss varies depending upon the speech CODEC. Higher compression ratios increase the susceptibility of packet loss. Packet loss should be less than 1 percent for an acceptable quality of voice communication.

End-to-end delay is another factor that affects voice quality because too much delay will make the packets drop. Several factors that influence end-to-end delay are processing delay, propagation delay and network transmission delay that will be further investigated later in the dissertation. Jitter is the delay difference between the slowest packet and the fastest packet, which will cause packet loss when the arriving intervals are longer than the time that the dejittering buffer can retain. In order to control jitter, buffer usage can be necessary; however, max delay of 150 ms for voice transmission must be guaranteed at the same time.

Voice over IP uses a set of protocols such as Session Initiation Protocol (SIP) and H.323 which makes voice communication established between the end users and that voice quality requirements is met such that in Public Switched Telephone Networks (PSTN). In this dissertation, we will further investigate the protocol stacks and the way of their working principles in Voice over IP. We will also provide a comparison for SIP and H.323 in terms of their qualitative aspects and quantitative aspects and also an explanation for the factors that account for the similarities and the differences between these two protocols.

1.2 Background

There is a tremendous demand on real-time multimedia delivery over wireless Internet due to the dramatic increase in wireless communication and the growth of the Internet. However, real-time multimedia over wireless Internet has many challenges. First of all, the inherent best-effort characteristic of packet-switched networks makes it difficult to provide guaranteed QoS for real-time multimedia delivery. Secondly, wireless channels have much higher packet-loss rate, bit-error rate, and channel instability compared to wired channels. Thirdly, the real-time communication demands strict time limitations on the network end-to-end delay and delay jitter.

Figure 1.1 shows a basic VoIP system, including PC-to-PC, PC-to-PSTN and wireless networks.

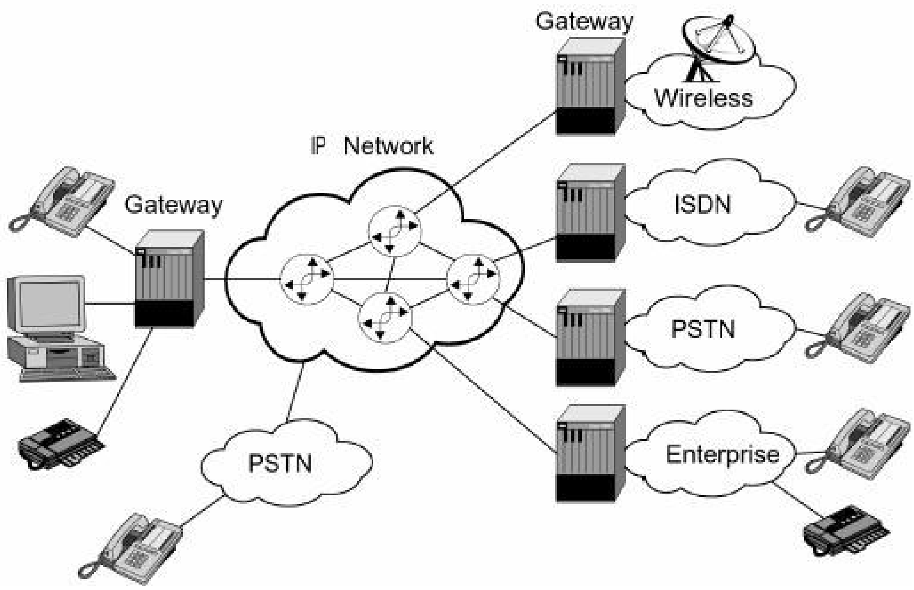


Figure 1.1 : Basic VoIP implementation

There are several works that aims to improve the voice quality such as PLC (Packet Loss Concealment) mechanism, jitter buffer management techniques, packet segmentation strategies and so on.

An adaptive PLC (Packet Loss Concealment) mechanism under different packet loss rates is proposed by Razvi Doomun (Razvi Doomun, 2007). This adaptive PLC scheme delivers an acceptable speech quality across varying packet loss rates. Doomun claims that it is possible to reconstruct lost packets by dynamically changing the packet concealment techniques. Packet repetition, noise insertion, pitch replication, and waveform substitution methods are used for the adaptive recovery mechanisms. According to the optimized system model, one can adaptively select the most adequate PLC technique which has the minimal computational complexity while maintaining the acceptable quality of service.

In wireless networks, packet loss and jitter (variation of the network delay) are the most challenging factors that has a direct effect on the voice quality. Packet loss causes voice clipping, skips and long end-to-end delay which have a significant impact on the quality of communication. A buffer in the receiving end always compensates for jitter. If the jitter exceeds the size of the device buffer, there will be buffer overflow and as a result, a packet that arrives later than the length of the jitter buffer will be lost in the transmission path. By improving the buffer management strategies, one could achieve better voice quality. Baratvand, Tabandeh, Behboodi, and Ahmadi (2008) propose a different jitter buffer playout scheme, including circular buffer and multi buffer. In multi buffer, the optimum buffer size is equal to size of packet payload because each buffer is assigned to one packet. In circular buffer, the optimum buffer size should be the two times of packet payload size. In circular buffer there is no mandatory constraint for size of buffer unlike Multibuffer, that buffer size should be integer multiple of packet payload.

Li, Yan and Xu (2008) proposed a novel link layer packet segmentation method to enhance the performance of VoIP which do not affect non-real-time data traffic much. They organize the link layer to support packet segmentation that allows large packets can be divided into small parts, so that decreases the queuing delay and defines real-time voice packets the highest priority in terms of transerring them first. To do so, they enhanced an exponential algorithm to decide when and how the packets should be segmented.

Hirannaiah, Jasti, and Pendse (2007) propose an algoritm to dinamically select the voice codecs in order to improve the performance of adative jitter buffer algorithm. The audio codecs are changed from a higher bit rate codec G.711 to a lower bit rate codec G.723.1 during an established call session, reducing the packet loss and improving the call performance.

1.3 Thesis Organization

This paper is organized as follows: In this Chapter, we present a brief introduction for VoIP and also the most common challenges of implementing voice communication into wireline or wireless networks were discussed. Common voice protocols, such as H.323, Session Initiation Protocol (SIP), Megaco and MGCP and video protocols such as H.261, H.263, H.264 will be described in Chapter 2 and Chapter 3, respectively. Chapter 4 covers the CODEC selection and factors affecting VoIP Quality of Service (QoS). The proposed scenarios and the simulation results with OPNET will be drawn in Chapter 5. Finally, conclusions will be discussed in Chapter 6.

2. VOICE TRANSMISSION TECHNIQUES: VOIP PROTOCOLS

Voice over IP (VoIP) uses the Internet Protocol (IP) to transmit voice as packets over an IP network. Using VoIP protocols, voice communications can be achieved on any IP network regardless whether it is Internet, Intranet or Local Area Networks (LAN). In a VoIP enabled network, the voice signal is digitized, compressed and converted to IP packets and then transmitted over the IP network. VoIP signaling protocols are used to set up and tear down calls, carry information required to locate users and negotiate capabilities. The key benefits of Internet telephony (Voice over IP) are the very low cost, the integration of data, voice and video on one network, the new services created on the converged network and simplified management of end user and terminals.

There are a few VoIP protocol stacks which are derived by various standard bodies and vendors, namely H.323, SIP, MEGACO and MGCP.

H.323 is the ITU-T's standard, which was originally developed for multimedia conferencing on LANs, but was later extended to cover Voice over IP. The standard encompasses both point to point communications and multipoint conferences. H.323 defines four logical components: Terminals, Gateways, Gatekeepers and Multipoint Control Units (MCUs). Terminals, gateways and MCUs are known as endpoints.

Session Initiation Protocol (SIP) is the IETF's standard for establishing VoIP connections. SIP is an application layer control protocol for creating, modifying and terminating sessions with one or more participants. The architecture of SIP is similar to that of HTTP (client-server protocol). Requests are generated by the client and sent to the server. The server processes the requests and then sends a response to the client. A request and the responses for that request make a transaction.

Media Gateway Control Protocol (MGCP), an IETF standard based on Cisco and Telcordia proposals, defines communication between call control elements (Call Agents or Media Gateway) and telephony gateways. MGCP is a control protocol, allowing a central coordinator to monitor events in IP phones and gateways and instruct them to send media to specific addresses. In the MGCP architecture, the call control intelligence is located outside the gateways and is handled by the call control elements (the Call Agent). Also, the call control elements (Call Agents) will synchronize with each other to send coherent commands to the gateways under their control.

The Media Gateway Control Protocol (Megaco) is a result of joint efforts of the IETF and the ITU-T (ITU-T Recommendation H.248). Megaco/H.248 is a protocol for the control of elements in a physically decomposed multimedia gateway, which enables separation of call control from media conversion. Megaco/H.248 addresses the relationship between the Media Gateway (MG), which converts circuit-switched voice to packet-based traffic, and the Media Gateway Controller, which dictates the service logic of that traffic. Megaco/H.248 instructs an MG to connect streams coming from outside a packet or cell data network onto a packet or cell stream such as the Real-Time Transport Protocol (RTP). Megaco/H.248 is essentially quite similar to MGCP from an architectural point of view and the controller-to-gateway relationship, but Megaco/H.248 supports a broader range of networks.

The SS7/C7 is the traditional signaling protocol for the circuit switched voice networks. To integrate the SS7/C7 network with the IP network, a group of protocols are defined, namely SIGTRAN (Signaling Transpor protocol). The key transport protocol in the SIGTRAN stack, the Stream Control Transmission Protocol (SCTP), has been applied in a much broader base after its creation.

Common signaling protocols used for voice communication is illustrated in Table 2.1.

Table 2.1: Common signaling protocols used for VoIP

Signaling	
ITU-T H.323	H.323: Packet-based multimedia communications (VoIP) architecture
	H.225: Call Signaling and RAS in H.323 VoIP Architecture
	H.235: Security for H.323 based systems and communications
	H.245: Control Protocol for Multimedia Communication
	T.120: Multipoint Data Conferencing Protocol Suite
IETF	Megaco / H.248: Media Gateway Control protocol
	MGCP: Media Gateway Control Protocol
	RTSP: Real Time Streaming Protocol
	SIP: Session Initiation Protocol
	SDP: Session Description Protocol
	SAP: Session Announcement Protocol
CableLab	NCS: Netowrk-based Call Signaling Protocol
Cisco Skinny	SCCP: Skinny Client Control Protocol
Media/CODEC	G.7xx: Audio (Voice) Compression Protocols (G.711, G.721, G.722, G.723, G.726, G.727. G.728, G.729)
	H.261: Video CODEC for Low Quality Videoconferencing
	H.263: Video CODEC for Medium Quality Videoconferencing
	H.264 / MPEG-4: Video CODEC for High Quality Video Streaming
	Video CODEC for Medium Quality VideoconferencingRTP: Real Time Transport Protocol
	RTCP: RTP Control Protocol
Others	COPS: Common Open Policy Service
	SIGTRAN: Signaling Transport protocol stack for SS7/C7 over IP
	SCTP: Stream Control Transmission Protocol
	TRIP: Telephony Routing Over IP

2.1 H.323 Protocol

2.1.1 Protocol description

H.323 is a widely deployed International Telecommunication Union (ITU) standard, originally established in 1996. It is part of the H.32x series of protocols and describes a mechanism for providing real-time multimedia communication (audio, video, and data) over an IP network.

H.323 is a suite of protocols that initially supported only videoconferencing systems over LAN/WAN (Local Area Networks/Wide Area Networks) topologies and protocols. H.323 has evolved into a packet-based signaling standart that provides a foundation for audio, video and data communications across IP-based networks, including the Internet. The second version of H.323 was released in 1998, H.323v2 changed the videoconferencing focus of H.323 to multimedia communications.

Functions of H.323 can be simply classified into four groups: call control, multimedia management, bandwidth management and interface between LANs and other H.323 non-compliant endpoints.

Components of an H.323 network include media-terminating devices such as phones, video conferencing terminals, gateways, and multipoint conferencing units (MCU, for hosting meetings). Devices in this group are categorized as endpoints in the H.323 network. Other components include gatekeepers and H.323 border elements. Gatekeepers provide services such as a network dial plan and bandwidth management for endpoints. The H.323 border element connects two H.323 networks to provide call routing and authorization between the networks. Table 2.2 shows the basic components of H.323 signaling stack. We will notice some familiar elements, such as CODECs and some new elements, such as H.225, H.235 and H.245.

Table 2.2: H.323 protocol stack

<table>
<tr><td>Audio CODECs</td><td>Video CODECs</td><td colspan="4">Terminal Control and Management</td><td>Data</td></tr>
<tr><td>G.711
G.722
G.723.1
G.728
G.729</td><td>H.261
H.263
H.264</td><td rowspan="2">RTCP</td><td rowspan="2">H.225
Terminal
to
Gateway
Signaling
(RAS)</td><td rowspan="2">H.235
Authentication,
Privacy and
Integrity</td><td rowspan="2">H.245
Control
Channel</td><td>T.120
Series

T.124</td></tr>
<tr><td colspan="2">RTP</td><td>T.125</td></tr>
<tr><td colspan="4">UDP</td><td colspan="2">TCP</td><td rowspan="3">T.123</td></tr>
<tr><td colspan="6">Network Layer (IP)</td></tr>
<tr><td colspan="6">Link Layer (IEEE 802.3)</td></tr>
<tr><td colspan="7">Physical Layer (IEEE 802.3)</td></tr>
</table>

The protocols in the H.323 protocol suite are:

Call control and signaling protocols are listed below:

- H.225.0: Call signaling protocols and media stream packetization (uses a subset of Q.931 signaling protocol). H.225 defines the call signaling method (direct or gatekeeper routed) during Registration, Admission and Status

(RAS) operations. H.225.0 provides a mechanism for initiating calls between devices.

- H.225.0/RAS: Registration, Admission and Status provides controls on bandwidth utilization and endpoint location.
- H.245: Control protocol for multimedia communication. H.245 negotiates CODECs to ensure conflicts are efficiently settled. H.245 provides a mechanism for negotiating media types and characteristics between endpoints.

Audio processing codecs are listed below:

Audio and video codecs provide the method for encoding and decoding media streams.

- G.711: Pulse code modulation of voice frequencies
- G.722: 7 kHz audio coding within 64 kb/s
- G.723.1: Dual rate speech coders for multimedia communication transmitting at 5.3 and 6.3 kb/s
- G.728: Coding of speech at 16 kb/s using low-delay code excited linear prediction
- G.729: Coding of speech at 8kps using conjugate-structure algebraic-code-excite linear-prediction

Video processing codecs are listed below:

- H.261: Video codecs for audiovisual services at Px64kps.
- H.263: Video coding for low bit rate communication.

Data conferencing protocols are listed below:

- T.120: This is a protocol suite for data transmission between end points. It can be used for various applications in the field of Collaboration Work, such as white-boarding, application sharing, and joint document management. T.120 utilizes layer architecture similar to OSI (Open Systems Interconnection) model. Top layer (T.126, T.127) are based on the services of layer layers (T.121, T.125).

Media transportation protocols are listed below:

RTP and RTCP stacks provide a mechanism for transporting and managing media packet data over an IP network.

- RTP: Real time Transport Protocol
- RTCP: RTP Control Protocol

Security protocols are listed below:

- H.235: Security and encryption for H.series multimedia terminals. H.235 provides authentication, privacy and integrity for H.323-based systems.

Table 2.3 shows a detailed description for H.323 protocol stack.

Table 2.3: A detailed description of H.323 protocol stack

Component	Description
Audio CODECs	G.711: 64 Kbps PCM G.722: 32 Kbps SB-ADPCM G.723.1: 6.3 Kbps MPMLQ/5.38 Kbps CS-ACELP G.728: 16 Kbps LD-CELP G.729A/B: 8 Kbps CS-ACELP
Video CODECs	H.261 describes a video stream for transport using the Real Time Transport Protocol (RTP) with any of the underlying protocols that carry RTP. H.263 specifies the payload format for encapsulating an H.263 bit stream in RTP.
Network Layer	Internet Protocol (IP) is a connectionless transport protocol. IP operates in Layer 3 Networking in the OSI model.
Link Layer	802.3 is an IEEE standard that defines the physical media and the working characteristics of Ethernet. Ethernet is the widely-installed local area network (LAN) technology.
Data	The T.120 standard contains a series of communication and application protocols and services that provide support for real time, multipoint data communications.

Table 2.3: A detailed description of H.323 protocol stack (continued)

Session and Transport Layer	Real Time Protocol (RTP) operates independantly of the transport layer and is intended as a framework, not a separate layer. User Datagram Protocol (UDP) is a connectionless protocol that runs on top of the IP networks. Within networking, selected UDP port numbers (well-known ports) are reserved for frequently used, higher level processes. Regardless of which customer network or which endpoint, these ports typically remain the same. Some examples are: - 25: SMTP (Simple Mail Transfer Protocol) - 80: Web - 110: POP - 1718: Gatekeeper Discovery - 1719: Registrationn with the Gatekeeper - 1720: H.225 Destination Transmission Control Protocol (TCP) is a highly-reliable and connection-oriented Layer 4 protocol that is used when data integrity is more important than the transmission time.
Terminal Control and Management	Real-time Transport Control Protocol (RTCP) is used by the RTP to control and synchronize streaming audio and video. RTCP provides feedback information to the source that can be used to adapt the flow to changing network conditions. H.225 governs H.225 session establishment and packetization. H.225 defines the call signaling method (direct or gatekeeper routed) during Registration, Admission and Status (RAS) operations. H.225 establishes the first connection, typically to port 1720, and uses the Q.931 signaling method. Q.931 defines and specifies call signaling and call setup acknowledgements and requirements. H.225 also uses Fast Connect or Fast Call Setup, which carries the H.245 signaling information within the H.225 message. H.245 negotiates audio and video CODECs to ensure conflicts are efficiently settled. H.245 also transmits DTMF (Dual Tone Multi Frequency) codes, lamp indicator control and other control signaling information required by an H.323 device, in addition to opening and closing media channels. H.235 provides authentication, privacy and integrity for H.323 based systems. H.235 applies to both simple point-to-point and multipoint conferences for any terminals that use H.245 as a control protocol.

2.1.2 Major components of H.323

H.323 is a flexible standard that can be applied to voice-only handsets and full multimedia video-conferencing stations. H.323 has broad industry support, which has established H.323 as the standard for audio and video communications. Figure 2.1 shows a general overview for major components of H.323 protocol stack.

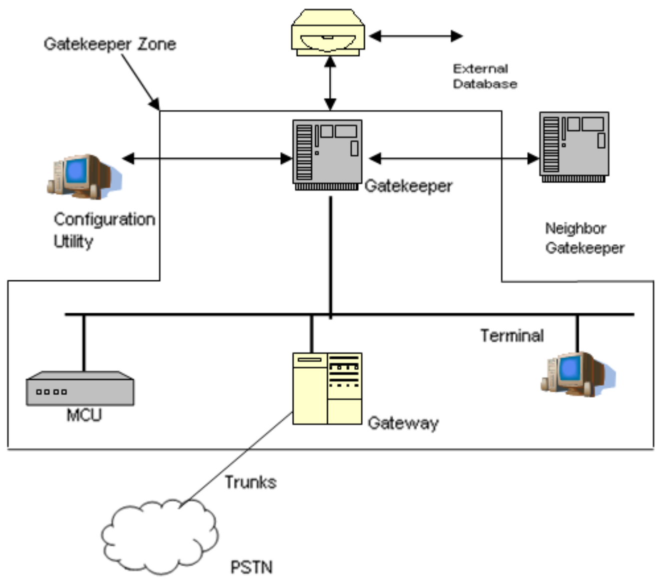


Figure 2.1 : Major components of H.323

H.323 consists of the key components listed below:

- H.323 gateways: Acts as a bridge to the IP network. It also maps the destination telephone number to the destination gateway IP address.
- Gatekeepers: Manages a gatekeeper zone. A gatekeeper zone consists of terminals, gateways, and MCUs managed by a single gatekeeper.
- Multipoint Control Units (MCUs): Accept the audio stream from the conferees, encode it into a comman signaling format, and regenerate the signal back to the participants.
- IP Terminals and Clients: Includes endpoints on the network, such hard or soft telephones (portable or stationary) and wireless devices (802.11a and 802.11b, as well as 802.11g). These devices bring voice and data communications to the end user. A call server completes the call processing.

H.323 Gateways transform audio received from a telephone device or telecommunications system into a format that the data network can use. Generally they have built-in intelligence to select the CODECs and adjust the protocols and timing between dissimilar computer systems or voice over data networks. H.323 Gateways also connect divergent networks by translating signaling protocols and converting media formats. They are optional in an H.323 conference.

Gatekeepers provide call control, call routing, address translation, media access and bandwidth management. They map destination telephone numbers to destination endpoint IP addresses, monitor and control the threshold for the number of simultaneous conversations on the network. H.323 Gatekeepers provide a centralized service to which H.323 devices register and map LAN aliases to an IP address and provide address lookups. They also handle the dialing plan and provide accounting, billing and charge back capabilities. Without gatekeepers, each device must be manually configured for inter-device communications.

The required and optional gatekeeper functions are listed in Table 2.4 and Table 2.5.

Table 2.4: Required gatekeeper functions

Function	Description
Address Translation	Provides for translation of alias address, such as URL (Uniform Resource Locator) , to transport address, such as IP; uses registration and update messages to build translation tables.
Admission Control	Provides LAN access authorization, using Admission Request, Confirmation and Reject (ARQ/ARC/ARJ) messages. Access can be based on call authorization, bandwidth or other criteria; Admissions Control can also be a null function admitting all requests.
Bandwidth Control	Supports Bandwidth Request, Confirmation and Reject (BRQ/BCF/BRJ) messages, as well as a null function accepting all bandwidth change requests.
Zone Management	Provides the functions above for teminals, MCUs and Gateways registered within its management zone.
Registration, Admission and Status (RAS)	H.225.0/RAS (Registration, Admission and Status) 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.

Table 2.5: Optional gatekeeper functions

Function	Description
Call Control Signaling	Can process Q.931 call control signals in point-to-point conference and can send Q.931 signals directly to endpoints.
Call Authorization	Can reject a call from a terminal based on Q.931; however, criteria for authorization pass or fail is outside the capability of H.323.
Bandwidth Management	Can reject a call if insufficient bandwidth or if terminal requests additional bandwidth; however, criteria for determining available bandwidth is outside capability of H.323.
Call Management	Can maintain a list of ongoing H.323 calls to determine whether a called terminal is busy or to provide information for Bandwidth Management.

MCUs (Multi-point Control Units) consist of Multipoint Controller and Multipoint Processor. Multipoint Controller (MC) performs H.245 negotiations between devices to determine common audio and video processing capabilities. Multipoint Processors (MPS) encodes and routes audio, video and data streams between device endpoints. The MC controls the MPS. There can be zero or more Multipoint Processors. The Multipoint Processors can co-exist on same device with gatekeeper software.

IP Terminals and Clients are considered client endpoints on the network. IP terminals and clients must support voice communications, while video and data support is optional.

2.1.3 H.323 call setup messages

While transporting information, H.323 proceeds in phases from endpoint to endpoint. Various messages are transmitted during the process. Some of these messages are ARQ (Admission Request Message), ACF (Admission Confirmation Message) and ARJ (Admission Reject Message).

In H.323 call setup scenarios, different scenarios may occur, based on whether a Gatekeeper is involved. Figure 2.2 shows the H.323 call setup with a Gatekeeper.

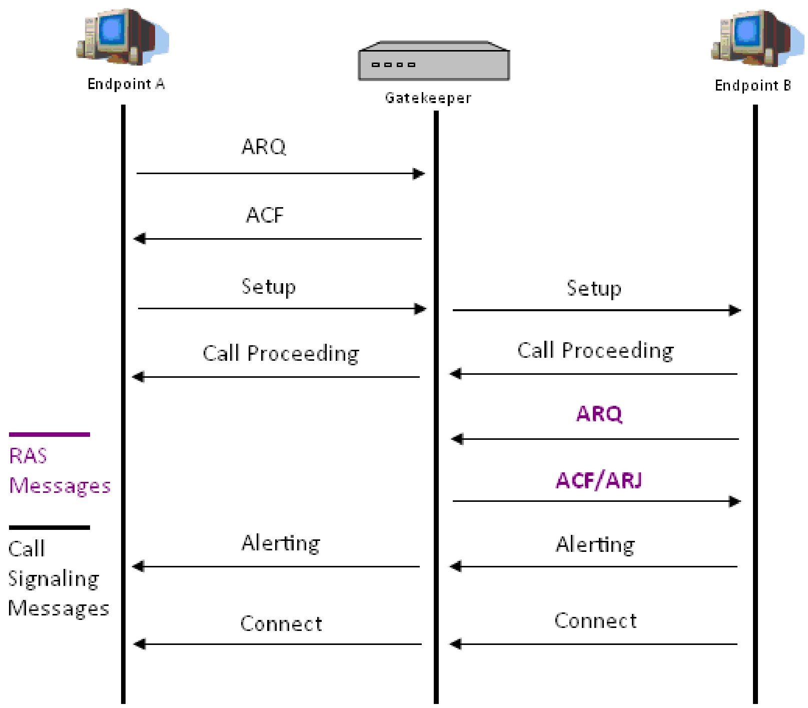


Figure 2.2 : H.323 call setup

- Endpoint A sends a RAS ARQ message on the RAS channel to the Gatekeeper for registration and requests the use of direct call signaling.
- The Gatekeeper sends an ACF message to Endpoint A to confirm admission and indicates that Endpoint A can use direct call signaling.
- Endpoint A sends a call signaling Setup message to Endpoint B to request a connection.
- Endpoint B responds with a Call Proceeding message to Endpoint A.
- Endpoint B registers with the Gatekeeper and sends a RAS ARQ message to the Gatekeeper on the RAS channel.
- The Gatekeeper sends a RAS ACF message to Endpoint B to confirm the registration.
- Endpoint B alerts Endpoint A of the connection establishment by sending an Alerting message.
- Endpoint B sends a Connect message to Endpoint A to confirm the establishment of the connection and the call is established.

2.2 Megaco/H.248: Media Gateway Control Protocol

2.2.1 Protocol description

Control of elements in a physically decomposed multimedia gateway, enabling the separation of call control from media conversion. The Media Gateway Control Protocol (Megaco) is a result of joint efforts of the IETF and the ITU-T Study Group 16. Therefore, the IETF defined Megaco is the same as ITU-T Recommendation H.248.

Megaco is the official international standard for decomposed gateway acrhitectures. In addition, MEGACO is the successor to MGCP. Several of the functions of MGCP were incorporated into MEGACO.

MEGACO is an ASCII-based protocol and considered as similar to HTTP. One of the benefits of H.248 over H.323 is that H.248 gathers information from servers, instead of gateways. This allows for lower-cost gateways with more features.

Media Gateway Controllers in a network that uses MEGACO signaling can also be used while SIP is being used in the same network.

Megaco/H.248 addresses the relationship between the Media Gateway (MG), which converts circuit-switched voice to packet-based traffic, and the Media Gateway Controller (MGC, sometimes called a call agent or softswitch, which dictates the service logic of that traffic). Megaco/H.248 instructs an MG to connect streams coming from outside a packet or cell data network onto a packet or cell stream such as the Real-Time Transport Protocol (RTP). Megaco/H.248 is essentially quite similar to MGCP from an architectural point of view and the controller- to-gateway relationship, but Megaco/H.248 supports a broader range of networks, such as ATM.

There are two basic components in Megaco/H.248: terminations and contexts. Terminations represent streams entering or leaving the MG (for example, analog telephone lines, RTP streams, or MP3 (MPEG-1 Audio Layer III) streams). Terminations have properties, such as the maximum size of a jitter buffer, which can be inspected and modified by the MGC.

Terminations may be placed into contexts, which are defined as occuring when two or more termination streams are mixed and connected together. The normal, "active" context might have a physical termination (say, one DS0 in a DS3) and one ephemeral one (the RTP stream connecting the gateway to the network). Contexts are created and released by the MG under command of the MGC. A context is created by adding the first termination, and is released by removing (subtracting) the last termination.

A termination may have more than one stream, and therefore a context may be a multistream context. Audio, video, and data streams may exist in a context among several terminations.

2.2.2 Major components of MEGACO

The key to MEGACO and its functionality is the use of its components, which are listed later in this chapter.

Media Gateway Controller (MGC) provides a central point of intelligence for the Media Gateways and controls one or more MGs. MGC communicates to Media Gateway and Signaling Gateway via TCP/IP.

Media Gateway (MG) controls and processes media streams between networks and functions primarily as a slave to execute commands from the MGC.

Signaling Gateway (SG) provides interoperability between the legacy SS7 and the newly-defined Stream Control Transmission Protocol (SCTP). SG acts as signaling server created for the PSTN.

Figure 2.3 illustrates a diagram for major components of MEGACO.

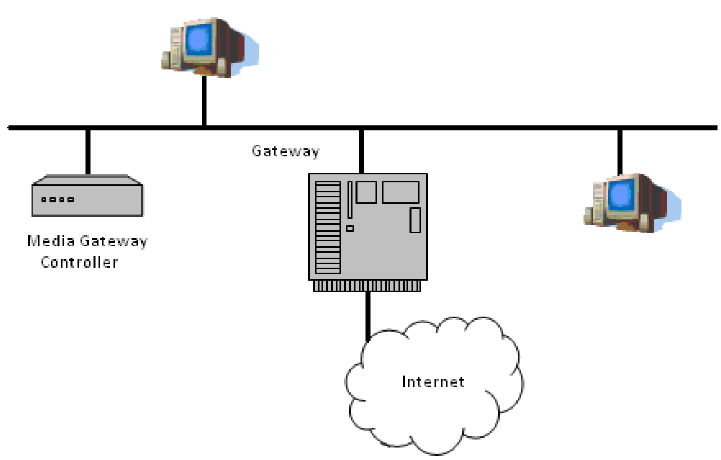


Figure 2.3 : Major components of MEGACO

2.2.3 Megaco call setup messages

All Megaco/H.248 messages are in the format of ASN.1 text messages. Megaco/H.248 uses a series of commands to manipulate terminations, contexts, events, and signals. The following is a list of the commands:

- Add -- The Add command adds a termination to a context. The Add command on the first Termination in a Context is used to create a Context.
- Modify -- The Modify command modifies the properties, events and signals of a termination.
- Subtract -- The Subtract command disconnects a Termination from its Context and returns statistics on the Termination's participation in the Context. The Subtract command on the last Termination in a Context deletes the Context.
- Move -- The Move command automatically moves a Termination to another context.
- AuditValue -- The AuditValue command returns the current state of properties, events, signals and statistics of Terminations.
- AuditCapabilities -- The AuditCapabilities command returns all the possible values for Termination properties, events and signals allowed by the Media Gateway.
- Notify -- The Notify command allows the Media Gateway to inform the Media Gateway Controller of the occurrence of events in the Media Gateway.
- ServiceChange -- The ServiceChange Command allows the Media Gateway to notify the Media Gateway Controller that a Termination or group of Terminations is about to be taken out of service or has just been returned to service. ServiceChange is also used by the MG to announce its availability to an MGC (registration), and to notify the MGC of impending or completed restart of the MG. The MGC may announce a handover to the MG by sending it ServiceChange command. The MGC may also use ServiceChange to instruct the MG to take a Termination or group of Terminations in or out of service.

Figure 2.4, 2.5 and 2.6 shows an example of tpycal Megaco messages, such as Add, Modify and Subtract, respectively.

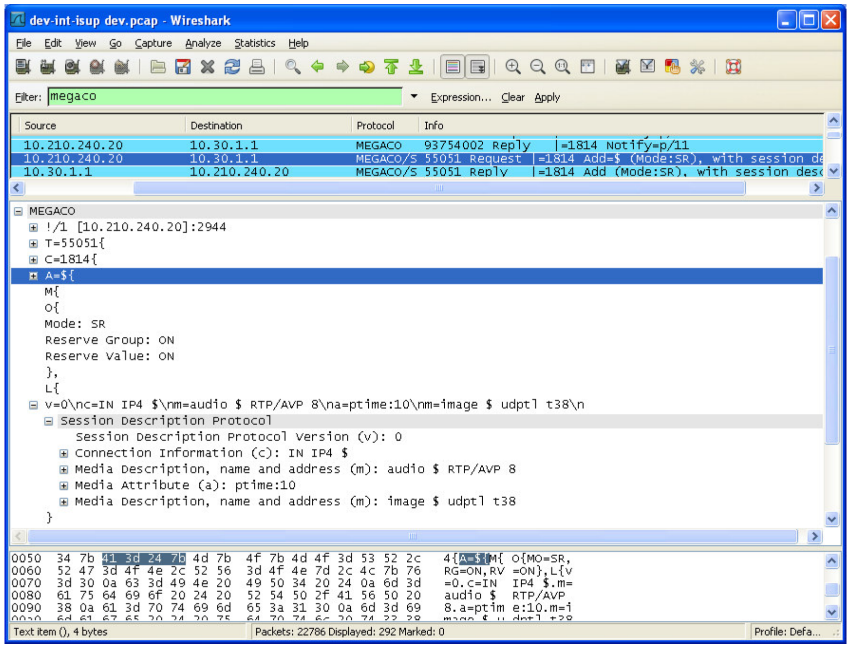


Figure 2.4 : Megaco Add (A) message

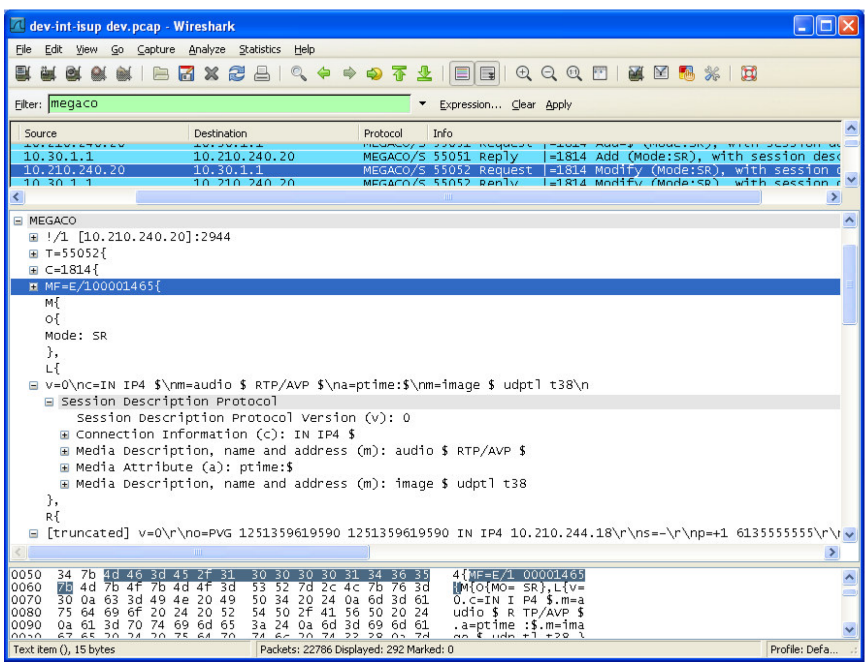


Figure 2.5 : Megaco Modify (MF) message

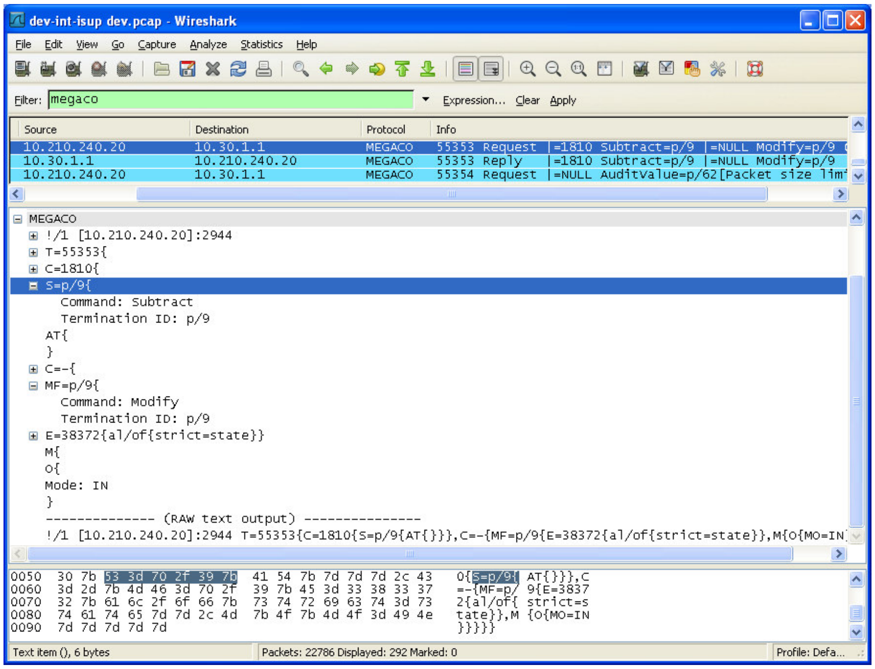


Figure 2.6 : Megaco Subtract (S) message

All of these commands are sent from the MGC to the MG, although ServiceChange can also be sent by the MG. The Notify command, with which the MG informs the MGC that one of the events the MGC was interested in has occurred, is sent by the MG to the MGC.

Figure 2.7 and 2.8 shows a diagram of basic call flow for a call setup and call teardown, respectively.

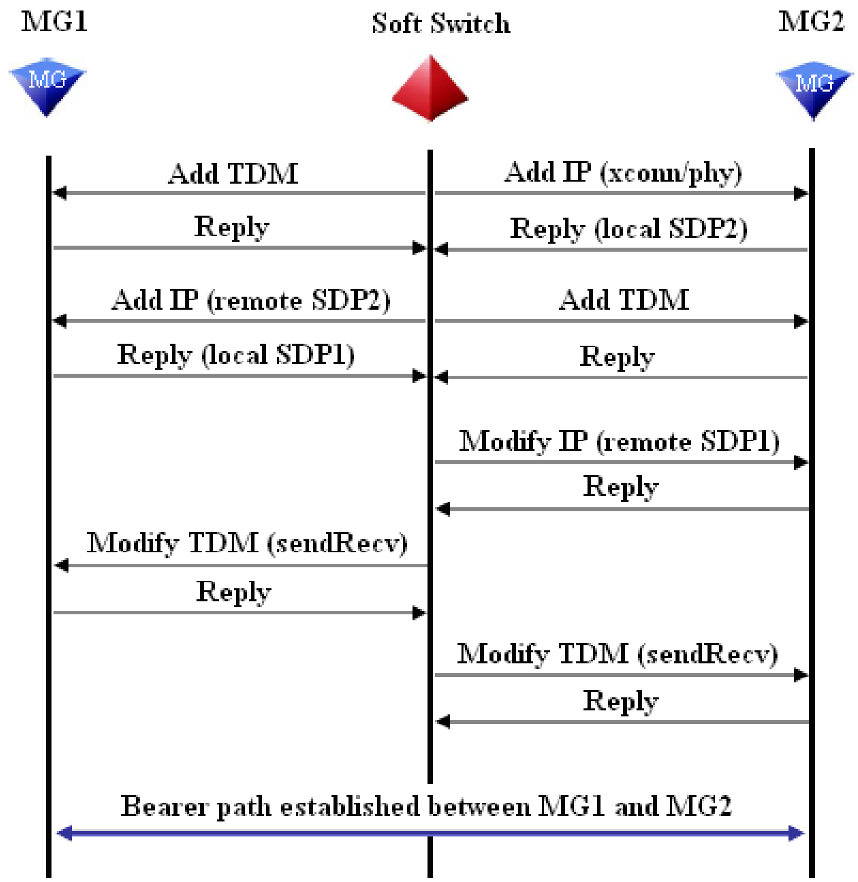


Figure 2.7 : Basic call flow for a call setup

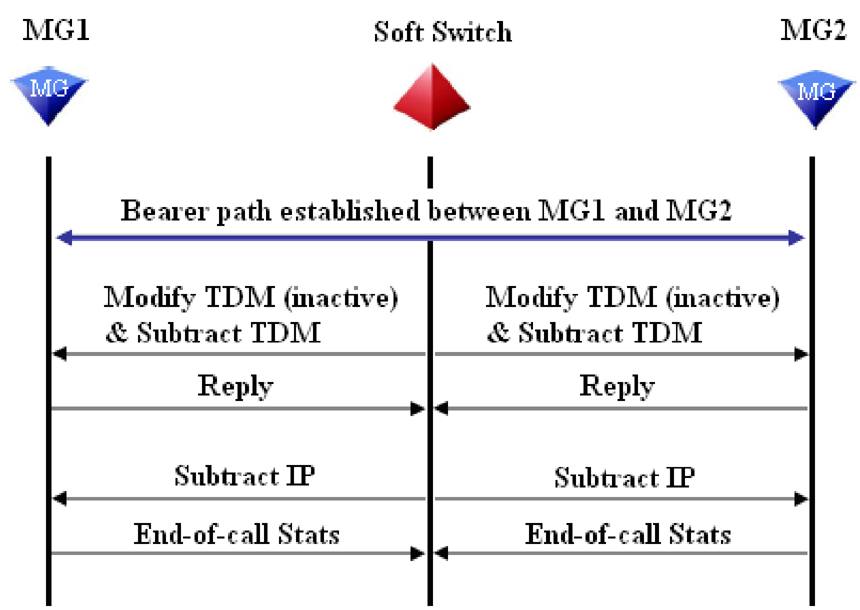


Figure 2.8 : Basic call flow for a call teardown

2.3 MGCP (Media Gateway Control Protocol)

2.3.1 Protocol description

Media Gateway Control Protocol (MGCP) is a VoIP protocol used between elements of a decomposed multimedia gateway which consists of a Call Agent, containing the call control "intelligence", and a media gateway containing the media functions, e.g., conversion from TDM voice to Voice over IP.

Media gateways contain endpoints on which the Call Agent can create, modify and delete connections in order to establish and control media sessions with other multimedia endpoints. A media gateway is typically a network element that provides conversion between the audio signals carried on telephone circuits and data packets carried over the Internet or over other packet networks. The Call Agent can instruct the endpoints to detect certain events and generate signals. The endpoints automatically communicate changes in service state to the Call Agent. Furthermore, the Call Agent can audit endpoints as well as the connections on endpoints.

MGCP assumes a call control architecture where the call control "intelligence" is outside the gateways and handled by Call Agents. It assumes that Call Agents will synchronize with each other to send coherent commands and responses to the gateways under their control. MGCP does not define a mechanism for synchronizing Call Agents. MGCP is, in essence, a master/slave protocol, where the gateways are expected to execute commands sent by the Call Agents.

MGCP assumes a connection model where the basic constructs are endpoints and connections. Endpoints are sources and/or sinks of data and can be physical or virtual. Creation of physical endpoints requires hardware installation, while creation of virtual endpoints can be done by software. Connections may be either point to point or multipoint. A point to point connection is an association between two endpoints with the purpose of transmitting data between these endpoints. Once this association is established for both endpoints, data transfer between these endpoints can take place. A multipoint connection is established by connecting the endpoint to a multipoint session. Connections can be established over several types of bearer networks.

In the MGCP model, the gateways focus on the audio signal translation function, while the Call Agent handles the call signaling and call processing functions. As a consequence, the Call Agent implements the "signaling" layers of the H.323 standard, and presents itself as an "H.323 Gatekeeper" or as one or more "H.323 Endpoints" to the H.323 systems.

MGCP is actually is a merger of the Internet Protocol Device Control (IPDC) and Signal Gateway Control Protocol (SGCP). IPDC is a suite of protocols that can, individually or together, perform connection control, media control and signaling transports between the circuit-switched network and the Internet. IPDC was developed by a consortium formed by Level 3 Communications. SGCP is a UDP-based protocol designed to address the concept of a network that combined voice and data on a single packet-switched IP network operating at a low level. SGCP was developed by Bellcore, now called Telcordia Technologies, and Cisco Systems.

The MGCP is a text based protocol. The transactions are composed of a command and a mandatory response. There are eight types of commands which is listed in Table 2.6.

Table 2.6: Types of MGCP commands

MGC --> MG	CreateConnection: Creates a connection between two endpoints; uses SDP to define the receive capabilities of the participating endpoints.
MGC --> MG	ModifyConnection: Modifies the properties of a connection; has nearly the same parameters as the CreateConnection command.
MGC <--> MG	DeleteConnection: Terminates a connection and collects statistics on the execution of the connection.
MGC --> MG	NotificationRequest: Requests the media gateway to send notifications on the occurrence of specified events in an endpoint.
MGC <-- MG	Notify: Informs the media gateway controller when observed events occur.
MGC --> MG	AuditEndpoint: Determines the status of an endpoint.
MGC --> MG	AuditConnection: Retrieves the parameters related to a connection.
MGC <-- MG	RestartInProgress: Signals that an endpoint or group of endpoints is taken in or out of service.

2.3.2 Major components of MGCP

Major Components of MGCP resemble IETF SIP in its functionality. MGCP consists of two key components: Call Agent and Gateways.

Call Agent is responsible from signaling, call processing and the Gateway. It is a web-based protocol and a software-based program known as a Gateway Controller.

Gateways are the network elements that provide an audio signal to the data packet conversion, which is transmitted on telephony circuitry, the Internet, or other packet networks. Endpoint-to-packet network or endpoint-to-endpoint connection occurs in the same gateway.

2.4 SIP (Session Initiation Protocol)

Session Initiation Protocol (SIP) is an alternative to H.323 that was initially created for the distribution of multimedia content. Unlike H.323, which specifies a complete, vertically integrated system, SIP supports a variety of architectures and protocols. SIP uses a media-description language and is modeled after the Hypertext Transfer Protocol, or HTTP, which operates in the application layer of the Open Systems Interconnection (OSI) communications model. SIP is based on a request-response model or INVITE to a Uniform Resource Locator (URL), or Internet address, for establishment of a session.

2.4.1 Protocol description

Session Initiation Protocol (SIP) is an application-layer control protocol that can establish, modify, and terminate multimedia sessions such as Internet telephony calls. SIP can also invite participants to already existing sessions, such as multicast conferences. Media can be added to (and removed from) an existing session. SIP transparently supports name mapping and redirection services, which supports personal mobility - users can maintain a single externally visible identifier regardless of their network location.

SIP supports five facets of establishing and terminating multimedia communications:

- User location: determination of the end system to be used for communication;
- User availability: determination of the willingness of the called party to engage in communications;
- User capabilities: determination of the media and media parameters to be used;
- Session setup: "ringing", establishment of session parameters at both called and calling party;

- Session management: including transfer and termination of sessions, modifying session parameters, and invoking services.

SIP is a component that can be used with other IETF protocols to build a complete multimedia architecture, such as the Real-time Transport Protocol (RTP) for transporting real-time data and providing QoS feedback, the Real-Time Streaming Protocol (RTSP) for controlling delivery of streaming media, the Media Gateway Control Protocol (MEGACO) for controlling gateways to the Public Switched Telephone Network (PSTN), and the Session Description Protocol (SDP) for describing multimedia sessions. Therefore, SIP should be used in conjunction with other protocols in order to provide complete services to the users. However, the basic functionality and operation of SIP does not depend on any of these protocols.

SIP provides a suite of security services, which include denial of- service prevention, authentication (both user-to-user and proxy-to-user), integrity protection, and encryption and privacy services.

SIP works with both IPv4 and IPv6. For Internet telephony sessions, SIP works as follows: Callers and callees are identified by SIP addresses. When making a SIP call, a caller first locates the appropriate server and then sends a SIP request. The most common SIP operation is the invitation. Instead of directly reaching the intended callee, a SIP request may be redirected or may trigger a chain of new SIP requests by proxies. Users can register their location(s) with SIP servers. SIP addresses (URLs) can be embedded in Web pages as shown in Figure 2.9 and therefore can be integrated as part of such powerful implementations as Click to talk.

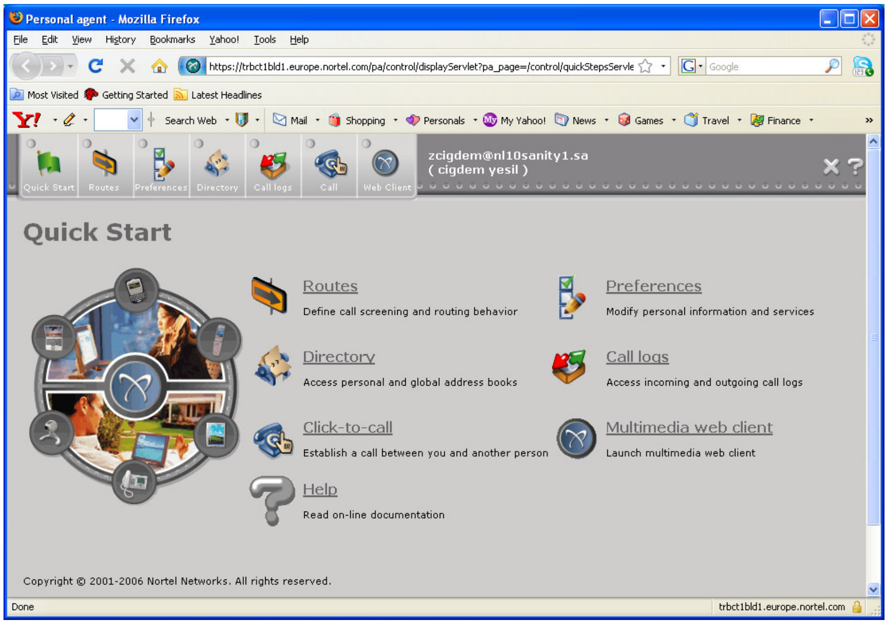


Figure 2.9 : Personal agent webpage

To facilitate the interconnection of the PSTN with IP, SIP-T (SIP for telephones) is defined by IETF in RFC 3372. SIP-T allows traditional IN-type services to be seamlessly handled in the Internet environment. It is essential that SS7 information be available at the points of PSTN interconnection to ensure transparency of features not otherwise supported in SIP. SS7 information should be available in its entirety and without any loss to the SIP network across the PSTN-IP interface. SIPT defines SIP functions that map to ISUP (ISDN User Part) interconnection requirements.

SIP messages can be transmitted either over TCP or UDP. SIP messages are text-based and use the ISO (International Organization for Standardization) 10646 character set in UTF-8 (Unicode Transformation Format-8bit) encoding. Lines must be terminated with CRLF (Carriage Return and Line Feed). Much of the message syntax and header field are similar to HTTP. Messages can be request messages or response messages. A request message has the format in Figure 2.10:

Method	Request URI	SIP version

Figure 2.10 : Request message format.

- Method -- The method to be performed on the resource. Possible methods are Invite, Ack, Options, Bye, Cancel, and Register.
- Request-URI -- A SIP URL or a general Uniform Resource Identifier; this is the user or service to which this request is being addressed.

- SIP version -- The SIP version being used.

The request line can be seen as “Request-Line: INVITE sip:1571@bskyb.com SIP/2.0” in Figure 2.11.

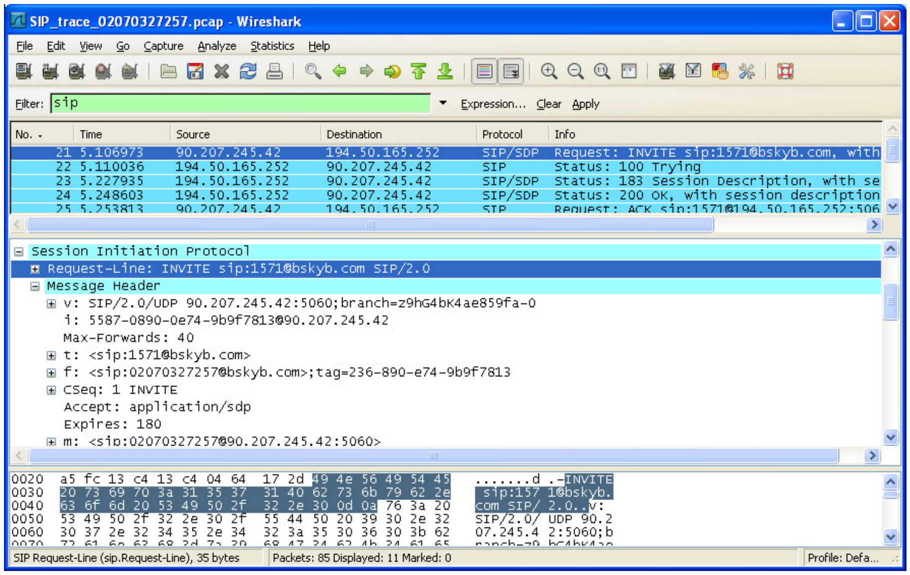


Figure 2.11 : Format of a request message

The format of the Response message header is shown in the Figure 2.12:

SIP version	Status code	Reason phrase

Figure 2.12 : Response message header format.

- SIP version -- The SIP version being used.
- Status-code -- A 3-digit integer result code of the attempt to understand and satisfy the request.
- Reason-phrase -- A textual description of the status code.

The status line can be seen as “Status-Line: SIP/2.0 100 Trying” in Figure 2.13.

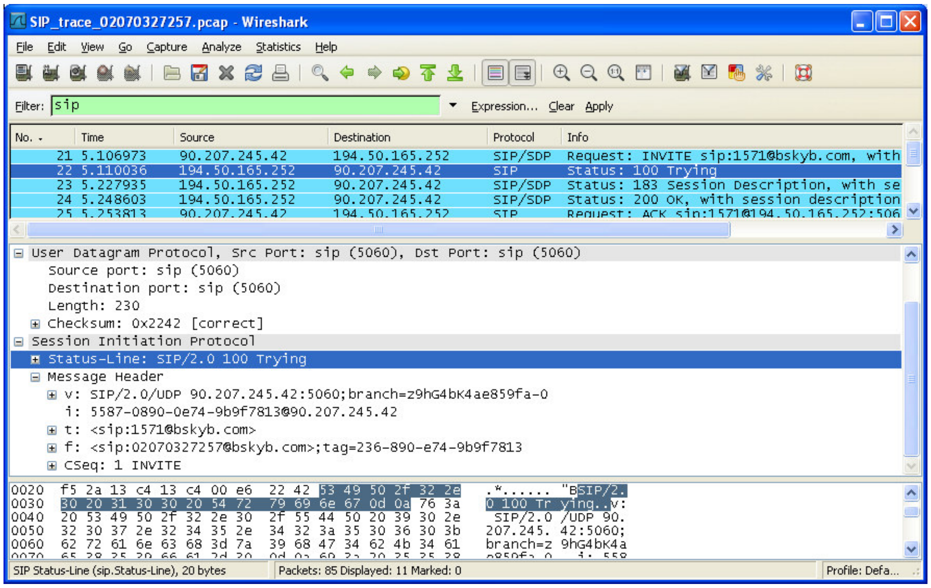


Figure 2.13 : Format of a status message

Figure 2.14 shows a graphical analysis of a basic SIP call.

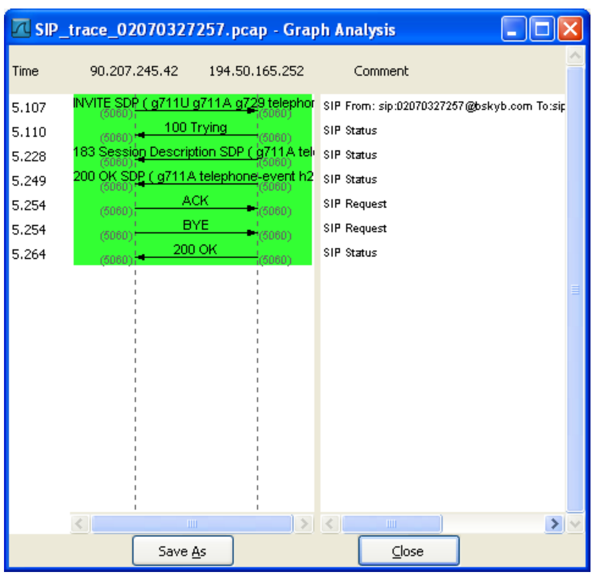


Figure 2.14 : A graphical analysis of a basic SIP call

2.4.2 Benefits of SIP

Benefits of SIP include those listed below:

- Integration: Works well with various web technologies using URLs, supporting MIME (Multipurpose Internet Mail Extensions), and carrying images, MP3s, and Java applets, in addition to using email routing mechanisms.

- Scalability: Supports several types of proxy servers.
- Extensibility: Consists of several mechanisms for extension of protocol.
- Flexibility: Does not dictate architecture, usage patterns, or deployment scenarios, but rather provides a framework within which it operates.

  - Relies on other protocols and techniques to provide QoS.

  - Remains totally independent of the voice path.

  - Leverages separate protocols that can be used without making any core protocol changes.

- Mobility:

  - Supports personal mobility, letting networks identify end users regardless of location on the network.

  - Gives end users the ability to originate and receive calls and access services on any network device, such as PC, laptop, or IP phone, regardless of location.

2.4.3 SIP protocols

The protocols that SIP can use to establish a session are listed below:

- Session Description Protocol (SDP)
- Session Announcement Protocol (SAP)
- Real-Time Streaming Protocol (RTSP)

SDP (Session Description Protocol) is session description protocol for multimedia sessions. SDP began as a component of the Session Announcement Protocol (SAP) but can be used with RTSP, SIP, as well as a standalone format for describing multicast sessions.

SDP is not a transport protocol. Instead, SDP is a simple, text-based format used to convey information to SIP entities to let them join and participate in the session; for example: Purpose of the session, Name of session, Time of the session, Media type pertaining to the session, such as video and audio, Formatted information for the video or audio session, and Pertinent IP addresses and port numbers for the session.

Figure 2.15 presents an example of an SDP message during a negotiation. Basically, in an SDP message, the pertinent IP address and port numbers for audio and/or video (image) communication are stated and supported CODECs are negotiated between the endpoints.

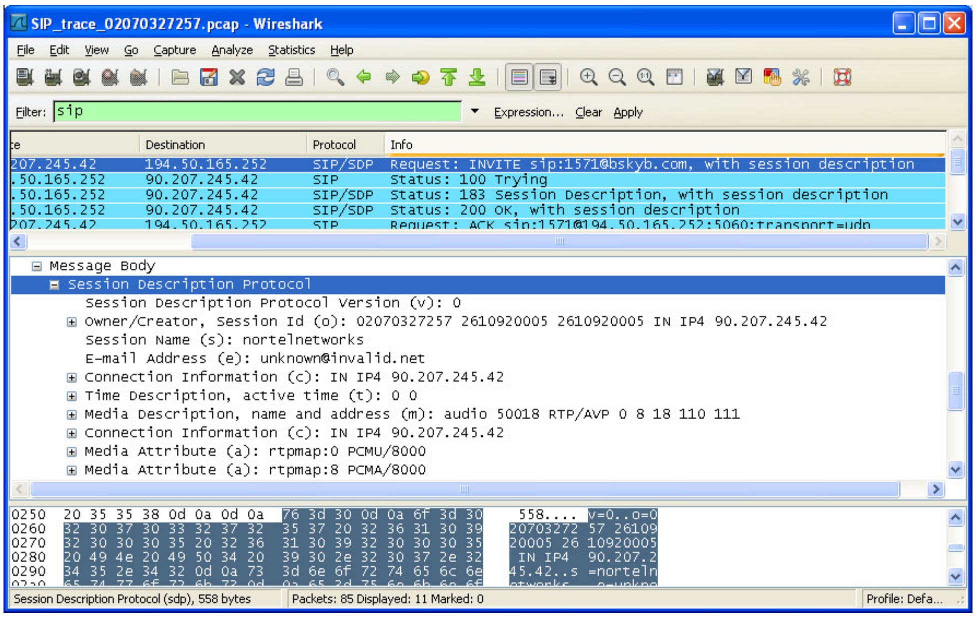


Figure 2.15 : Content of an SDP message

SAP (Session Announcement Protocol) delivers SDP packets using multicasting where participants are not known in advance. A multicast session is announced by sending multicast packets to a well-known multicast group carrying an SDP description of the session to occur.

Session Announcement Protocol is responsible of creating, modifying, and terminating sessions. It contains SDP as the payload and compares to a television schedule. SAP announces a conference session by periodically multicasting an announcement packet to a well-known multicast address and port.

RTSP (Real-Time Streaming Protocol) supports the control of delivered multimedia stream to include pause, fast forward, reverse, and absolute positioning within the media stream, recording, and device control.

2.4.4 Major components of SIP

SIP is based on a Request/Response or Invite model. The major components of SIP are listed below:

- User Agent: User Agent Client (UAC) and User Agent Server (UAS)
- Server: Proxy Server, Redirect Server, Registrar Server

Figure 2.16 shows a schema of SIP major components.

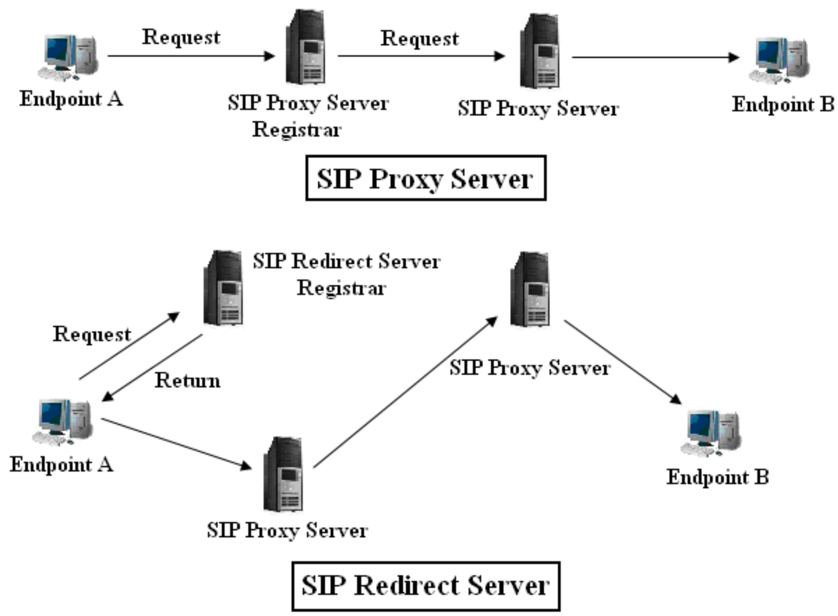


Figure 2.16 : Major components of SIP

The User Agent Server (UAS) contacts the user when a SIP request is received and returns an appropriate response on behalf of the user. The User Agent (UA) contains both a User Agent Client (UAC) and User Agent Server (UAS).

Server accepts request messages from the client and forwards accordingly; and sends back responses to the client's requests.

- Proxy Server: Proxy Server functions as a server and a client to make requests on behalf of other clients. It supports requests locally or forwards on to other servers. Proxy server interprets and can rewrite a SIP request message prior to forwarding to another server or to a user agent.
- Redirect Server: Redirect Server accepts a SIP request, maps the address in the request to a new address, and returns the appropriate message to the client. It does not forward SIP requests to other servers and does not accept calls.

- Registrar: Registrar accepts REGISTER requests and is typically co-located with a proxy or redirect server; can offer location services. It registers SIP parties in a SIP domain that is within the administrative entity for a SIP provider.

2.4.5 Session management (SIP messages)

The purpose of session management is to control the attributes of the end-to-end call. Various acknowledgements and responses occur during a SIP Request between User Agent (UA), SIP Proxy Servers, SIP Redirect Servers, Registrars, and User Agent Servers (UAS).

It is important to know the SIP Requests, in addition to some of the key SIP Response Type, that occur during the call setup process.

Examples of SIP Requests are listed below:

- INVITE: Starts a session.
- ACK: Confirms that the User Agent (UA) has received a final response to an INVITE.
- BYE: UA uses BYE to tell the server to release the call.
- CANCEL: Cancels a pending request, but not the completed request.
- OPTIONS: Only requests information about capabilities.
- REGISTER: Provides the user’s location to a SIP server.

Examples of SIP Responses are listed below:

- 1xx Trying
- 180 Ringing
- 2xx Successful Connection
- 3xx Redirection

305 Use Proxy

- 4xx Request Failure

407 Proxy Authentication Required

415 Unsupported Media Type

483 Too Many Hops

484 Address Incomplete

486 Busy Here

- 5xx Server Failure

502 Bad Gateway

504 Gateway Time-out

505 Version Not Supported

- 6xx Global Failures

600 Busy Everywhere

603 Decline

A SIP INVITE is required to set up a Media session. The information contained in the INVITE is different based on the type of Media session, message body length, and various other data. An example of the INVITE sent by msafin@atp.com to arod@wta.com is shown in Figure 2.17.

INVITE sip: arod@wta.com SIP/2.0	SIP URL of the called party
Via: SIP/2.0/UDP msafin.atp.com	Indicates path taken by the request
From: Marat <sip:msafin@atp.com>	Initiator of request
To: Andy <sip:arod@wta.com>	Recipient of request
Call-ID: 1113244434@atp.com	Uniquely IDs a particular invitation
Cseq: 1 INVITE	Command Sequence - Uniquely ID's a request within a Call-ID
Subject: Your tennis game	Nature of the call
Cont-Type: application/SDP	Indicates the Media type in message
V=0	Protocol version
0=Marat 76329381938 IN IP4 192.168.3.2	User, session ID, network type, address
S=marat203	Session name
C-IN IP4 192.168.3.2	Connection information
M-audio 5004 RTP/AVP 0 12 18	Media name, port address, CODEC options available from the client

Figure 2.17 : An example of the INVITE message

SIP operations follow a different set of procedures when inviting a participant to a session, as compared to an H.323 Call Setup. Sometimes, the INVITE has to be redirected and a User Agent located, as shown in Figure 2.18 below. This is one of the valuable advantages of this protocol.

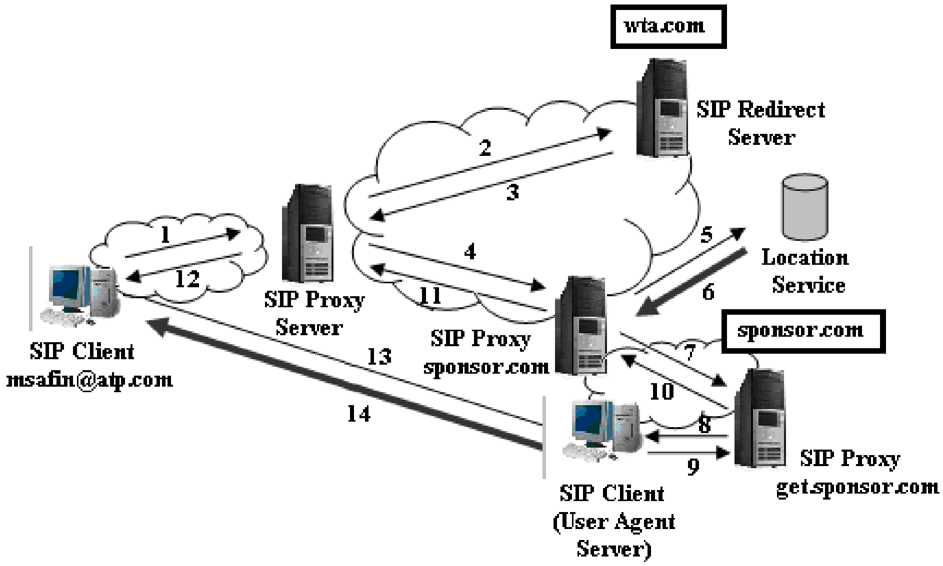


Figure 2.18 : INVITE redirection

Here is the explanation of INVITE Redirection step by step:

1) msafin@atp.com sends a SIP INVITE message to atp.com proxy to place a call to arod@wta.com.

2) atp proxy forwards request to wta.com server.

3) wta server determines arod is temporarily at sponsor.com. wta.com server sends redirect to atp.com proxy to try sponsor.com.

4) atp proxy sends INVITE to arod@sponsor.com.

5) sponsor.com server consults its database.

6) sponsor.com determines arod is located at get.sponsor.com.

7) sponsor.com sends new URL to arod@get.sponsor.com and sends INVITE to get.sponsor.com proxy.

8) get.sponsor proxy forwards INVITE to arod's PC after SIP used REGISTER action to ensure location of arod through an address binding procedure; arod responds to request.

9-10-11) Response is forwarded back through proxies to original caller.

12) ACK is sent.

13) Session is established and media is transmitted.

2.4.6 Comparing H.323 and SIP

Distinct differences between the H.323 and the SIP and how they support VoIP are shown in Figure 2.19.

Aspect	SIP	H.323
Origin	Networking community (IETF)	Telecom Community (ITU)
Architecture	Element	Stack
Encoding	HTTP model	ASN.1
Network Intelligence and Services	Provided by Servers (Proxy, Redirect, Registrar)	Provided by Gatekeepers
Clients	Intelligent	Intelligent
Emphasis	Multimedia, multicast, telephony	Telephony, video, whiteboard, document sharing
Transport	Mostly UDP	V1=TCP; V2=UDP
Address	SIP URLs	Aliases

Service	SIP	H.323
Call Hold	Yes	Yes
Call Transfer	Yes	Yes
Call Forward	Yes	Yes
Call Waiting	Yes	Yes
Third-Party call	Yes	No
Conference	Yes	Yes
Click-to-Dial	Yes	Yes
Capability Exchange	Yes	Yes

QoS	SIP	H.323
Call Setup Delay	1.5 rt	2.5 rt
Packet Loss Recovery	Yes	Yes
Loop Detection	Via Hops	Path Value
Fault Tolerance	Yes	Backup

Figure 2.19 : Comparing H.323 and SIP

2.5 Real-Time Transport Protocol-RTP

Real-time transport protocol (RTP) is a key protocol in any VoIP solution. It is used to carry applications that are time sensitive, and thus need to be delivered in real-time. This includes voice, which if delivered too early or too late, will result in a degradation of voice quality at the receiving end. RTP is defined by the IETF in RFC 1889.

When one refers to VoIP, one is essentially referring to the encapsulation of voice samples into RTP packets, for transmission over an IP network. RTP packets are usually encapsulated in UDP, to be transported to the desired destination.

A Media Gateway (MG) is the network element in a VoIP network that is responsible for packetizing the voice traffic. The media gateway will perform the RTP encapsulation, and handle the RTP packet flows from other media gateways.

The RTP protocol stack is shown below in Figure 2.20. In the diagram, RTP packets flow between two media gateways, but RTP can also be used for transporting voice packets between other types of trunk gateways, as well as between other types of network elements such as line gateways and VoIP clients.

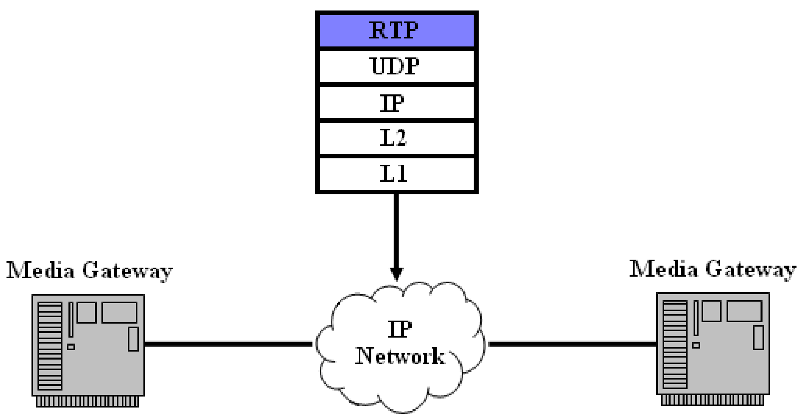


Figure 2.20 : Basic diagram showing the RTP protocol stack

Each direction has a separate and unique RTP flow. Each flow has a source and destination IP address as well as a source and destination UDP port number. All of this information is negotiated at call setup using protocols such as H.248, MGCP or SIP.

RTP identifies the type of payload that is carrying, which will often be the audio codec that is being used. It also provides timestamp information and sequence numbering. Figure 2.21 shows the format of an RTP packet.

2	3	4	8	9	16bit	32bit
V	P	X	CSRC count	M	Payload type	Sequence number
Timestamp						
SSRC: Synchronization source						
CSRC:Contributing source (variable 0-15 items 32bits each)						

Figure 2.21 : Real-time transport protocol packet header

- V -- Version (2 bits). Identifies the RTP version. The current RFC version is 2.
- P -- Padding (1 bit). If the padding bit is set, the packet contains one or more bytes of padding at the end of the packet. The very last byte indicates the number of bytes of padding.
- X -- Extension bit (1 bit). If the extension bit is set, the fixed header is followed by a header extension. Although it is rarely used, a header extension allows applications to carry additional information in the RTP header.
- CSRC count (4 bits) -- Contains the number of CSRC identifiers that follow the fixed header.
- M -- Marker (1 bit). The interpretation of the marker is defined by a profile. It is intended to allow significant events such as frame boundaries to be marked in the packet stream.
- Payload type (7 bits) -- Identifies the format of the RTP payload and determines its interpretation by the application. A profile specifies a default static mapping of payload type codes to payload formats. Additional payload type codes may be defined dynamically through non-RTP means. The packet will be processed, depending on the value in this field. For example, if the payload type is 18, the system knows that this packet has been compressed using the G.729 codec.
- Sequence number (16 bits) -- Increments by one for each RTP data packet sent, and may be used by the receiver to detect packet loss and to restore packet sequence.
- Timestamp (32 bits) -- Reflects the sampling instant of the first octet in the RTP data packet. The sampling instant must be derived from a clock that

increments monotonically and linearly in time to allow synchronization and jitter calculations.

- SSRC (32 bits) -- Synchronization source. This identifier is chosen randomly, with the intent that no two synchronization sources within the same RTP session will have the same SSRC identifier.
- CSRC (1-15 items, 32 bits each) -- Contributing source identifiers list. Identifies the contributing sources for the payload contained in this packet.

Table 2.7 lists some of the common payload types used in VoIP communications.

Table 2.7: Common RTP payload types

Payload Type	CODEC
0	G.711 u-law (PCMU)
8	G.711 A-law (PCMA)
13	Comfort Noise
18	G.729
101	DTMF Relay

An example of an RTP packet is provided in Figure 2.22. The payload type for this packet is 8, indicating that the payload is voice encoded using the G.711 A-law (PCMA) codec. The last line is the RTP payload, which contains the actual encoded voice sample.

(Payload: 1A24270E86B59C6ACFB7B98B0927240488B1846CC1B5B285...)

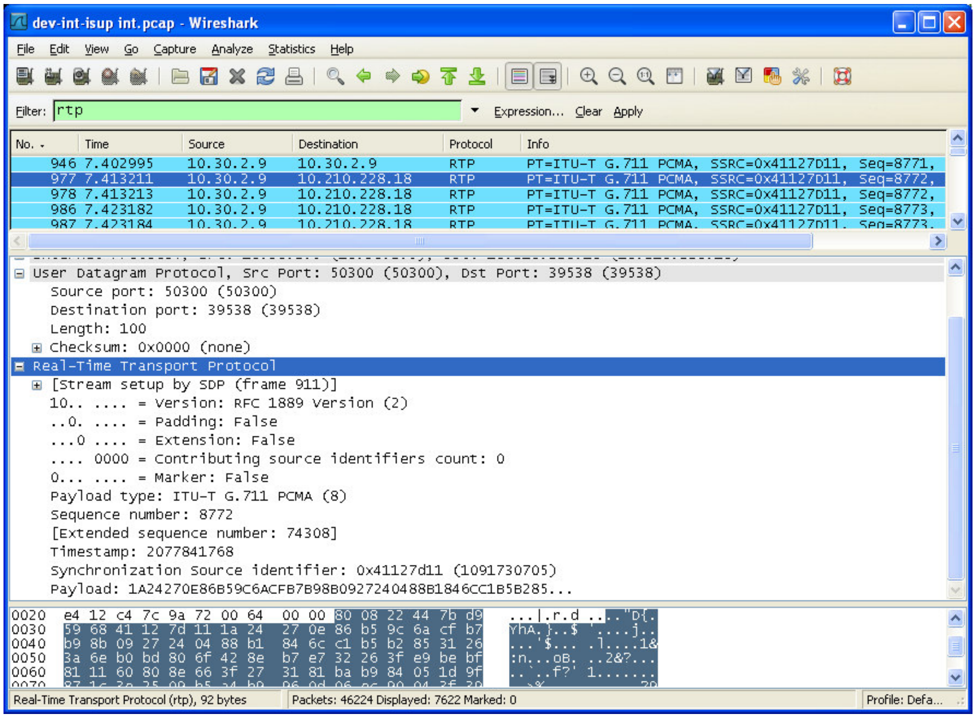


Figure 2.22 : Example of an RTP packet

2.6 Real-Time Control Protocol (RTP Control Protocol-RTCP)

The Real-time Transport Control Protocol (RTCP) is defined in RFC 1889. It is designed to carry and exchange control information about the participants that are involved in the RTP flows. RTCP also gathers statistical information about the flows which can be used to monitor quality of service.

Real-time Transport Control Protocol (RTCP) is a sister protocol to RTP that provides out-of-band control information for an RTP flow. RTCP gathers statistics on a media connection and information such as the number of packets and bytes sent, the number of packets lost, the amount of jitter and the round trip delay. An application can use this information to improve the quality of the service by perhaps limiting the traffic flow or by switching to a higher or lower compression codec.

There are several different types of RTCP packets such as SR (Sender Report), RR (Receiver Report), SDES (Source Description), BYE and APP RTCP packets. This specification defines several RTCP packet types to carry a variety of control information:

- SR: Sender report, for transmission and reception statistics from participants that are active senders

- RR: Receiver report, for reception statistics from participants that are not active senders

- SDES: Source description items, including CNAME

- BYE: Indicates end of participation

- APP: Application specific functions

An RTP flow is established on an even numbered UDP port (N). The RTCP flow, associated with this RTP flow, is established on the next higher odd numbered UDP port, by incrementing the RTP port number by one (N+1). An RTCP packet is sent every second.

Sender Report (SR) packets contain information such as the number of packets and bytes sent, the number of packets lost, the amount of jitter and the round trip delay. A Sender Report packet (SR packet) is issued if the media gateway has sent an RTP packet during the interval since issuing the last report; otherwise a Receiver Report packet (RR packet) is issued. An example of SR packet is provided in Figure 2.23.

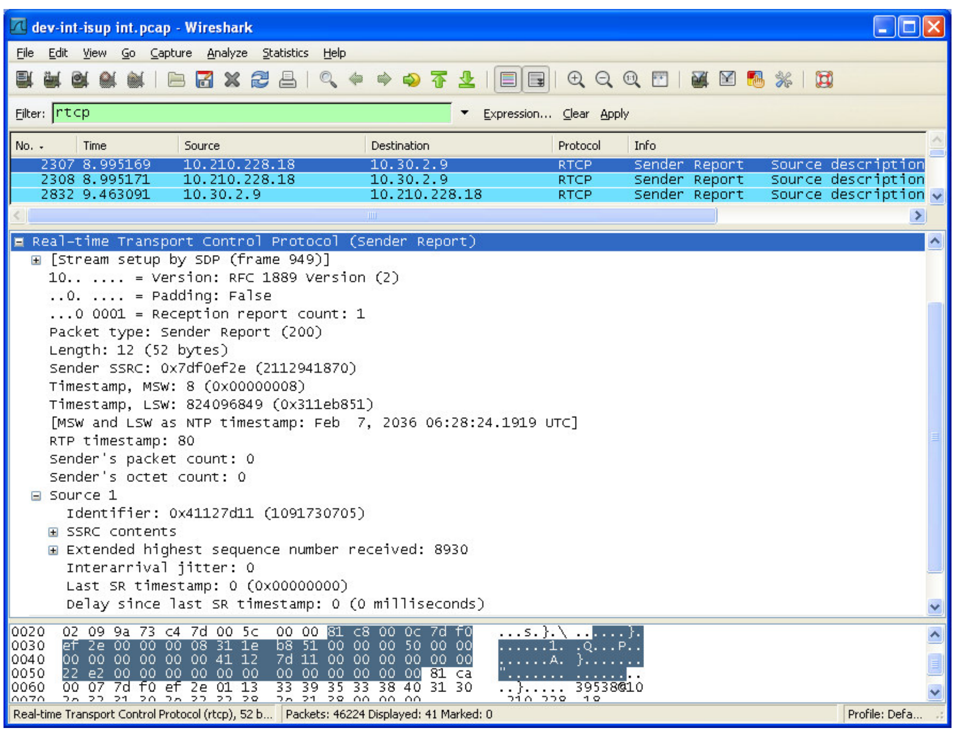


Figure 2.23 : Example of a SR packet

The Source Description (SDES) packet is used to carry a unique identifier for an RTP source called a canonical name (CNAME). The CNAME (Canonical end-point identifier) is the only mandatory parameter in the Source Description packet (SDES packet) and is the only parameter to be transmitted by this feature. It is a variable length text string encoded according to UTF-8 (RFC 2279). For the characters transmitted by this feature, this is equivalent to US-ASCII encoding. The transmitted CNAME parameter is of the form:

<udpPort>@<medialpAddress>

The <udpPort> field corresponds to the local UDP port for RTP traffic in decimal, and the <medialpAddress> field is the local media IP address in dotted decimal form. For example, if local RTP UDP port is 49152 and the local media IP address is 10.1.2.134, then the CNAME transmitted is 49152@10.1.2.134. An example of a SDES packet is provided in Figure 2.24.

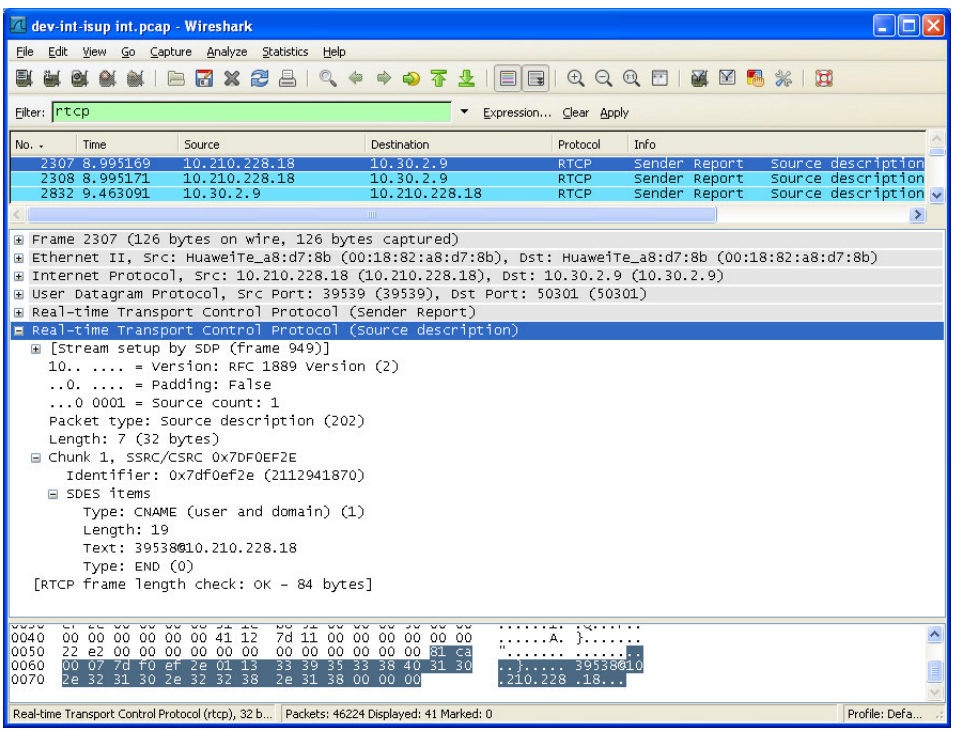


Figure 2.24 : Example of a SDES packet

Protocol structure of RTCP is shown in Figure 2.25.

2	3	8	16bit
Version	P	RC	Packet type
Length			

Figure 2.25 : Protocol structure of RTCP

Version -- Identifies the RTP version which is the same in RTCP packets as in RTP data packets. The version defined by this specification is two (2).

P -- Padding. When set, this RTCP packet contains some additional padding octets at the end which are not part of the control information. The last octet of the padding is a count of how many padding octets should be ignored. Padding may be needed by some encryption algorithms with fixed block sizes. In a compound RTCP packet, padding should only be required on the last individual packet because the compound packet is encrypted as a whole.

RC -- Reception report count, the number of reception report blocks contained in this packet. A value of zero is valid. Packet type contains the constant 200 to identify this as an RTCP SR packet.

Length -- The length of this RTCP packet in 32-bit words minus one, including the header and any padding. (The offset of one makes zero a valid length and avoids a possible infinite loop in scanning a compound RTCP packet, while counting 32- bit words avoids a validity check for a multiple of 4.)

3. VIDEO CONFERENCING CODECS

3.1 H.261 Video Codec for Low Quality Video-conferencing

H. 261 is the video coding standard which was developed by the ITU (International Telecom Union) in 1990. It has been specifically designed for video telecommunication applications, meant for videoconferencing, videotelephone applications over ISDN (Integrated Services Digital Network) telephone lines. H.261 was designed for datarates which are multiples of 64Kbit/s, and is sometimes called p x 64Kbit/s (p is in the range of 1-30). These datarates suit ISDN lines, for which this video codec was originally designed for. One or both B-channels of a narrowband ISDN connection can transfer video data, for example, in addition to speech. This requires that both partners connected via the channel have to use the same coding of video data. In a narrowband ISDN connection, exactly two B-channels and one D-channel are available at the user interface. The European ISDN hierarchy also allows a connection with 30 B-channels. The primary applications of ISDN were videophone and videoconferencing systems. These dialogue applications require that coding and decoding be carried out in real time. H.261 uses real-time transport protocol (RTP) to transport video stream. The basic approach to H. 261 Compression is summarized in Figure 3.1.

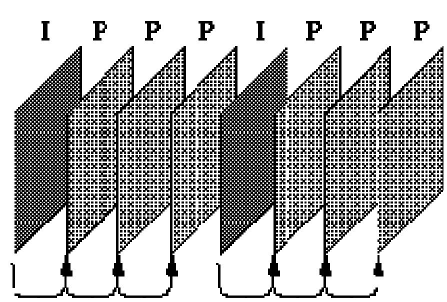


Figure 3.1 : I- P-Frames structure of H.261

The coding algorithm is a hybrid of inter-picture prediction, transform coding and motion compensation. The data rate of the coding algorithm was designed to be able to be set to between 40 Kbps and 2 Mbps. INTRA coding encodes blocks of 8x8 pixels each only with reference to themselves and sends them directly to the block transformation process. Figure 3.2 shows the structure of I-Frames.

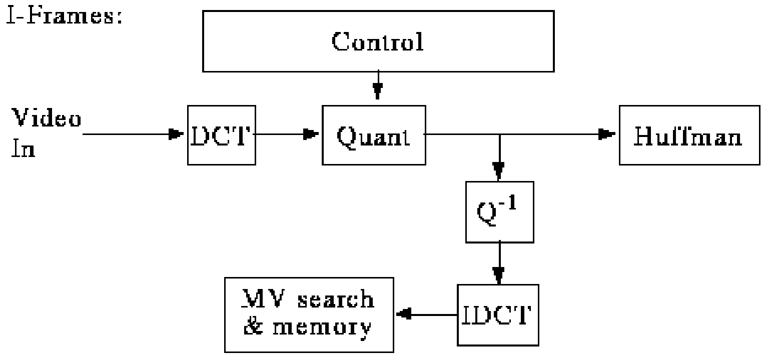


Figure 3.2 : Detailed I-Frames structure

On the other hand, INTER coding frames are encoded with respect to another reference frame. Figure 3.3 shows a detailed block diagram of P-Frames. The inter-picture prediction removes temporal redundancy. The transform coding removes the spatial redundancy. Motion vectors are used to help the codec compensate for motion. To remove any further redundancy in the transmitted bitstream, variable length coding is used. H261 supports motion compensation in the encoder as an option. In motion compensation a search area is constructed in the previous (recovered) frame to determine the best reference macroblock.

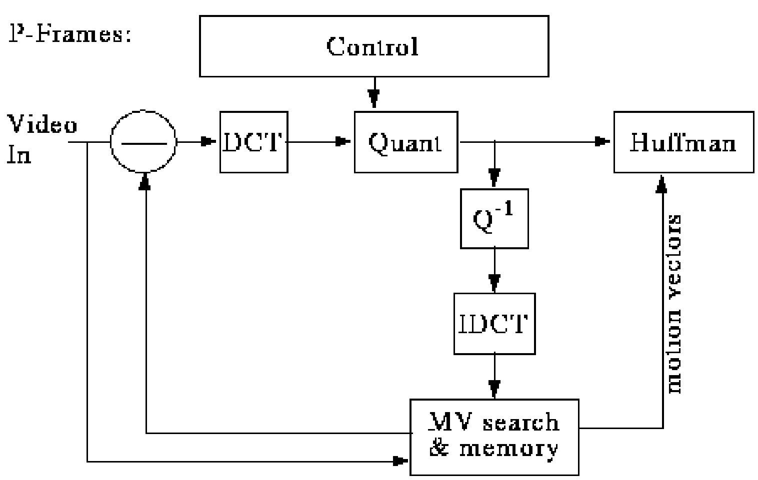


Figure 3.3 : Detailed P-Frames structure

The term intra frame coding refers to the fact that the various lossless and lossy compression techniques are performed relative to information that is contained only within the current frame, and not relative to any other frame in the video sequence. In other words, no temporal processing is performed outside of the current picture or frame. This mode will be described first because it is simpler, and because non-intra coding techniques are extensions to these basics. Figure 3.4 shows a block diagram of a basic video encoder for intra frames only. It turns out that this block diagram is very similar to that of a JPEG (Joint Photographic Experts Group) still image video encoder, with only slight implementation detail differences.

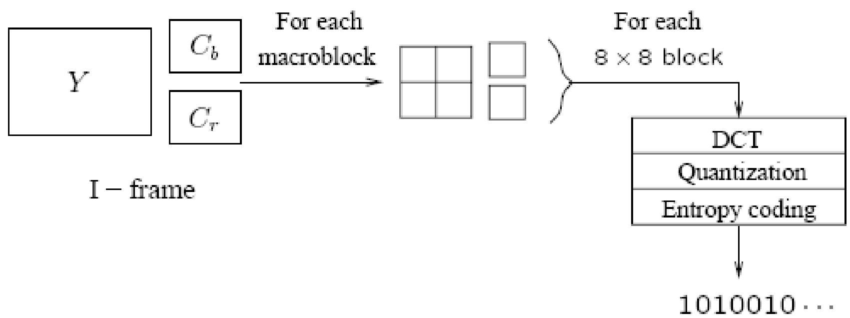


Figure 3.4 : I-Frame coding

The previously discussed intra frame coding techniques were limited to processing the video signal on a spatial basis, relative only to information within the current video frame. Considerably more compression efficiency can be obtained however, if the inherent temporal, or time-based redundancies, are exploited as well. Anyone who has ever taken a reel of the old-style super-8 movie film and held it up to a light can certainly remember seeing that most consecutive frames within a sequence are very similar to the frames both before and after the frame of interest. Temporal processing to exploit this redundancy uses a technique known as block-based motion compensated prediction, using motion estimation. A block diagram of the basic encoder and decoder with extensions for non-intra frame coding techniques is given in Figure 3.5 and Figure 3.6 respectively. Of course, this encoder can also support intra frame coding as a subset.

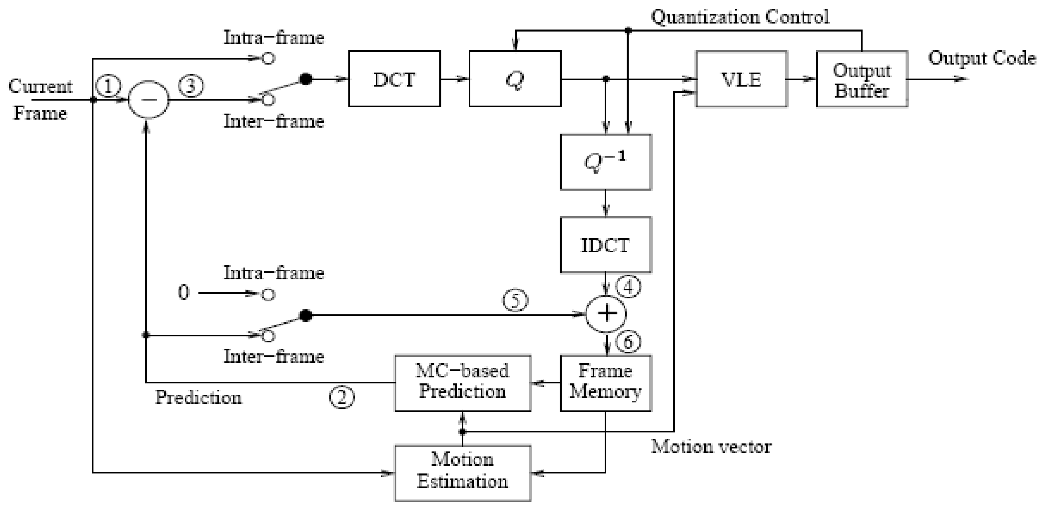


Figure 3.5 : Coding for video conferencing (H.261 encoder)

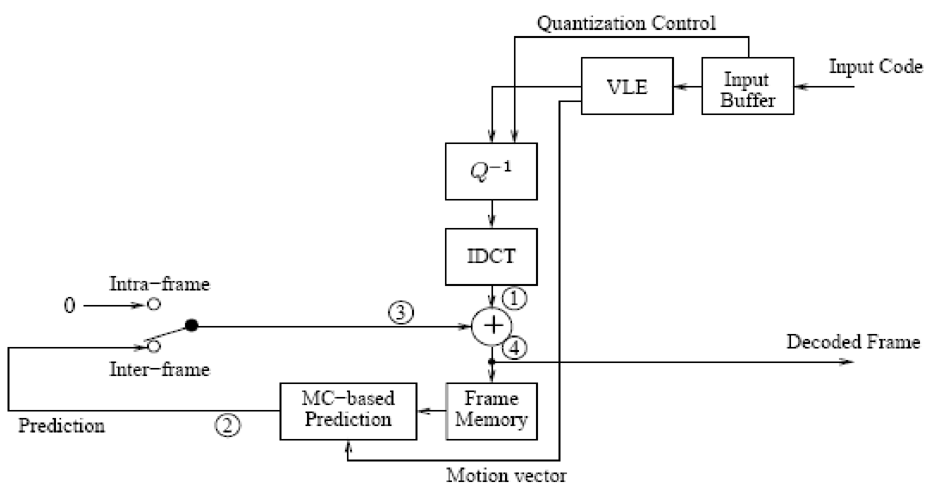


Figure 3.6 : Coding for video conferencing (H.261 decoder)

Starting with an intra, or I frame, the encoder can forward predict a future frame. This is commonly referred to as a P frame, and it may also be predicted from other P frames, although only in a forward time manner. As an example, consider a group of pictures that lasts for 6 frames. In this case, the frame ordering is given as I,P,P,P,P,P,I,P,P,P,P. Each P frame in this sequence is predicted from the frame immediately preceding it, whether it is an I-frame or a P-frame. As a reminder, I frames are coded spatially with no reference to any other frame in the sequence. P-coding can be summarized as follows:

- A Coding Example (P-frame)
- Previous image is called reference image.

- Image to code is called target image.
- Actually, the difference is encoded.

Comparison of target frame and the reference frame is explained in Figure 3.7.

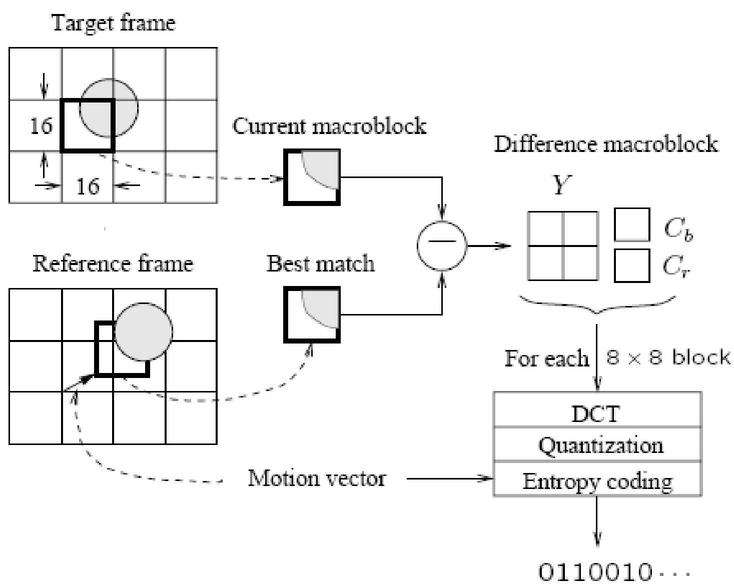


Figure 3.7 : Target frame – reference frame difference

H261 supports two image resolutions, QCIF (Quarter Common Interchange format) which is (144x176 pixels) and CIF (Common Interchange format), which is (288x352). Table 3.1 shows the supported resolutions in H.261 codec.

Table 3.1: Picture formats supported by H.261

Picture Format	Luminance pixels	Luminance lines	H.261 support	Uncompressed bitrate (Mbps)			
				10 frames/s		30 frames/s	
				Grey	Colour	Grey	Colour
QCIF	176	144	Yes	2.0	3.0	6.1	9.1
CIF	352	288	Optional	8.1	12.2	24.3	36.5

The video multiplexer structures the compressed data into a hierarchical bitstream that can be universally interpreted. The hierarchy has four layers:

- Picture layer: corresponds to one video picture (frame)
- Group of blocks: corresponds to 1/12 of CIF pictures or 1/3 of QCIF
- Macroblocks : corresponds to 16x16 pixels of luminance and the two spatially corresponding 8x8 chrominance components
- Blocks: corresponds to 8x8 pixels

Figure 3.8 and 3.9 shows the structure of Picture, GOB, MB, and Block Layers.

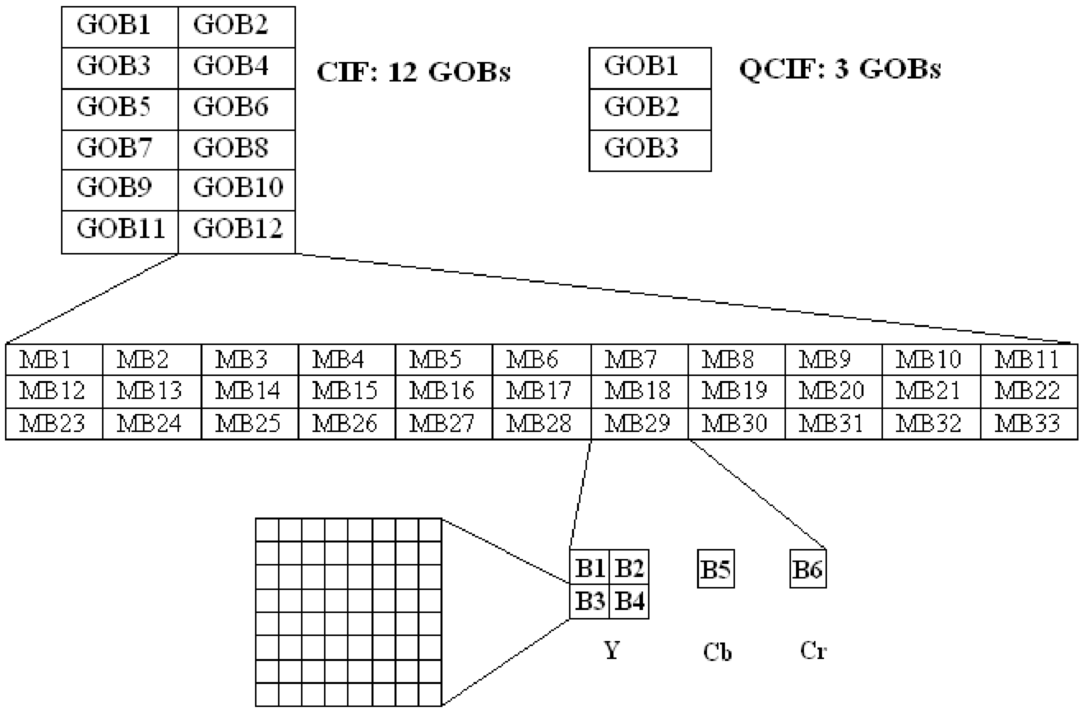


Figure 3.8 : Picture, GOB, MB, and block

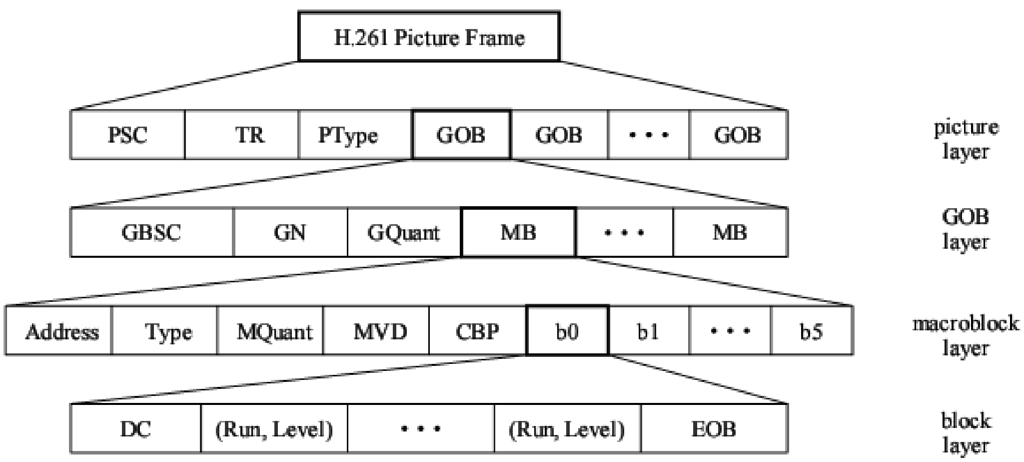


Figure 3.9 : Syntax of H.261 video bitstream

PSC - Picture Start Code (20 bits): A word of 20 bits. Its value is 0000 0000 0000 0001 0000.

TR – Temporal Reference (5 bits): A 5-bit number which can have 32 possible values. It is formed by incrementing its value in the previously transmitted picture header by one plus the number of non-transmitted pictures (at 29.97 Hz) since that last transmitted one. The arithmetic is performed with only the five LSBs (Least Significant Bit).

PType – Picture Type (6 bits): Information about the complete picture:

Bit 1 Split screen indicator, “0” off, “1” on;

Bit 2 Document camera indicator, “0” off, “1” on;

Bit 3 Freeze picture release, “0” off, “1” on;

Bit 4 Source format, “0” QCIF, “1” CIF;

Bit 5 Optional still image mode HI_RES defined in Annex D; “0” on, “1” off;

Bit 6 Spare.

GOB – Group of Blocks

GBSC – GOB Start Code (16 bits): A word of 16 bits, 0000 0000 0000 0001.

GN – Group Number (4 bits): Four bits indicating the position of the group of blocks. The bits are the binary representation of the number. Group numbers 13, 14 and 15 are reserved for future use. Group number 0 is used in the PSC.

GQuant – GOB Quantizer (5 bits): A fixed length codeword of 5 bits which indicates the quantizer to be used in the group of blocks until overridden by any subsequent MQUANT. The codewords are the natural binary representations of the values of QUANT (see 4.2.4) which, being half the step sizes, range from 1 to 31.

MB – Macroblock

MBA – MB Address (Variable Length): A variable length codeword indicating the position of a macroblock within a group of blocks. For the first transmitted macroblock in a GOB, MBA is the absolute address. For subsequent macroblocks, MBA is the difference between the absolute addresses of the macroblock and the last transmitted macroblock. MBA is always included in transmitted macroblocks.

MType – MB Type (Variable Length): Variable length codewords giving information about the macroblock and which data elements are present. MTYPE is always included in transmitted macroblocks.

MQuant – MB Quantizer (5 bits): MQUANT is present only if so indicated by MTYPE. A codeword of 5 bits signifying the quantizer to be used for this and any following blocks in the group of blocks until overridden by any subsequent MQUANT. Codewords for MQUANT are the same as for GQUANT.

MVD – Motion Vector Data (Variable Length): Motion vector data is included for all MC macroblocks. MVD is obtained from the macroblock vector by subtracting the vector of the preceding macroblock. For this calculation the vector of the preceding macroblock is regarded as zero in the following three situations:

1) evaluating MVD for macroblocks 1, 12 and 23;

2) evaluating MVD for macroblocks in which MBA does not represent a difference of 1;

3) MTYPE of the previous macroblock was not MC.

MVD consists of a variable length codeword for the horizontal component followed by a variable length codeword for the vertical component. Advantage is taken of the fact that the range of motion vector values is constrained. Each VLC (Variable Length Coding) word represents a pair of difference values. Only one of the pair will yield a macroblock vector falling within the permitted range.

CBP – Coded Block Pattern (Variable Length): CBP is present if indicated by MTYPE. The codeword gives a pattern number signifying those blocks in the macroblock for which at least one transform coefficient is transmitted. The pattern number is given by:

$$32 \cdot P1 + 16 \cdot P2 + 8 \cdot P3 + 4 \cdot P4 + 2 \cdot P5 + P6 \qquad \textbf{(3.1)}$$

where $Pn = 1$ if any coefficient is present for block n, else 0.

3.2 H.263 Video Codec for Medium Quality Video-conferencing

The H.263 is the video coding standard developed by the International Telecommunications Union (ITU) for video-conferencing and video-telephony applications. It was developed to transport video streams at bandwidths of 20 Kbps to 24 Kbps and was based on H.261 codec. H.263 requires half of the bandwidth to achieve the same video quality compared to H.261 video codec and this makes H.263 to be largely replaced among the H.261. H.263 uses RTP protocol to transport video streams over the network. Outlined block diagram of the H.263 video codec is shown in Figure 3.10.

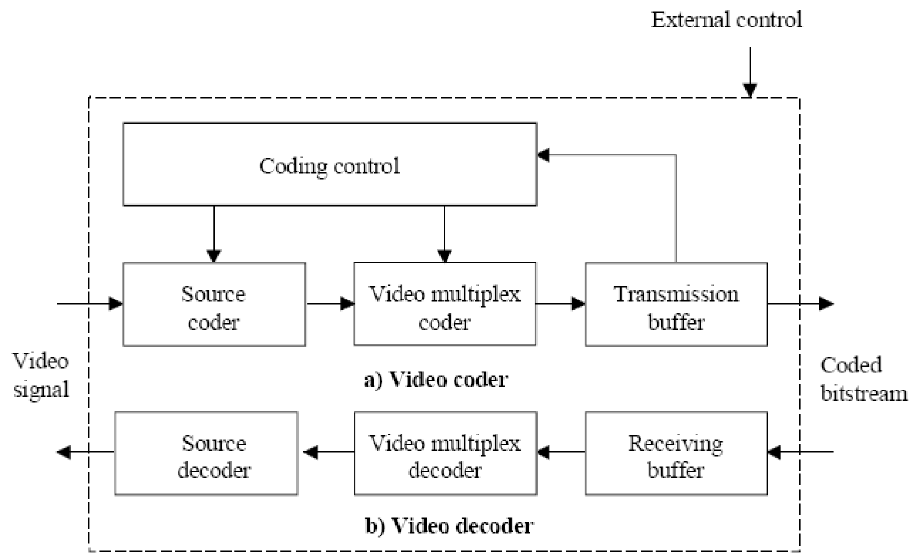


Figure 3.10 : Outline block diagram of the video codec

Figure 3.11 below shows a block diagram of an H.263 baseline encoder. Motion-compensated prediction first reduces temporal redundancies. Discrete Cosine Transform (DCT)-based algorithms are then used for encoding the motion-compensated prediction difference frames. The quantized DCT coefficients, motion vectors, and side information are entropy coded using variable length codes (VLC's).

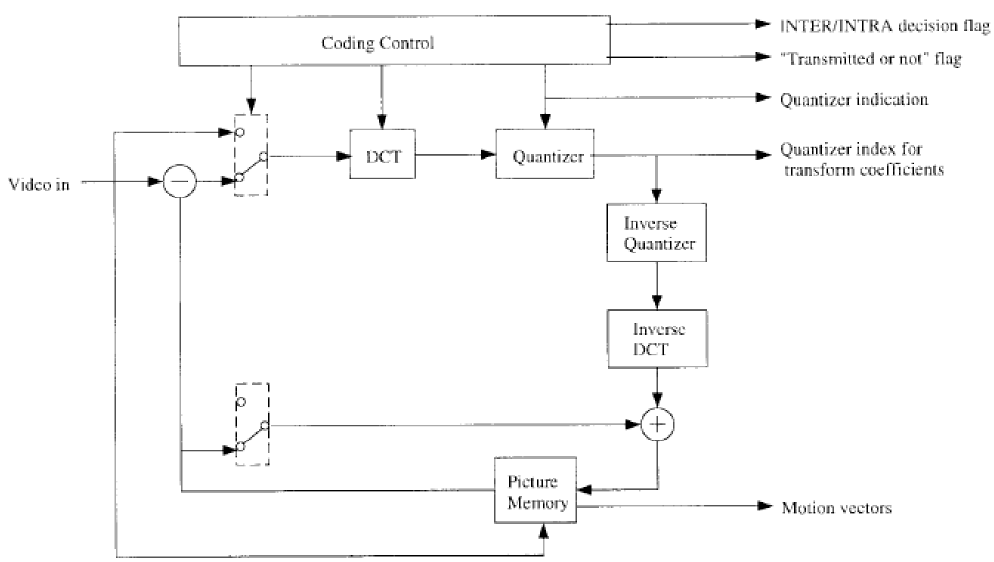


Figure 3.11 : H.263 video encoder block diagram

Since H.263 was based on the H.261 codec, the coding algorithm is similar to that used for H.261, but with some improvements on performance and error recovery processes. The decoder has motion compensation capability, allowing optional incorporation of this technique in the coder. Half pixel precision is used for the motion compensation, as opposed to ITU-T Rec. H.261 where full pixel precision and a loop filter are used. Variable length coding is used for the symbols to be transmitted.

In addition to the basic video source coding algorithm, there are also eighteen negotiable coding options are included for improved compression performance and the support of additional capabilities (Continuous Presence Multipoint and Video Multiplex mode, Unrestricted Motion Vector mode, Syntax-based Arithmetic Coding mode, Advanced Prediction mode, PB-frames mode, Forward Error Correction, Advanced INTRA Coding mode, Deblocking Filter mode, Slice Structured mode, Supplemental enhancement information, Improved PB-frames mode, Reference Picture Selection mode, Temporal, SNR and Spatial Scalability mode, Reference Picture Resampling mode, Reduced-Resolution Update mode, Independent Segment Decoding mode, Alternative INTER VLC mode, Modified Quantization mode). The most important four are Unrestricted Motion Vectors, Syntax-based arithmetic coding, Advance prediction, and forward-and-backward frame prediction similar to MPEG (Moving Pictures Experts Group) called P-B frames.

In baseline H.263, motion vectors can only reference pixels that are within the picture area. Because of this, macroblocks at the border of a picture may not be well predicted. When the Unrestricted Motion Vector mode (Annex D) is used, motion vectors can take on values in the range [-31.5, 31.5] instead of [-16, 15.5] and are allowed to point outside the picture boundaries. The longer motion vectors improve coding efficiency for larger picture formats, i.e. 4CIF or 16CIF. Also, by allowing motion vectors to point outside the picture, a significant gain is achieved if there is movement along picture edges. This is especially useful in the case of camera movement or background movement.

In Syntax-based Arithmetic Coding mode (Annex E), all VLC coding operations are replaced by syntax-based arithmetic coding operations. Since both VLC and arithmetic coding are lossless coding schemes, the resulting picture quality is not affected, yet the bit rate can be reduced by approximately 5%, due to the more efficient arithmetic codes.

Advanced Prediction mode (Annex F) allows to use four motion vectors per macroblock, one for each of the four 8x8 luminance blocks. Furthermore, overlapped block motion compensation is used for the luminance macroblocks, and motion vectors are allowed to point outside the picture as in the Unrestricted Motion Vector mode. Use of this mode improves inter picture prediction and yields a significant improvement in subjective picture quality for the same bit rate by reducing blocking artifacts.

In PB-Frames mode (Annex G), the frame structure consists of a P-picture and a B-picture. The quantized DCT coefficients of the B- and P-pictures are interleaved at the macroblock layer such that a P-picture macroblock is immediately followed by a B-picture macroblock. The P-picture is forward predicted from the previously decoded P-picture. The B-picture is bi-directionally predicted from the previously decoded P-picture and the P-picture currently being decoded. With this coding option, the picture rate can be increased considerably without substantially increasing the bit rate. However, an Improved PB-frames mode is also provided (Annex M). The original PB-frames mode is retained herein only for purposes of compatibility with systems made prior to the adoption of the Improved PB-frames mode. Figure 3.12 shows the Forward prediction, Backward prediction and Bidirectional prediction structure.

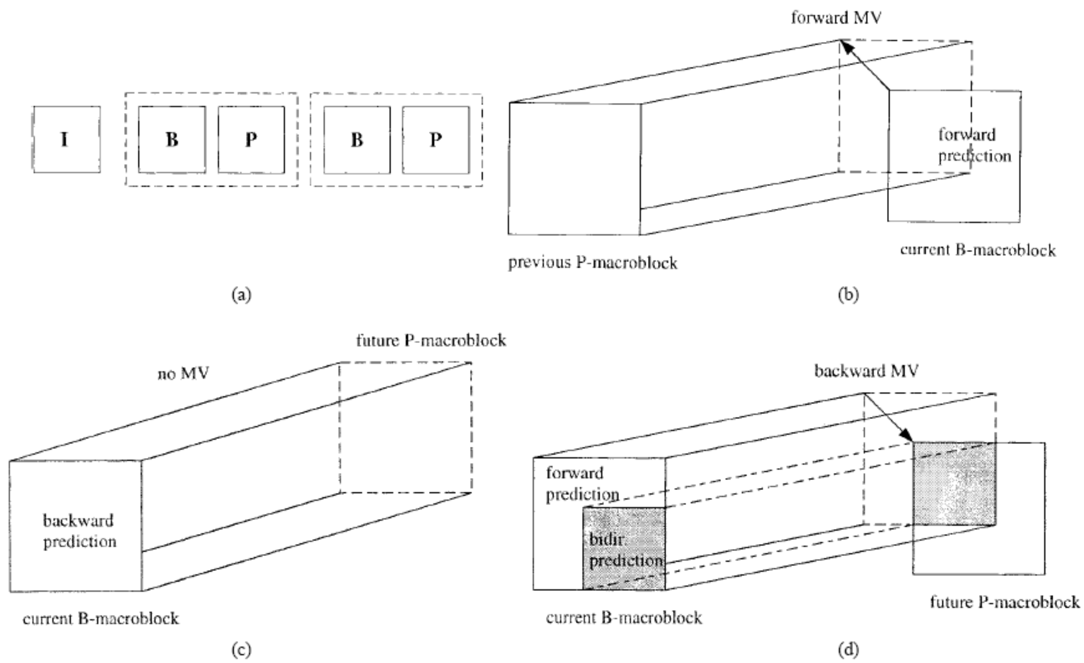


Figure 3.12 : Improved PB frames. (a) Structure. (b) Forward prediction. (c) Backward prediction. (d) Bidirectional prediction

H.263 supports five resolutions. In addition to QCIF and CIF that were supported by H.261 codec, there are SQCIF (Sub Quarter Common Interchange Format), 4CIF and 16CIF as well. SQCIF is approximately the half the resolution of QCIF, while 4CIF and 16CIF are 4 and 16 times the resolution of CIF respectively. The support of 4CIF and 16CIF means the codec can compete with other higher bitrate video coding standards such as the MPEG standards. The differences between the H.261 and H.263 coding algorithms are listed in Table 3.2.

Table 3.2: Difference between H.261 and H.263 video codecs

Picture Format	Luminance pixels	Luminance lines	H.261 support	H.263 support	Uncompressed bit rate (Mbps)			
					10 frames/s		30 frames/s	
					Grey	Color	Grey	Color
SQCIF	128	96	No	Yes	1.0	1.5	3.0	4.4
QCIF	176	144	Yes	Yes	2.0	3.0	6.1	9.1
CIF	352	288	Optional	Optional	8.1	12.2	24.3	36.5
4CIF	704	576	No	Optional	32.4	48.7	97.3	146.0
16CIF	1408	1152	No	Optional	129.8	194.6	389.3	583.9

The picture structure is shown in Figure 3.13 for the QCIF resolution. Each picture in the input video sequence is divided into macroblocks, consisting of four luminance blocks of 8 pixelsx8 lines followed by one Cb block and one Cr block, each consisting of 8 pixelsx8 lines. A group of blocks (GOB) is defined as an integer number of macroblock rows, a number that is dependent on picture resolution. For example, a GOB consists of a single macroblock row at QCIF resolution.

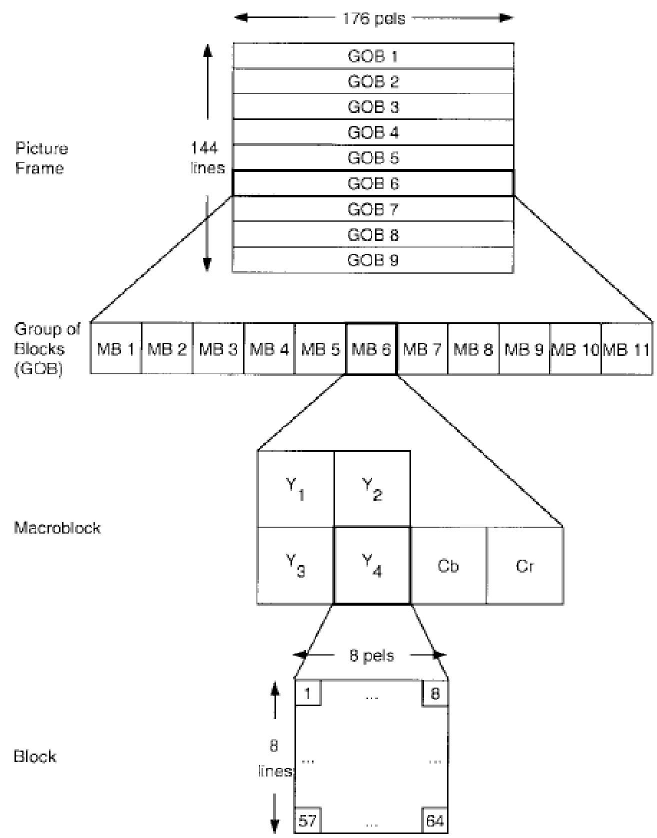


Figure 3.13 : H.263 picture structure at QCIF resolution

H.263 supports inter-picture prediction that is based on motion estimation and compensation. The coding mode where temporal prediction is used is called an inter mode. In this mode, only the prediction error frames - the difference between original frames and motion-compensated predicted frames - need be encoded. If temporal prediction is not employed, the corresponding coding mode is called an intra mode.

Motion-compensated prediction assumes that the pixels within the current picture can be modeled as a translation of those within a previous picture, as shown in Figure 3.14. In baseline H.263, each macroblock is predicted from the previous frame. This implies an assumption that each pixel within the macroblock undergoes the same amount of translational motion.

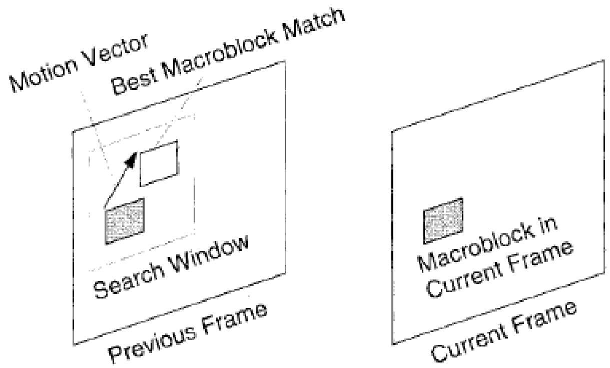


Figure 3.14 : H.263 source coding algorithm: motion compensation

H.263 video streams need to be packetized for transportation over networks. The transport protocol for H.263 streams is the Real Time Transport Protocol (RTP). Three formats (mode A, mode B and mode C) are defined for the H.263 payload header. In mode A, an H.263 payload header of four bytes is present before the actual compressed H.263 video bitstream. It allows fragmentation at GOB boundaries. In mode B, an 8- byte H.263 payload header is used and each packet starts at MB boundaries without the PB-frames option. Finally, a 12- byte H.263 payload header is defined in mode C to support fragmentation at MB boundaries for frames that are coded with the PB-frames option.

3.3 H.264 Video Codec for High Quality Video-conferencing

The H.264 codec was jointly developed by two standards bodies which are ITU and the ISO/IEC (International Organization for Standardization / International Electro-technical Commission). As a result, H.264 can be interpreted as a joint effort of the ITU H.264 and the ISO MPEG-4, Part 10. H.264 is also known by its more generic name AVC, for Advanced Video Codec. The H.264 video codec has a very broad range of applications that covers all forms of digital compressed video from low bit-rate Internet streaming applications to HDTV broadcast and Digital Cinema applications with nearly lossless coding. The next standard in the MPEG series, MPEG4, is enabling a new generation of internet-based video applications whilst the ITU-T H.263 standard for video compression is now widely used in videoconferencing systems. H.264/MPEG-4 has achieved a significant improvement in the rate-distortion efficiency –providing a factor of two in bit-rate savings compared with MPEG-2 Video, which is the most common standard used for video storage and transmission. The coding gain of H.264 over H.263 is in the range of 25% to 50%, depends on the types of applications.

The H.264/MPEG-4 design covers a Video Coding Layer (VCL), which efficiently represents the video content, and a Network Abstraction Layer (NAL), which formats the VCL representation of the video and provides header information in a manner appropriate for conveyance by particular transport layers (such as Real Time Transport Protocol) or storage media. All data are contained in NAL units, each of which contains an integer number of bytes. A NAL unit specifies a generic format for use in both packet-oriented and bitstream systems. The format of NAL units for both packet-oriented transport and bitstream delivery is identical - except that each NAL unit can be preceded by a start code prefix in a bitstream-oriented transport layer. Structure of H.264/AVC video encoder is shown in Figure 3.15.

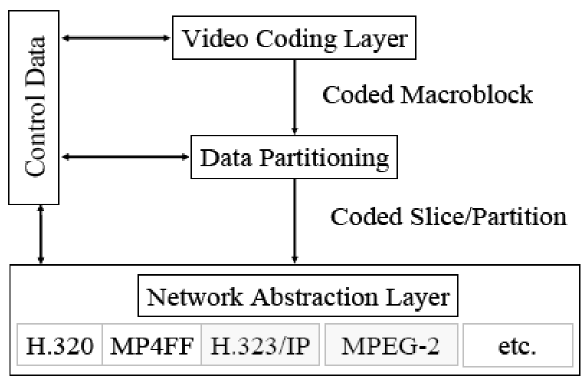


Figure 3.15 : Structure of H.264/AVC video encoder

Relative to prior video coding methods, as exemplified by MPEG-2 video, some highlighted features of the design that enable enhanced coding efficiency include the following enhancements of the ability to predict the values of the content of a picture to be encoded:

Variable block-size motion compensation with small block sizes: This standard supports more flexibility in the selection of motion compensation block sizes and shapes than any previous standard, with a minimum luma motion compensation block size as small as 4x4.

Quarter-sample-accurate motion compensation: Most prior standards enable half sample motion vector accuracy at most. The new design improves up on this by adding quarter sample motion vector accuracy, as first found in an advanced profile of the MPEG-4 Visual (part 2) standard, but further reduces the complexity of the interpolation processing compared to the prior design.

Motion vectors over picture boundaries: While motion vectors in MPEG-2 and its predecessors were required to point only to areas within the previously-decoded reference picture, the picture boundary extrapolation technique first found as an optional feature in H.263 is included in H.264/AVC.

Multiple reference picture motion compensation: Predictively coded pictures (called "P" pictures) in MPEG-2 and its predecessors used only one previous picture to predict the values in an incoming picture. The new design extends upon the enhanced reference picture selection technique found in H.263++ to enable efficient coding by allowing an encoder to select, for motion compensation purposes, among a larger number of pictures that have been decoded and stored in the decoder. The same extension of referencing capability is also applied to motion-compensated bi-prediction, which is restricted in MPEG-2 to using two specific pictures only (one of these being the previous intra (I) or P picture in display order and the other being the next I or P picture in display order).

Decoupling of referencing order from display order: In prior standards, there was a strict dependency between the ordering of pictures for motion compensation referencing purposes and the ordering of pictures for display purposes. In H.264/AVC, these restrictions are largely removed, allowing the encoder to choose the ordering of pictures for referencing and display purposes with a high degree of flexibility constrained only by a total memory capacity bound imposed to ensure decoding ability. Removal of the restriction also enables removing the extra delay previously associated with bi-predictive coding.

Decoupling of picture representation methods from picture referencing capability: In prior standards, pictures encoded using some encoding methods (namely bi-predictively-encoded pictures) could not be used as references for prediction of other pictures in the video sequence. By removing this restriction, the new standard provides the encoder more flexibility and, in many cases, an ability to use a picture for referencing that is a closer approximation to the picture being encoded.

Weighted prediction: A new innovation in H.264/AVC allows the motion-compensated prediction signal to be weighted and offset by amounts specified by the encoder. This can dramatically improve coding efficiency for scenes containing fades, and can be used flexibly for other purposes as well.

Improved "skipped" and "direct" motion inference: In prior standards, a "skipped" area of a predictively-coded picture could not motion in the scene content. This had a detrimental effect when coding video containing global motion, so the new H.264/AVC design instead infers motion in "skipped" areas. For bi-predictively coded areas (called B slices), H.264/AVC also includes an enhanced motion inference method known as "direct" motion compensation, which improves further on prior "direct" prediction designs found in H.263+ and MPEG-4 Visual.

Directional spatial prediction for intra coding: A new technique of extrapolating the edges of the previously-decoded parts of the current picture is applied in regions of pictures that are coded as intra (i.e., coded without reference to the content of some other picture). This improves the quality of the prediction signal, and also allows prediction from neighboring areas that were not coded using intra coding (something not enabled when using the transform-domain prediction method found in H.263+ and MPEG-4 Visual).

In-the-loop deblocking filtering: Block-based video coding produces artifacts known as blocking artifacts. These can originate from both the prediction and residual difference coding stages of the decoding process. Application of an adaptive deblocking filter is a well-known method of improving the resulting video quality, and when designed well, this can improve both objective and subjective video quality. Building further on a concept from an optional feature of H.263+, the deblocking filter in the H.264/AVC design is brought within the motion-compensated prediction loop, so that this improvement in quality can be used in inter-picture prediction to improve the ability to predict other pictures as well. In addition to improved prediction methods, other parts of the design were also enhanced for improved coding efficiency, including:

Small block-size transform: All major prior video coding standards used a transform block size of 8x8, while the new H.264/AVC design is based primarily on a 4x4 transform. This allows the encoder to represent signals in a more locally adaptive fashion, which reduces artifacts known colloquially as "ringing". (The smaller block size is also justified partly by the advances in the ability to better predict the content of the video using the techniques noted above, and by the need to provide transform regions with boundaries that correspond to those of the smallest prediction regions.)

Hierarchical block transform: While in most cases, using the small 4x4 transform block size is perceptually beneficial, there are some signals that contain sufficient correlation to call for some method of using a representation with longer basis functions. The H.264/AVC standard enables this in two ways: 1) by using a hierarchical transform to extend the effective block size use for low frequency chroma information to an 8x8 array, and 2) by allowing the encoder to select a special coding type for intra coding, enabling extension of the length of the luma transform for low-frequency information to a 16x16 block size in a manner very similar to that applied to the chroma.

Short word-length transform: All prior standard designs have effectively required encoders and decoders to use more complex processing for transform computation. While previous designs have generally required 32-bit processing, the H.264/AVC design requires only 16-bit arithmetic.

Exact-match inverse transform: In previous video coding standards, the transform used for representing the video was generally specified only within an error tolerance bound, due to the impracticality of obtaining an exact match to the ideal specified inverse transform. As a result, each decoder design would produce slightly different decoded video, causing a "drift" between encoder and decoder representation of the video and reducing effective video quality. Building on a path laid out as an optional feature in the H.263++ effort, H.264/AVC is the first standard to achieve exact equality of decoded video content from all decoders.

Arithmetic entropy coding: An advanced entropy coding method known as arithmetic coding is included in H.264/AVC. While arithmetic coding was previously found as an optional feature of H.263, a more effective use of this technique is found in H.264/AVC to create a very powerful entropy coding method known as CABAC (context-adaptive binary arithmetic coding).

Context-adaptive entropy coding: The two entropy coding methods applied in H.264/AVC, termed CAVLC (context-adaptive variable-length coding) and CABAC, both use context-based adaptivity to improve performance relative to prior standard designs. Robustness to data errors/losses and flexibility for operation over a variety of network environments is enabled by a number of design aspects new to the H.264/AVC standard including the following highlighted features.

Parameter set structure: The parameter set design provides for robust and efficient conveyance header information. As the loss of a few key bits of information (such as sequence header or picture header information) could have a severe negative impact on the decoding process when using prior standards, this key information was separated for handling in a more flexible and specialized manner in the H.264/AVC design.

NAL unit syntax structure: Each syntax structure in H.264/AVC is placed into a logical data packet called a NAL unit. Rather than forcing a specific bitstream interface to the system as in prior video coding standards, the NAL unit syntax structure allows greater customization of the method of carrying the video content in a manner appropriate for each specific network.

Flexible slice size: Unlike the rigid slice structure found in MPEG-2 (which reduces coding efficiency by increasing the quantity of header data and decreasing the effectiveness of prediction), slice sizes in H.264/AVC are highly flexible, as was the case earlier in MPEG-1.

Flexible macroblock ordering (FMO): A new ability to partition the picture into regions called slice groups has been developed, with each slice becoming an independently-decodable subset of a slice group. When used effectively, flexible macroblock ordering can significantly enhance robustness to data losses by managing the spatial relationship between the regions that are coded in each slice. (FMO can also be used for a variety of other purposes as well.)

Arbitrary slice ordering (ASO): Since each slice of a coded picture can be (approximately) decoded independently of the other slices of the picture, the H.264/AVC design enables sending and receiving the slices of the picture in any order relative to each other. This capability, first found in an optional part of H.263+, can improve end-to-end delay in real-time applications, particularly when used on networks having out-of-order delivery behavior (e.g., internet protocol networks).

Redundant pictures: In order to enhance robustness to data loss, the H.264/AVC design contains a new ability to allow an encoder to send redundant representations of regions of pictures, enabling a (typically somewhat degraded) representation of regions of pictures for which the primary representation has been lost during data transmission.

Data Partitioning: Since some coded information for representation of each region (e.g., motion vectors and other prediction information) is more important or more valuable than other information for purposes of representing the video content, H.264/AVC allows the syntax of each slice to be separated into up to three different partitions for transmission, depending on a categorization of syntax elements. This part of the design builds further on a path taken in MPEG-4 Visual and in an optional part of H.263++. Here the design is simplified by having a single syntax with partitioning of that same syntax controlled by a specified categorization of syntax elements.

SP/SI synchronization/switching pictures: The H.264/AVC design includes a new feature consisting of picture types that allow exact synchronization of the decoding process of some decoders with an ongoing video stream produced by other decoders without penalizing all decoders with the loss of efficiency resulting from sending an I picture. This can enable switching a decoder between representations of the video content that used different data rates, recovery from data losses or errors, as well as enabling trick modes such as fast-forward, fast-reverse, etc.

Since the Real Time Transport Protocol (RTP) is the transport protocol for H.264/MPEG-4 video streams, the H.264 / MPEG- 4 packets are encapsulated by RTP frames.

4. CODEC SELECTION AND FACTORS AFFECTING QOS

Quality of Service (QoS) is the most important consideration for real time voice transmission. In this section, we will further investigate the main factors that have an essential effect on voice quality, which can be classified as CODEC selection, Voice Activity Detection (VAD) which is also known as Silence Suppression, delay, packet loss, jitter, echo, etc.

4.1 CODEC Selection

Codec selection is important to select a CODEC that meets the bandwidth and voice quality requirements.

G.711 is the preferred choise when bandwidth and cost are not an issue and is generally the default CODEC for Local Area Networks (LANs) because G.711 does not compress the audio.

G.729 A/B is generally the default CODEC for Wide Area Networks (WANs) because it requires less bandwidth than G.711 and delivers near-toll voice quality; for example, a G.729 A/B with a 30 ms sample size is an effective technique to reduce bandwidth limitations, while achieving acceptable voice quality.

G.723.1 is used when bandwidth, not voice quality is the main consideration.

Speech CODECs (compression algorithms), such as G.729 A/B and G.723.1, are designed to reduce the bandwidth required; however, when using these CODECs, parameters such as end-to-end delay (latency) and distorsion in voice quality should be considered. Although G.726 CODEC has less processing delay than G.729 A/B CODEC, G.729 A/B is generally used for VoIP because it delivers better voice quality. Table 4.1 shows some network requirements such as conversion time, Look Ahead Delay, Minimum Compression Algorithm Delay and Voice Quality of common voice CODECs.

Table 4.1: Network requirements of common voice codecs

CODEC	Conversion Time	Look Ahead Delay	Minimum Compression Algorithm Delay	Kbps	Voice Quality
G.711	.125 ms	None	.125 ms	64 Kbps	Toll quality
G.729 A/B	10 ms	5 ms	15 ms	8 Kbps	Near toll quality
G.723.1	30 ms	7.5 ms	37.5 ms	5.38/6.3 Kbps	Fair to good

4.2 Voice Activity Detection (VAD)

Most common voice conversations include about 50% of silence. Voice Activity Detection (VAD), also known as silence suppression, is a software application that lets a data network carrying voice traffic detect the absence of audio and converse bandwidth by preventing the transmission of silent packets over the network. Bandwidth savings vary depending upon the operating environment.

CODECs are available with and without VAD; for example, the G.729 A/B CODEC supports VAD, while G.711 does not.

The affects of VAD are unique to each call and is averaged over time and number of calls. The average utilization is more stable as the number of calls on a link increases, but a 30 to 40 % savings is typically assumed.

4.3 End-to-End Delay (Latency)

Although some CODECs have an higher compression algorithm delay (encoding and decoding time), all network experience some delay. The two types of delay are listed below:

- End-to-end delay (latency)
- Variable delay (jitter)

End-to-end delay is a non-variable (fixed) value. It is the time between the generation of a sound at one end and reception of the sound at the other end. Several factors that influence end-to-end delay are listed below:

- Processing: Time needed to encode, packetize and decode the transmission
- Propagation: Distance between source and destination

- Data Network Transmission: Link Speed (per network segment)

Delay should be less than 150 ms to avoid complaints, depending upon the types of applications used and customer requirements.

4.4 Variable Delay (Jitter)

Variable delay (Jitter) is dynamic and produces gaps in the conversation due to an uneven flow of data. Several factors that contribute to jitter are listed below:

- Performance of the network during peak conditions
- Packet contention for network devices
- Speed of links
- Size of voice or data packets
- Size of router buffers

Uncontrollable jitter impacts packet loss. An important technique to control jitter is to configure routers and switches to prioritize Voip packets ahead of other IP data packets.

In order to control jitter, buffer usage can be necessary; however, max delay of 150 ms for voice transmission must be guaranteed at the same time.

4.5 Packet Loss

Occasionally, packets fail to arrive at their destination and create gaps in conversation that degrade the voice quality; for example, clicks, muting, or unintelligible speech. Several factors that impact packet loss can listed as below:

- Congestion: The router discards packets to reduce congestion or the buffer overflows.
- Service Disruption: The network experiences an outage.
- Delay: Some packets take a longer route or experience delays that prevent them from reaching their destination at the appropriate time. (Packets which can not be delivered to destination within TTL (Time To Live) level, will be lost on the network.).

- CODEC Selection: Coding and decoding delay; for example, G.723.1 has a higher algorithmic delay than G.711, which can impact voice quality.

Packet loss varies depending upon the speech CODEC. Higher compression ratios increase the susceptibility of packet loss. Packet loss should be less than 1 percent.

4.6 Echo

Echo that is inaudible in the circuit-switched network can become noticeable with packet transmission because of the increased delay.

Round Trip Delay (RTD) is one of the main causes of echo. The maximum allowable amount of echo delay that VoIP can tolerate is 28 ms. It is important to detect and remove echo greater than 28 ms.

Factors that impact the severity of echo:

- Amplitude of echoed signal
- Time it takes signal to return to speaker

Devices to control echo:

- Echo canceller (Systems that use adaptive filters)
- Echo suppressor (By attenuating signal)

4.6.1 Echo canceller

An Echo Canceller is a voice operated device connected on the four wire side of a circuit. Its operating performance has been standardized in ITU-T G.165 [21]. However, rather than attenuating the echo, it calculates an estimate of what the echo will be, and then subtracts this from the actual returned signal. In this way, only the echo is removed, while the wanted signal is unaffected.

An echo canceller is a device that looks for echo (a delay on the return path that is strongly correlated with a signal seen on the incoming path).

An echo canceller uses an adaptive filter to calculate an estimate of the echo, then subtracts it from the actual return signal.

An echo canceller can improve the echo path loss of a connection by up to 26 to 30 dB with the adaptive filter. Any residual echo is removed using a non-linear processor that removes all signals below a certain threshold. Figure 4.1 shows the operating principles of an echo canceller. Echo canceller will cancel the echo on its local tail circuit, i.e. the echo being returned from the hybrid towards S_{in}.

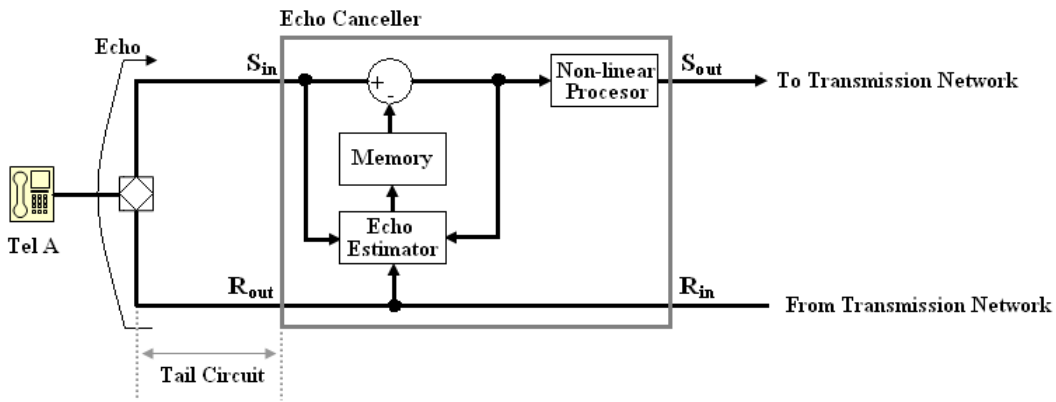


Figure 4.1 : Echo canceller

The echo canceller operates by developing a mathematical model of the echo that it anticipates will occur and subtracting this from the signalin the echo return path i.e. the signal into S_{in}.

Essentially this model is based upon two main factors: Firstly, the difference in the signal level out of R_{out} and the echo returned to S_{in}, also known as the Echo Return Loss (ERL), and secondly, the round trip delay of the tail circuit.

In practice, this cancellation process is not perfect, and a small amount of echo, known as residual echo usually remains at the output of the subtractor. To remove this residual echo, an additional function known as the non-linear processor (NLP), is used.

4.6.2 Echo suppressor

Echo suppressors are relatively simple devices used for echo control. They have been standardized in ITU-T Recommendation G.164 [20] and basically operates as follows: An echo suppressor, or voice switch, is a device that detects a signal on the incoming or outgoing path and switches attenuation into the other path to reduce the level of any returning signal. This suppression technique can be used in speakerphones, headsets, and wireless handsets.

An echo suppressor is designed to remove echo as shown in Figure 4.2. It does this by placing a large loss (attenuation) in the transmit path when it detects a signal on the receive path. While this reduces the level of returned echo, its main drawback is that it will also prevent the speech from Telephone A reaching the remote party when both parties are talking simultaneously (termed "double talking"). To reduce this effect, during double-talk, the echo suppressor must be able to operate in a second mode, where the transmit attenuation is reduced to a very small value and a small amount of attenuation is also placed in the receive path. In this way, while the received speech is now reduced in level slightly, the echo is effectively attenuated twice due to both the receive and transmit attenuation. Furthermore, during double-talk, the echo will tend to be masked out by the speech from Telephone A anyway.

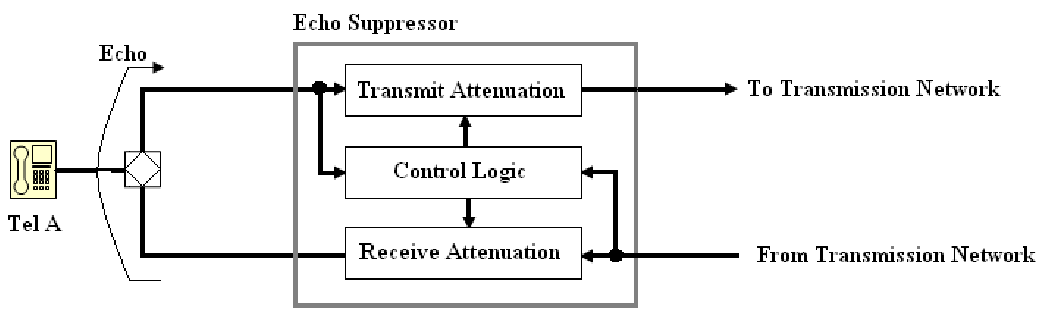


Figure 4.2 : Echo suppressor

Nevertheless, echo suppressors, while reducing echo, impair the quality of the speech communication, and for this reason have been superseded to a large extent by the superior echo canceller.

4.7 Common Models for Voice Quality

When the voice quality and delay budget profiles are outlined, it is important to monitor the voice quality of the VoIP network. Two common models used to measure voice quality are described in Table 4.2.

Table 4.2: Common models for voice quality

MEAN OPINION SCORE (MOS)	E-MODEL
Based upon subjective ratings	Based upon R-factor calculation
Generally used with circuit-switched environments	Generally used with VoIP environments
Uses a scale of five, where: 5 is considered unobtainable 4 is the lower boundary 3 or lower is clearly unacceptable	Scale ranges from 50 to 94, where: 94 is considered unobtainable 80 is recommended 70 is the minimum Below 50 is clearly unacceptable

Figure 4.3 compares the E-Model and MOS, where E-Model typically ranges from 50 to 94, while the MOS uses a scale of 5.

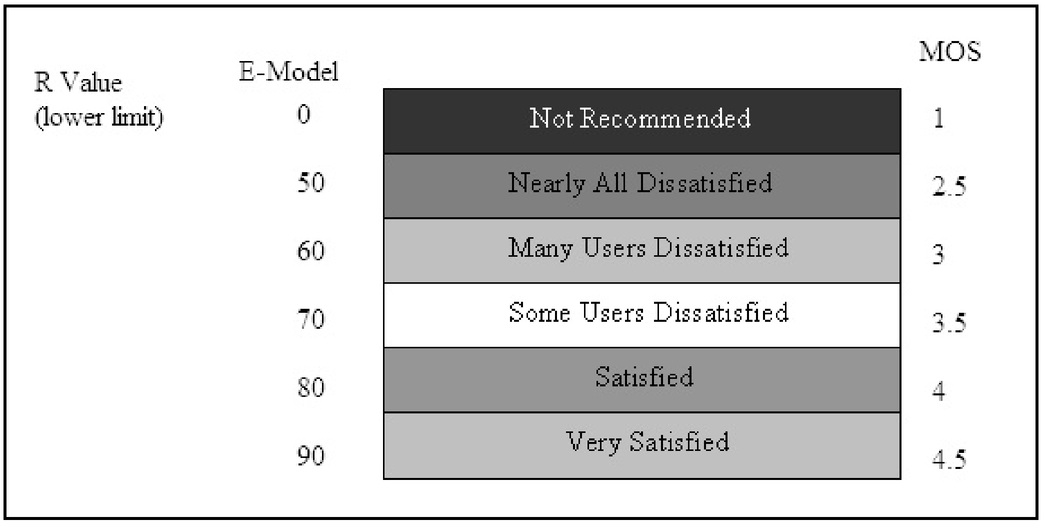


Figure 4.3 : Comparison of E-Model and MOS

The formula used to calculate the R-Factor is shown below:

$$R = R_0 - l_s - l_d - l_e + A \qquad \textbf{(4.2)}$$

Where;

R_0 = Base R-Value, such as noise

l_s = Impairments simultaneous to speech, such as echo

l_d = Impairments delayed after speech

l_e = Impairments due to the effects of special equipment, such as CODECs

A = Advantage factor, or user advantages, such as mobility

4.8 Bandwidth

Bandwidth is a significant parameter that affects QoS. The bandwidth required is directly related to the voice sample size and speech used by CODEC.

Generally the smaller the voice sample, the greater the number of packets needed. There are two types of bandwidth described below:

Available Bandwidth: With available bandwidth, the service purchased is typically available, but not guaranteed. For example, a customer purchased 256 Kbps. During light network traffic times, a user can achieve 256 Kbps, but under heavy network traffic conditions, the bandwidth is not achieved consistently. Many network operators oversubscribe the bandwidth on their network to maximize the return on their network infrastructure or leased bandwidth. This still does not ensure consistent bandwidth, as all users compete for available bandwidth.

Guaranteed Bandwidth: Some providers offer a Guaranteed Bandwidth service that guarantees a minimum bandwidth and burst bandwidth (duration of excess bandwidth use). With Guaranteed Bandwidth, subscribers get preferential treatment (QoS bandwidth guarantee) over the Available Bandwidth subscribers.

Figure 4.4 describes the formula to calculate the bandwidth per call basis.

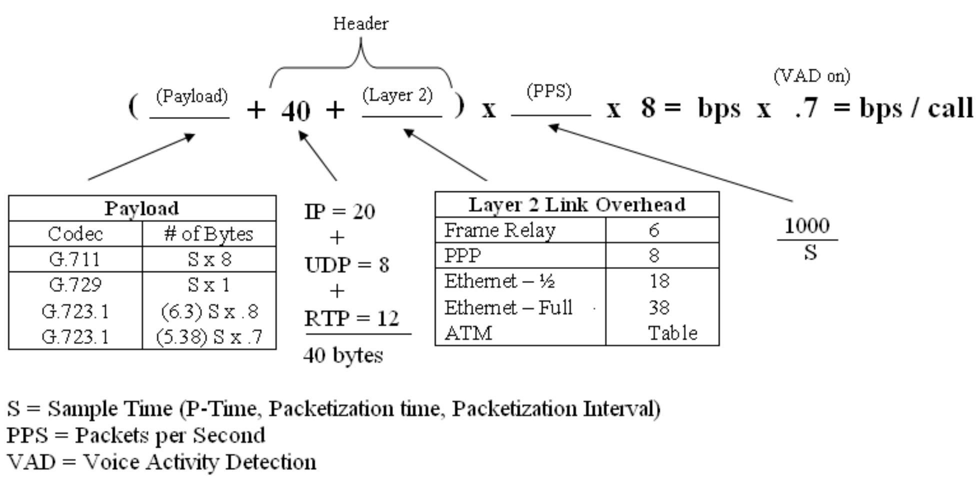


Figure 4.4 : Calculating bandwidth per call

4.8.1 Calculating bandwidth without Layer 2

This section describes how to calculate bandwidth for different compression standarts, such as G.711 and G.729 A/B, without the Layer 2 Data Link. To calculate the approximate bandwidth without Layer 2, the following formula should be used:

$$(Payload + IPHeader) \times PPSpercall \times 8bitsperbyte \qquad \textbf{(4.2)}$$

Where:

- Sample = Voice sample time measured in milliseconds (ms)
- Packet = Payload (bytes) + IP header (bytes)
- Payload = If G.711: Sample (ms) x 8 (CODEC compression value)

  If G.729 A/B: Sample (ms) x 1 (CODEC compression value)
- IP Header = 40 bytes (generally)
- Packets Per Second per call (PPS per call) = 1000 / Sample (ms)
- Effective bandwidth
- Bits per second (bps) / 1000 = kilobits per second (Kbps)

4.8.2 Calculating bandwidth with Layer 2

Some Layer 2 link technologies require more bandwidth than others. For example, assuming a 40-byte base IP header:

- Frame Relay: Add 6 bytes to the base header. If using longer Frame Relay addresses, this value is higher.
- PPP: Add 8 bytes to the base header size.
- Ethernet: Half-dublex Ethernet: Add 18 bytes to the base header size.

  Full-dublex Ethernet: Add 38 bytes to the base header size.

  If using 802.1Q tagging, this value is higher.

To calculate the approximate bandwidth for different compression standarts with the Layer 2, follwing formula should be used:

$$(Payload + IPHeader + L2) \times PPSpercall \times 8bitsperbyte \quad \textbf{(4.3)}$$

Where:

- Sample = Voice sample time measured in milliseconds (ms)
- Packet = Payload (bytes) + IP header (bytes)
- Payload = If G.711: Sample (ms) x 8 (CODEC compression value)

  If G.729 A/B: Sample (ms) x 1 (CODEC compression value)

  IP Header = 40 bytes (generally)
- Layer 2 Data Link Overhead = Frame Relay: Add 6 bytes

  PPP: Add 8 bytes

  Half-dublex Ethernet: Add 18 bytes

  Full-dublex Ethernet: Add 38 bytes
- Packets Per Second per call (PPS per call) = 1000 / Sample (ms)
- Effective bandwidth = bps / 1000 = Kbps

4.8.3 Calculating effective bandwidth with VAD

In voice transmission, most conversations include about 50 percent of silence. Voice Activity Detection (VAD), sometimes known as silence suppression, is a software application that lets a data network carrying voice traffic detect the absence of audio and conserve bandwidth by preventing the transmission of silent packets over the network.

CODECs are available with and without VAD; for example, the G.729 A/B CODEC supports VAD, while G/711 CODEC does not.

4.9 Methods to achieve QoS

QoS is an acrhitecture that delivers availability, reliability, and predictability for prioritizing traffic flows. Service Providers can use QoS to provide Service SLAs (Service Level Agreement) to customers and help manage their expectations.

Network operators achieve end-to-end QoS by ensuring that network elements apply consistent treatment to traffic flows as they traverse the network.

Some methods to achieve quality of service for VoIP traffic arelisted below:

- RSVP (Resource ReSerVation Protocol): Reserves dedicated portion of network's bandwidth.
- DiffServ (Differentiated Services): Defines different service classes and QoS mechanisms for packets.
- Ethernet 802: Provides congestion management at LAN port and user levels.
- Port-based Prioritization: Provides congestion at LAN port and user levels.
- IP Address Prioritization: Prioritizes packets originating from designated IP addresses; ideal for VoIP devices with static IP addresses.
- RTP Prioritization: Uses RTP port numbers for classification into a priority queue.
- VLANS: Separate voice and data traffic.
- Fragmentation: Divides packets before traversing bandwidth-limited connections.
- Policy Management: Prioritizes traffic via management or service policies.
- Traffic Shaping: Lets you determine which packets are dropped due to congestion and which packets receive priority.
- Queuing Mechanisms: Includes techniques, such as FIFO, PQ, CQ, FQ, and WFQ.

When implementing QoS, routing protocols should also be considered. Routing protocols can be very important when considering how VoIP calls are routed and how quickly fail-over occurs. When planning deployment of VoIP, the network designer must be aware of what types of situations trigger a routing table update with respect to the routing protocol. This helps to predict what path VoIP traffic can take when a network failure occurs.

Since we simulate the effects of selecting different queuing schemes over a wireline communication network, we will discuss different queuing mechanisms about how they work and how they prioritize the voice, video and data packets.

The purpose of queuing mechanisms is to alleviate congestion by managing how data is removed from the buffers and placed on the network.

4.9.1 First in first out (FIFO)

- Used traditionally for data traffic.
- Queues packets leaving the queue in the same order in which they arrive.

Since voice traffic has different requirements, other queuing mechanisms are sometimes used for congestion management to ensure the transmission of time-sensitive packets, such as voice.

4.9.2 Priority queuing (PQ)

- PQ uses multiple queues and classification to ensure that the most important traffic is serviced first.
- PQ works best when the majority of traffic sent across the interface can be separated into four distinct categories, from most to least important. This can be performed on a protocol, interface, host, or specific application basis.
- Lets the administrator identify and assign traffic to a high, medium, normal, or low priority queue.
- Services traffic in the high priority queue first and traffic in the lowest priority queue is last.

4.9.3 Custom queuing (CQ)

- Lets traffic be identified and assigned to a numbered queue to control how much data is sent each time to a particular queue.
- Sends up to 1500 bytes each time, by default.
- Lets the administrator to control the amount of bandwidth a queue consumes by raising or lowering the queue's threshold.

4.9.4 Fair queuing (FQ)

- Uses a round-robin approach, which prevents any one source from overusing its share of network capacity.
- Lets each flow passing through a network device have a fair share of network resources and prevents queues from not getting service because high-priority queues are being serviced.

4.9.5 Weighted fair queuing (WFQ)

- Provides better control of traffic flow across WAN circuits.
- Combines the functionality of priority queuing and fair queuing.
- Lets several sessions share the same link.
- Identifies logical conversations (high volume traffic between two peers) and ensures that those conversations do not monopolize the link.
- Ensures that low volume traffic is dispatched with other conversations.

Weighted Round Robin (WRR) and Weighted Fair Queuing (WFQ) are not recommended for VoIP traffic. A different method must be used to improve QoS, such as fragmentation. VoIP packets should be queued in a router or switch using a strict priority scheduler. If the router or switch does not support a priority scheduler (only supports a weighted scheduler), then the queue weight for VoIP traffic must be configured to 100 percent. If strict queuing is used, it is critical to engineer the WAN links supporting both the data and VoIP traffic with sufficient bandwidth. Queuing can cause significant CPU overhead.

5. OPNET (OPTIMIZED NETWORK ENGINEERING TOOL)

In this chapter, we will simulate three different scenarios. In scenario 1, effects of queuing schemes on a wireline network including voice (with voice CODEC G.711), video and data communication will be simulated and the service performance will be analyzed in terms of end-to-end delay and delay variation, comparing FIFO, Priority Queuing and WFQ.

In scenario 2, we will simulate the performance of the proposed network by changing the voice CODECs to G.711, G.711 silence, G.723.1 and G.729, respectively.

In scenario 3, we will implement a wireless network with a set of wireless clients to the existing wireline network and discuss the simulation results.

5.1 Why OPNET?

There are too many tools available in the market for simulating and measuring the traffic on a network as well as reporting the service performance of the voice, video or data communication on the simulated network in terms of traffic received/sent, end-to-end delay and jitter (delay variation). Few of the commercially available tools in the market are Cisco Netflow, Omegon, NetIQ Assessor, SNMPUtil and so on, Most of these tools have various functionalities including load measurement and identifying the bottlenecks in the network.

Networks are so complicated as per applications, protocols, device and link technologies, traffic flow and routing algorithms. There may be different configurations succeeding different characteristics and costs. Network design tools can help to find the correct combination of devices, protocols, routing algorithms and also show the network behaviour for a specific network configuration. OPNET is one of the well-known simulation tool which has advantages of anlyzing and simulating the network, when compared to other tools. It lets us to create a virtual network with different network elements (such as routers, switches, hubs, links, servers and vendor specific network elements) and supporting protocols and it is completely software based and can be run from any personal computer.

The simulation tool adopted in this dissertation is OPNET IT Guru Academic Edition 9.1. OPNET (Optimized Network Engineering Tool) is an object-orientated simulation tool for planning, modelling and performance analysis of simulation of network communication, network devices and protocols. OPNET Modeler has a number of models for network elements, and it has many different real-life network configuration capabilities. These makes real-life network environment simulations in OPNET very close to the reality and provide full phases of a study.

OPNET also includes features such as comprehensive library of network protocols and models, user friendly GUI (Graphical User Interface), data collection and analysis (graphical results and statistics). OPNET network modeling usually through three modeling hierarchical steps (Network modeling, node modeling and process modeling). First, a network topology needs to be defined include scale and size of the network (e.g., enterprise, campus, office and x span y span in degrees, meters, kilometers), the technologies need to be used (e.g., Ethernet, wireless), and nodes and links (e.g., 100Base, 1000Base). The node modeling deals with interrelation of processes, protocols and subsystems in and process mode describes the behaviour define the statistical features in a simulation model. Figure 5.1 shows an example workflow of OPNET Simulation Tool.

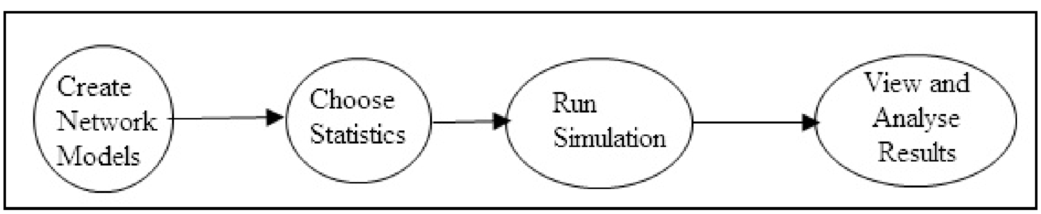


Figure 5.1 : Workflow of OPNET simulation tool

However, the current version of OPNET can only support SIP (Session Initiation Protocol) protocol, thus VoIP equipments such as VoIP gateway and gatekeeper product models are not included in OPNET, and this means the performance of VoIP gateway and gatekeeper are not measurable. Besides VoIP gateway and gatekeeper, OPNET can simulate voice traffic for both wired and wireless nodes. The statistical and graphical results for analysing the voice traffic transmission include the jitter, end-to-end delay, delay variation, and the amount of sent/received packets etc.

5.2 Network Elemets Responsible for Generating Background Traffic

5.2.1 Application definition object

Application Definition Object specifies applications using available application types. This is where the applications used in the network are defined. One can specify a name and the corresponding description in the process of creating new applications. For example, "Web Browsing (Heavy HTTP 1.1)" indicates a web application performing heavy browsing using HTTP 1.1. The specified application name will be used while creating user profiles on the "Profile Config" object. "Voice Encoder Schemes" specifies encoder Parameters (such as G.711, G.723.1, G.729, etc) for each of the encoder schemes used for generating voice traffic in the network.

In the proposed network there are three application servers, voice, video and data (ftp). Hence three applications are defined and configured on the application definition object. Table 5.1 shows a list of configurations on Application Definition Object to achieve voice, video and ftp communications in the network.

Table 5.1: Configuration on application definition object

APPLICATION	ATTRIBUTE	VALUE
VOICE	Silence Length (seconds)	default
	Talk Spurt Length (seconds)	default
	Symbolic Destination Name	Voice Destination
	Encoder Scheme	G.711
	Voice Frames per Packet	1
	Type of Service	Interactive Voice (6)
	RSVP Parameters	None
	Traffic Mix (%)	All Discrete
	Signaling	None
VIDEO	Frame Inter-arrival Time Information	10 frames/sec
	Frame Size Information	128x120 pixels
	Symbolic Destination Name	Video Destination
	Type of Service	Streaming Multimedia (4)
	RSVP Parameters	None
	Traffic Mix (%)	All Discrete
FTP	Command Mix (Get/Total)	50%
	Inter-Request Time (seconds)	constant (10)
	File Size(bytes)	Constant (1000000)
	Symbolic Server Name	FTP Server
	Type of Service	Best Effort (0)
	RSVP Parameters	None
	Back-End Custom Application	Not Used

5.2.2 Profile definition object

The "Profile Config" node can be used to create user profiles. These user profiles can then be specified on different nodes in the network to generate application layer traffic. The applications defined in the "Application Config" object are used by this object to configure profiles. Therefore, one must create applications using the "Application Config" object before using this object.

One can specify the traffic patterns followed by the applications as well as the configured profiles on this object. After the profiles are created, they can be applied to the workstations which would then access the application servers to generate the backgroung traffic.

In the proposed network, there are three profiles such as, VoIP_pro, VIDEO_pro and FTP_pro which are correlated with applications defined in Application Definition Object (VoIP_app, VIDEO_app and FTP_app). After creating profiles, workstations in the network are assigned to these profiles randomly to generate backgroung voice, video and data traffic. In Profile Dfinition Object, there are some attributes as described below:

- Profile Name: The symbolic name of the profile.
- Applications: The different applications that are executed within this profile and their characteristics.
- Operation Mode: It defines how applications will start. If set to serial ordered, all the application start one after each other in an ordered manner.
- Start Time: It defines when during the simulation the profile session will start.
- Duration: It defines the maximum amount of time allowed for the profile before it aborts. If set to end of simulation, the profile would abort only at the end of the simulation.
- Repeatibility: It specifies the parameters used to repeat the execution of the profile. When set to once at start time, the profile will not be repeated after its duration time.

Table 5.2 shows the profile list defined in Profile Definition Object.

Table 5.2: Profiles defined in profile definition object

Profile Name	Applications	Operation Mode	Start Time (seconds)	Duration (seconds)	Repeatabilit y
FTP_pro	FTP_app	Simultaneous	constant (100)	End of Simulation	Once at Start Time
VIDEO_pro	VIDEO_app	Simultaneous	constant (100)	End of Simulation	Once at Start Time
VoIP_pro	VoIP_app	Simultaneous	constant (100)	End of Simulation	Once at Start Time

In order to generate background traffic in the network, the following steps need to be taken respectively:

- Application definition Object needs to be modified to define the voice, video and ftp application using relevant voice encoder schemes.
- A new Profile Definition Object needs to be created to define three profiles with each one representing voice, video and ftp application respectively.

- Configuring servers to support required applications which are previously defined in Application Definition Object.
- Configuring LANs or workstations to support required profiles which are previously defined in Profile Definition Object.

In order to achieve that workstations generate traffic in the network, first, one of the pre-defined profiles are assigned to the workstations as Application Supported Profiles, and then applications which were defined on Application Definition Object are set as Application Supported Services on the servers.

5.2.3 QoS definition object

Qos Definition Object defines attribute configuration details for protocols supported at the IP layer. These specifications can be referenced by the individual nodes using symbolic names (character strings).

Queuing Profiles defines different queuing schemes such as FIFO, WFQ, Priority Queuing, Custom Queuing, MWRR (Modified Weighted Round Robin), MDRR (Modified Deficit Round Robin) and DWRR (Deficit Weighted Round Robin).

CAR (Committed Access Rate) Profiles defines different CAR profiles that can be used in the network.

Each queuing-based profile (e.g., FIFO, WFQ, PQ, CQ) contains a table in which each row represents one queue. Each queue has many parameters such as queue size, classification scheme (best effort, streaming multimedia, interactive voice, etc), RED (Random Early Detection) parameters , etc.

Note that the classification scheme can also be configured to contain many different criteria by increasing the number of rows. Some examples of setting queue priorities are:

- Weight for WFQ profile. Higher priority is assigned to the queue with a higher weight.
- Byte count for Custom Queuing profile. More traffic is served from the queue with a higher byte count.
- Priority label for Priority Queuing. Higher priority is assigned to the queue with a higher priority label.

The CAR profile defines a set of classes of service (COS). Each row represents a COS for which CAR policies have been defined.

5.3 Simulation Scenarios

5.3.1 Effects of using different queuing schemes on service performance

In this scenario, we simulate a network supporting voice, video and ftp traffic with FIFO, Priority Queuing (PQ) and Weighted Fair Queuing (WFQ). Figure 5.2 shows the diagram of our proposed network.

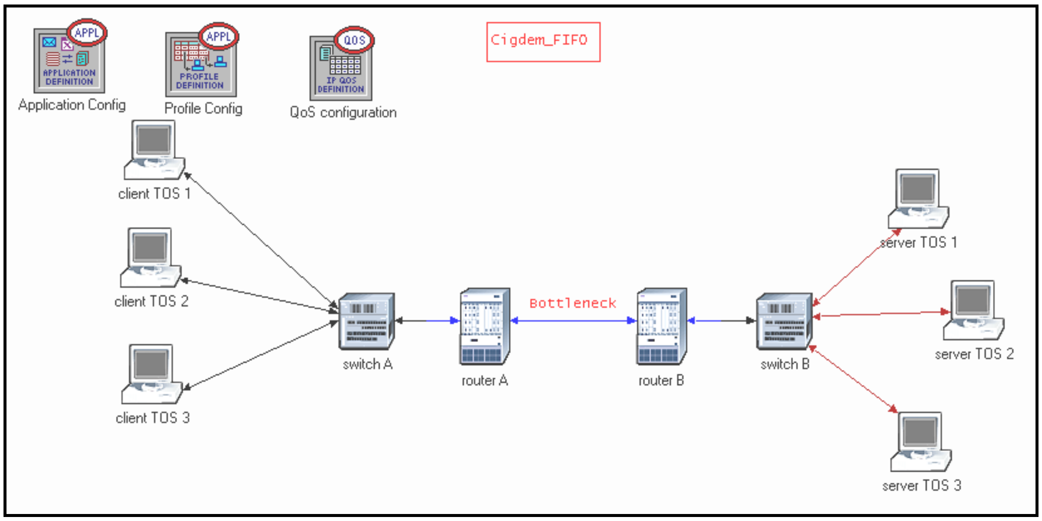


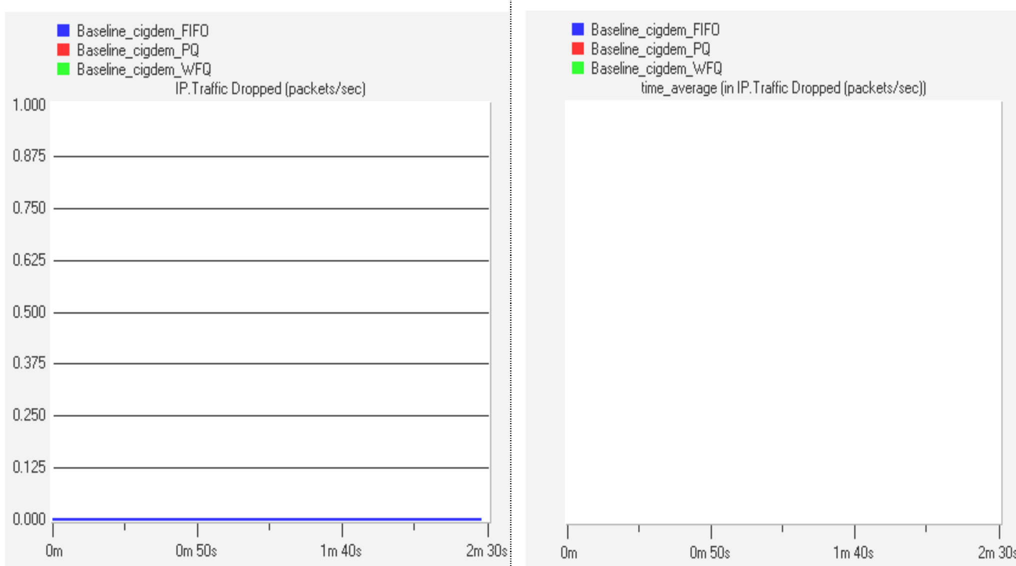
Figure 5.2 : Diagram of the proposed network

We can configure the link between the routers as it suppots different queuing schemes. Hence the routers can be configured in terms of QoS schemes. QoS specification parameters are available on a per interface basis on every router. The sub-attribute called "QoS info" in the "IP Routing Parameters" attribute is used to specify this information.

An incoming and outgoing CAR profile can be assigned to this interface as well as a queuing mechanism (FIFO, WFQ, PQ, CQ) with its queuing profile. Queuing Profiles are special schemes defining different queue configuration options. These are defined on the "QoS Attribute Configuration" object.

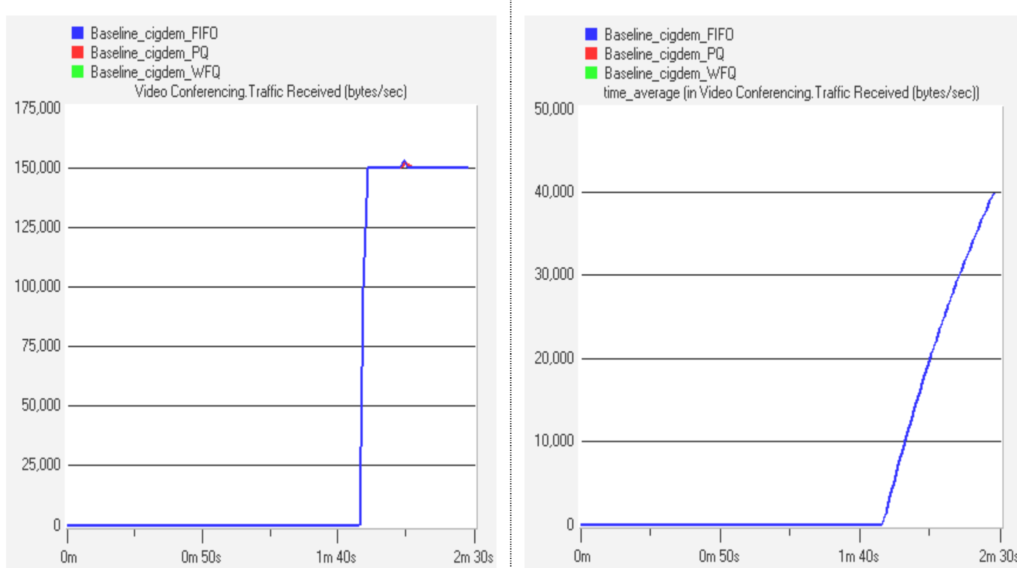
After selecting the specific queuing scheme, we can choose the indivudual statistics as IP-Traffic Dropped (packets/sec), FTP-Traffic Received (packets/sec), Video-Traffic Received (Bytes/sec), Voice-Packet Delay Variation, Voice-Packet End-to-

End delay and Voice Traffic received (Bytes/sec). Figure 5.3 and Figure 5.4 shows a graphical simulation result of IP Traffic Dropped (packets/sec) and Video conferencing Traffic Received (bytes/sec) in the network, respectively.



(a) IP Traffic Dropped (as is) | (b) IP Traffic Dropped (time average)

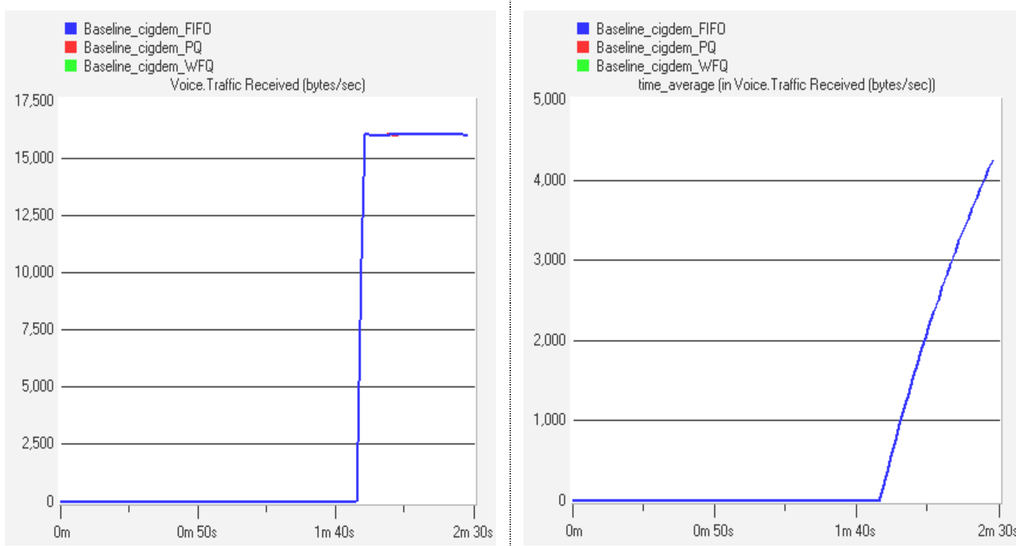
Figure 5.3 : IP traffic dropped (packets/sec)



(a) Video Conferencing Traffic Received | (b) Video Conferencing Traffic Received

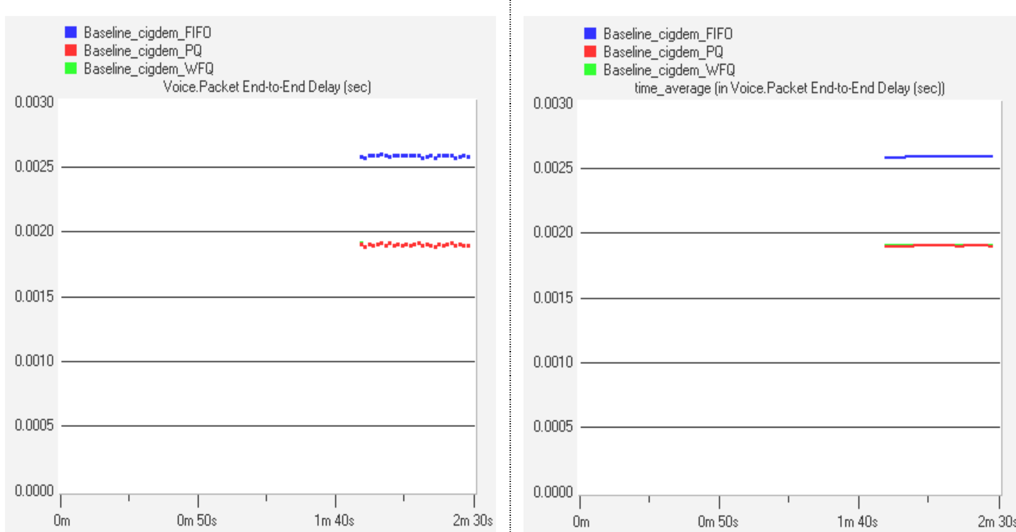
Figure 5.4 : Video conferencing traffic received (bytes/sec)

Figure 5.5 and Figure 5.6 represents the diagram of voice traffic in the network, Voice Traffic Received (bytes/sec) and Voice-packet End-to-End Delay (sec), respectively.



(a) Voice Traffic Received (as is) (b) Voice Traffic Received (average)

Figure 5.5 : Voice traffic received (bytes/sec) in the network



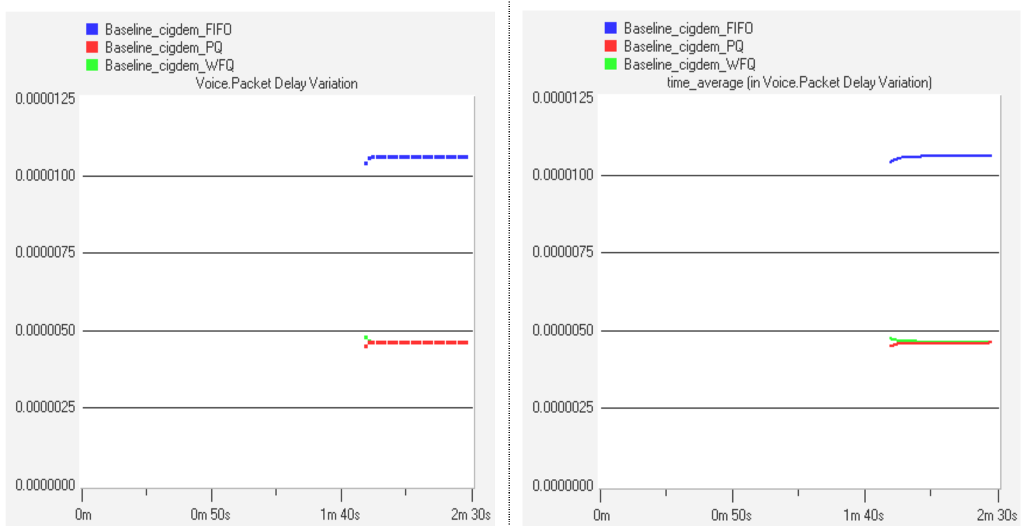
(a) Voice End-to-End Delay (as is) (b) Voice End-to-End Delay (average)

Figure 5.6 : Voice packet end-to-end delay (sec)

As discussed before in Chapter IV, voice end-to-end delay should be less than 150 msec. In Figure 5.6, Voice-packet end-to-end delay is in acceptable range when using Priority queuing and Weighted Fair Queuing, because we use some specific classification schemes in the QoS Attribute Configuration object, such as Best Effort

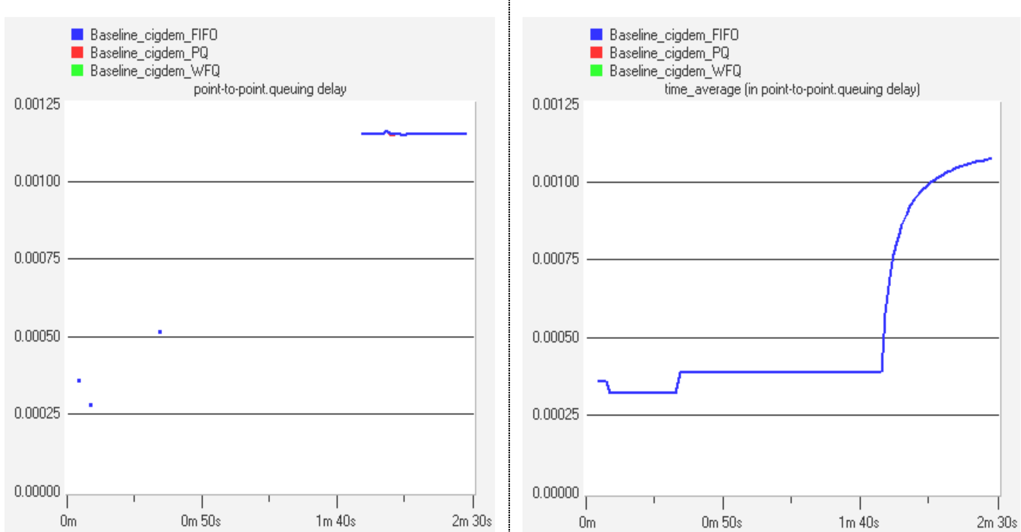
(0) for FTP traffic, Streaming Multimedia (4) for video traffic and Interactive Voice (6) for voice traffic. It can be seen that the delay shoots up at approximately 2 minutes. This indicates that some network elements have reached their maximum capacity to handle the traffic.

Voice-packet Delay Variation (jitter) and queuing delay graphics can be seen in Figure 5.7 and Figure 5.8, respectively.



(a) Voice Delay Variation (as is) (b) Voice Delay Variation (average)

Figure 5.7 : Voice packet delay variation



(a) Queuing Delay on Router (as is) (b) Queuing Delay on Router (average)

Figure 5.8 : Queuing delay for the link between west-router and east-router

As previously explained in Chapter IV, uncontrollable jitter (delay variation) impacts packet loss. An important technique to control jitter is to configure routers and switches to prioritize Voip packets ahead of other IP data packets, as we do so by giving the voice packets highest priority (Interactive Voice), while setting FTP traffic as Best Effort (0).

As seen in Figure 5.8, queuing delay in FIFO scenario is much more higher than that WFQ and Priority Queuing scenarios.

5.3.2 Effects of using different codec schemes on service performance

In this scenario, we simulate a network supporting voice, video and ftp traffic with different Voice Encoder Schemes, such as G.711, G.711_silence, G.729 and G.723.1. The diagram of our proposed network is shown in Figure 5.9.

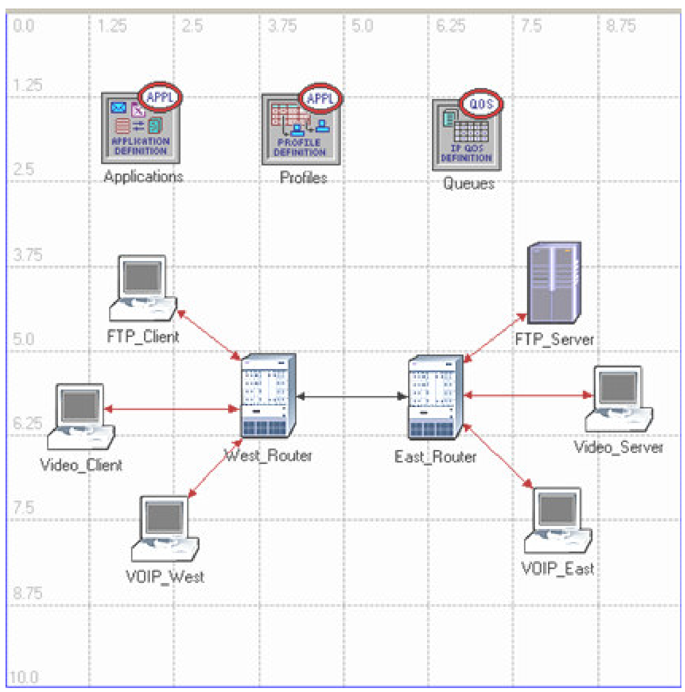
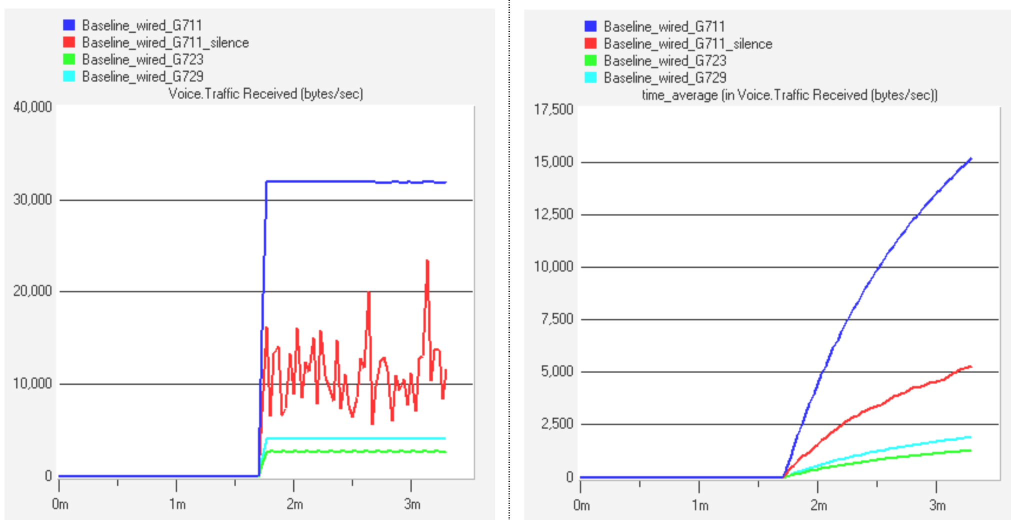


Figure 5.9 : Network diagram of scenario 2

We will analyze the graphics of Voice Traffic Received, End-to-End Delay and Delay Variation.

Figure 5.10 shows a diagram of Voice Traffic Received (bytes/sec).

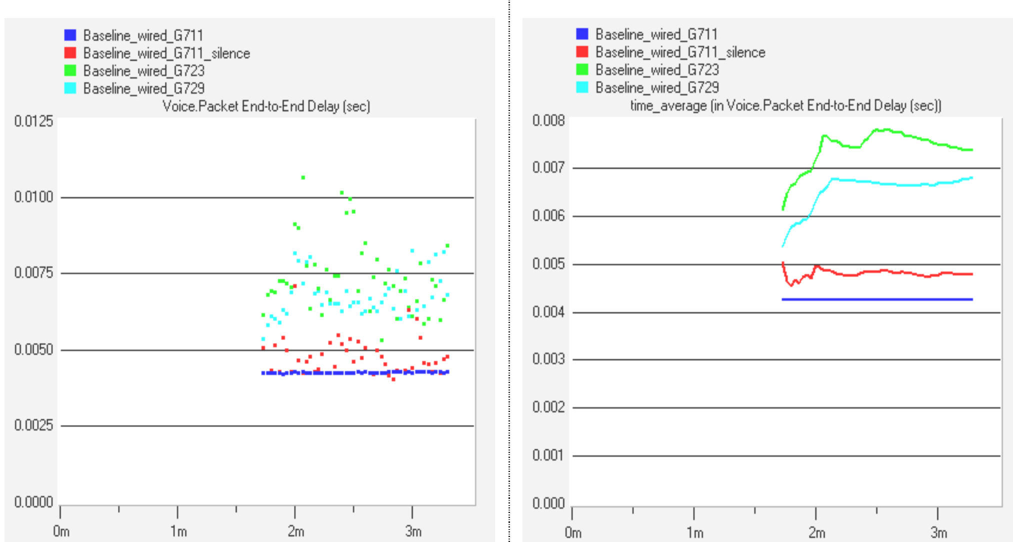


(a) Voice Traffic Received (as is) | (b) Voice Traffic Received (average)

Figure 5.10 : Voice traffic received (bytes/sec)

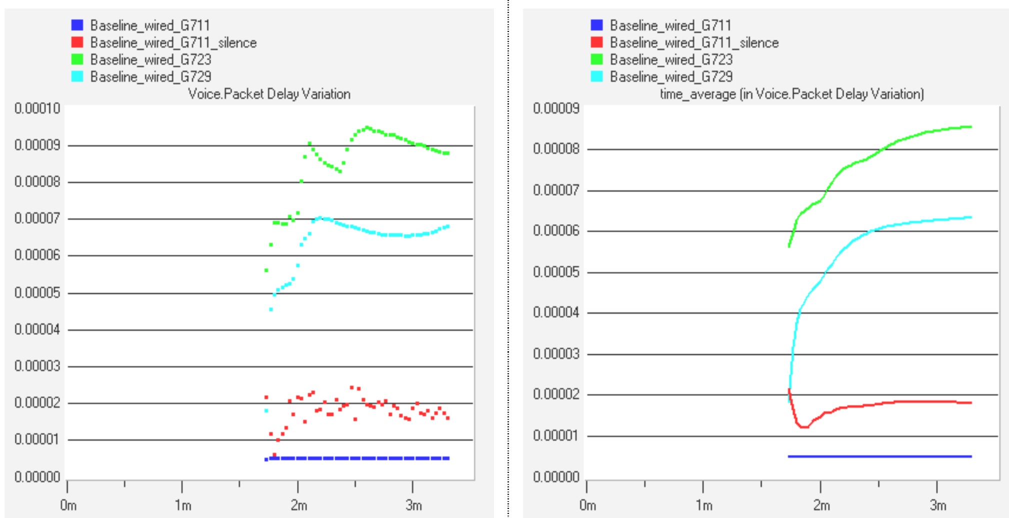
In Figure 5.11 and Figure 5.12, Voice End-to-End delay (sec) and Delay variation are represented, respectively. All voice encoder schemes are in range of the acceptable voice quality in terms of end-to-end delay, which is less than 150 msec.

The codec schemes' comparison values of voice delay variation (jitter) are shown in Figure 5.12. All these codec schemes yield acceptable voice jitters which are less than 1ms.



(a) Voice End-to-End Delay (as is) | (b) Voice End-to-End Delay (average)

Figure 5.11 : Voice packet end-to-end delay (sec)



(a) Voice Delay Variation (as is) | (b) Voice Delay Variation (average)

Figure 5.12 : Voice delay variation

5.3.3 Effects of implementing a wireless network to an existing network

In this scenario, we will simulate an integration of wireless and wireline networks with an Internet Cloud. IP Cloud attributes defined are Packet Discard Ratio and Packet Latency. Packet Discard Ratio specifies the percentage of packets dropped (ratio of packets dropped to the total packets submitted to this cloud multiplied by 100), while Packet Latency (secs) specifies the average latency (in seconds) experienced by packets traversing through this cloud.

Figure 5.13 shows a general diagram of this scenario with two subnets one representing wireless network and the other representing the wired network.

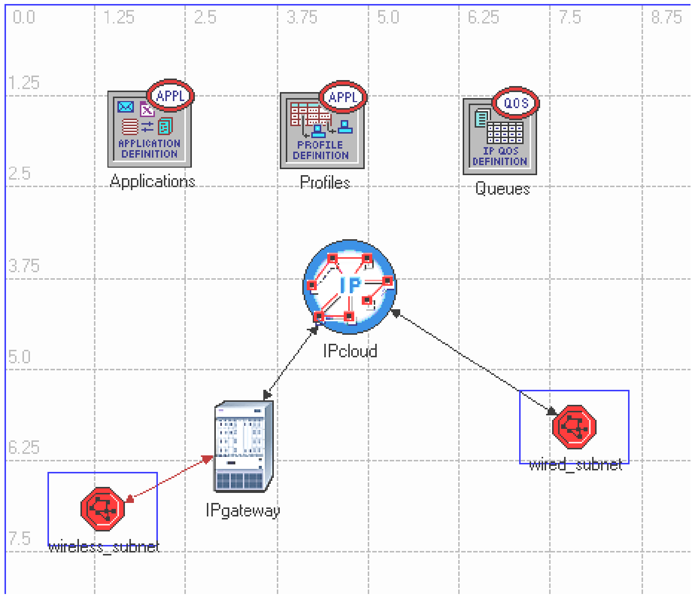
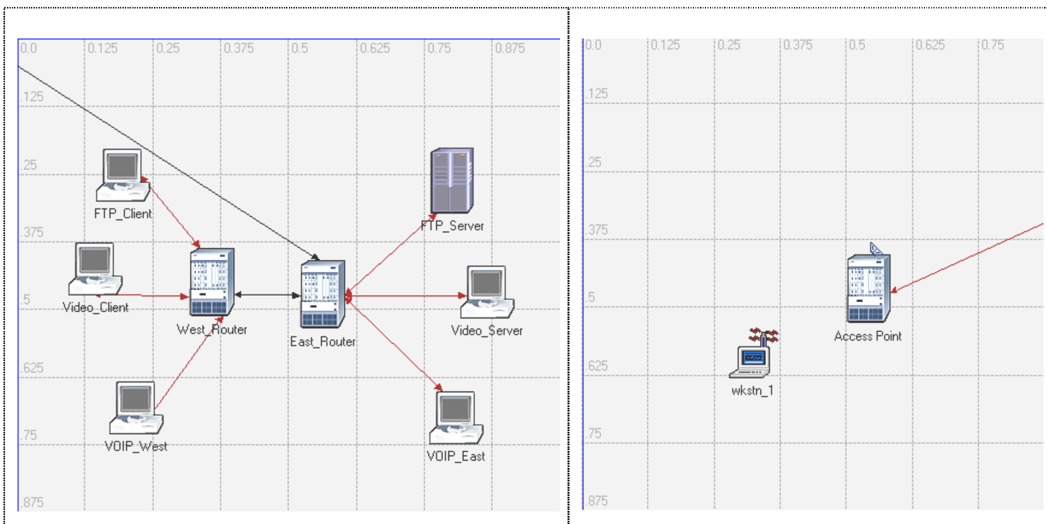


Figure 5.13 : General diagram of proposed network

Figure 5.14 (a) and (b) shows the structure of wired and wireless subnets in the proposed network.



(a) wired_subnet (b) wireless_subnet

Figure 5.14 : Network diagram of wired_subnet and wireless_subnet

When you edit the attributes of any wireless LAN node (station, client, server, access point) in the network, you will eventually see a set of Wireless LAN Parameters. A brief explanation and some suggested (or default) values are listed in Table 5.3.

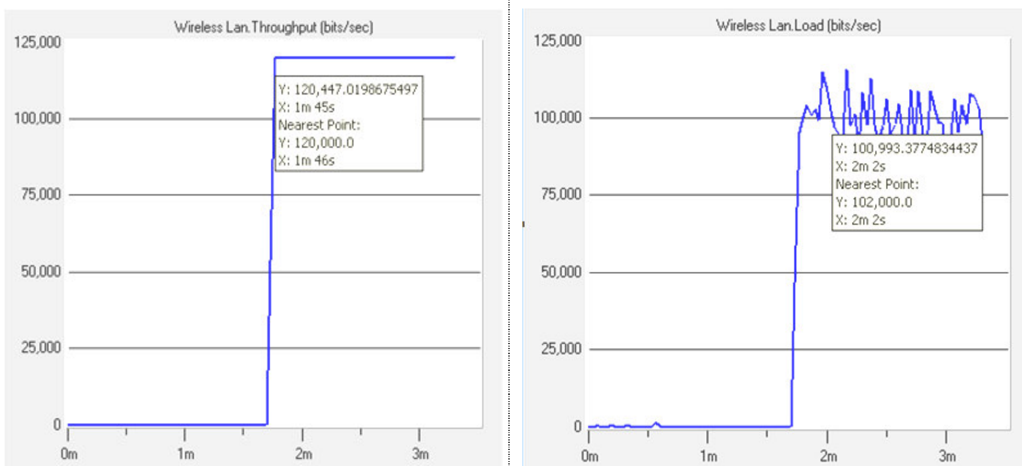
Table 5.3: Wireless LAN parameters

Parameter	Value	Description
WLAN MAC Address	Auto Assigned	Let OPNET automatically choose appropriate MAC addresses
RTS Threshold	None	If the data to send is greater than the RTS Threshold than RTS/CTS will be used; otherwise Basic Access will be used for that frame.
Fragmentation Threshold	None	If the data to send is greater than the Fragmentation Threshold than the data will be fragmented and sent as multiple frames. Note the comment about maximum size (read the *Help* by clicking on ?)
Data Rate	11Mb/s	Physical layer data rate. IEEE 802.11b has a maximum of 11Mb/s.
Physical Characteristics	Direct Sequence	Physical layer used. Direct Sequence is common in IEEE 802.11b.
Packet Reception - Power Threshold	7.33 E-14	The minimum power at which the receiver can successfully receive. The suggested value (with the default transmit power) gives a range of approximately 300m.
Short Retry Limit	7	Number of retransmission attempts for a frame using Basic Access before the MAC gives up and drops the frame
Long Retry Limit	4	Number of retransmission attempts for a frame using RTS/CTS before the MAC gives up and drops the frame
Access Point Functionality	Disabled	Whether this node acts as an AP or not
Channel Settings	(...)	Allows you to set the Bandwidth and Minimum Frequency of the channel (defaults to 1000 and BSS Based)
Buffer Size	1024000	Amount of data received from higher layer that can buffered before it is dropped. If the application (or traffic generator) sends too fast, then the buffer may fill up leading to drops.
Max Receive Lifetime	0.5	Time to wait to reassemble a frame if fragmentation is used
Large Packet Processing	Drop	If Drop then all packets larger than 2304Bytes are dropped (read the help; it interacts with Fragmentation)
BSS Identifier	Not Used	Needed if you have more than 1 BSS
PCF Parameters	Disabled	If enabled, PCF is used instead of DCF

We will analyze the impacts of increasing number of wireless clients under different voice encoder schemes, such as G.711 and G.729. We will also discuss the maximum number of wireless clients which can be added to the network, while keeping the service quality within the acceptable levels.

First we will simulate a network with 1 wireless client and a wireline network capable of handling voice, video and ftp traffic simulataneously.

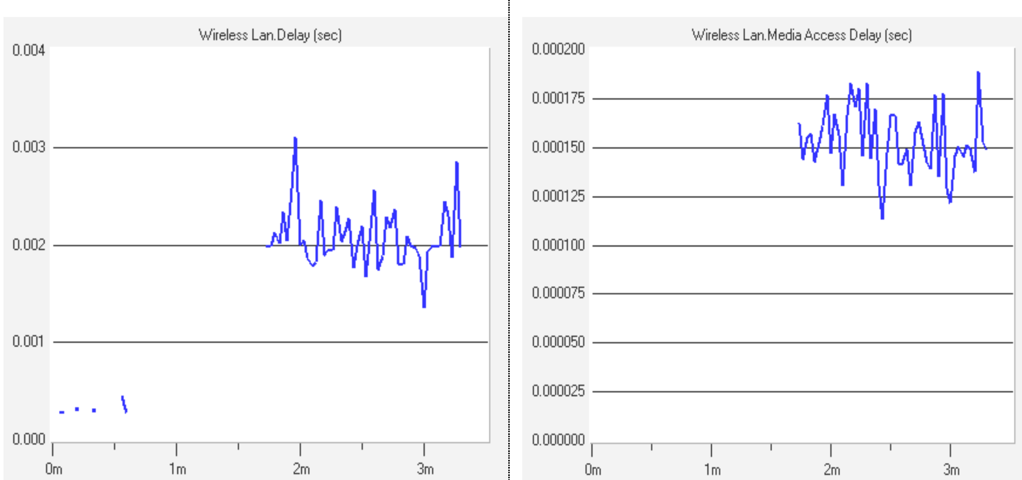
Figure 5.15 shows WLAN Throughput and WLAN Load with G.711 codec and 1 wireless client.



(a) WLAN Throughput (1 client) (b) WLAN Load (1 client)

Figure 5.15 : WLAN throughput and load for G.711 with 1 wireless client

Figure 5.16 shows WLAN Delay and WLAN Media Access Delay (sec).

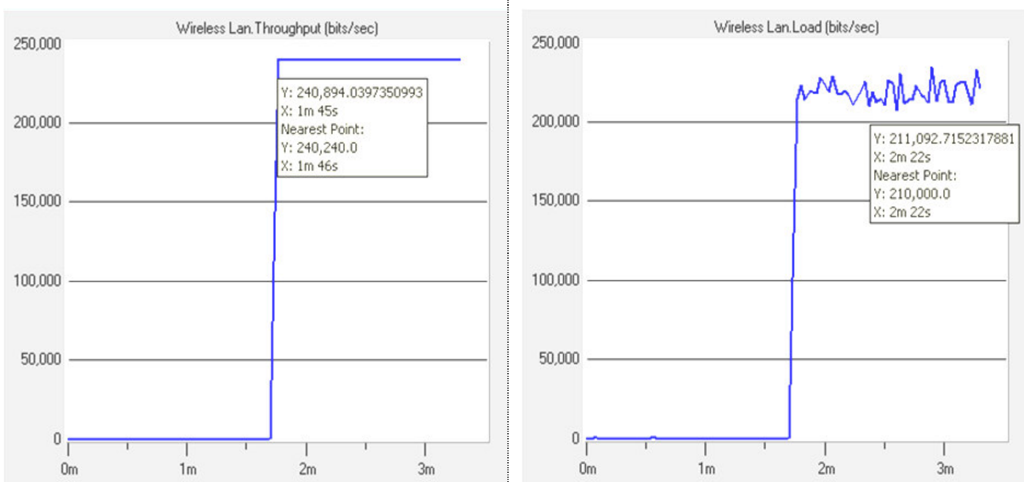


(a) wkstn_1 WLAN Delay (b) wkstn_1 WLAN Media Access Delay

Figure 5.16 : WLAN delay for G.711 with 1 wireless client

As seen in Figure 5.15 (a), WLAN Throughput is about 120 Kbps with G.711 encoder scheme, while WLAN delay in is acceptable level, 2.3 msec.

Now the number of wireless clients will be two and we will analyze the delay and throughput again. Figure 5.17 shows the WLAN Troughput and WLAN Load for 2 wireless clients added to the network.

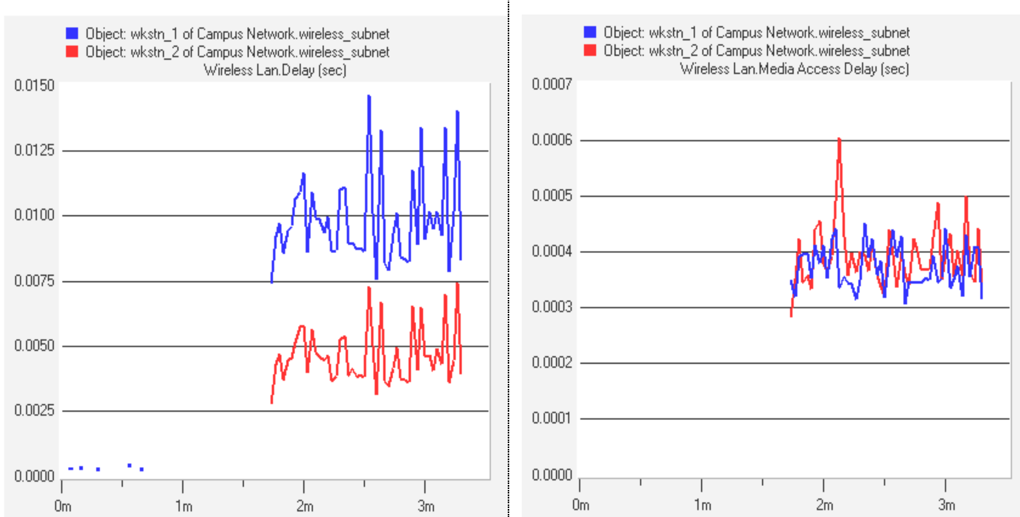


(a) WLAN Througput (2 clients) | (b) WLAN Load (2 clients)

Figure 5.17 : WLAN throughput and load for G.711 with 2 wireless clients

WLAN Throughput for 2 wireless clients is now 240 Kbps. WLAN delay is 10 msec for wkstn_1 and 5 msec for wkstn_2, which is still acceptable for voice quality.

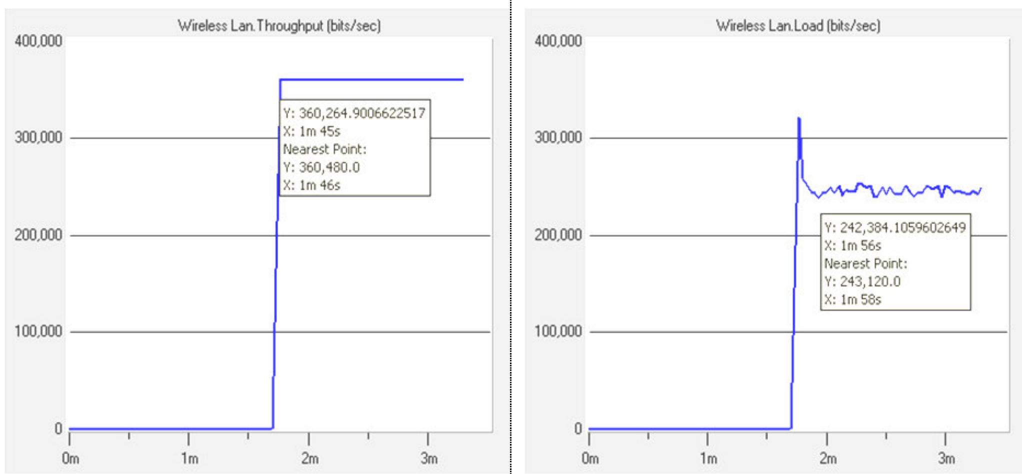
Figure 5.18 shows WLAN delay and WLAN media access delay in seconds.



(a) WLAN Delay (2 clients) | (b) WLAN Media Access Delay (2 clients)

Figure 5.18 : WLAN delay for G.711 with 2 wireless clients

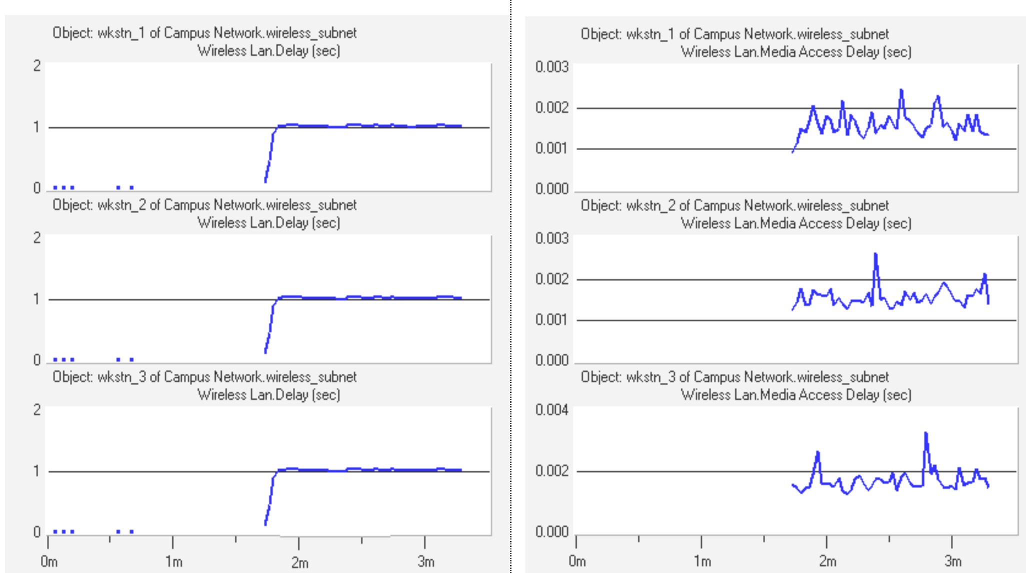
Figure 5.19 and Figure 5.20 shows the graphics for 3 wireless workstations with G.711. WLAN Throughput is now 360 Kbps as seen in Figure 5.19 (a).



(a) WLAN Througput (3 clients)	(b) WLAN Load (3 clients)

Figure 5.19 : WLAN throughput and load for G.711 with 3 wireless clients

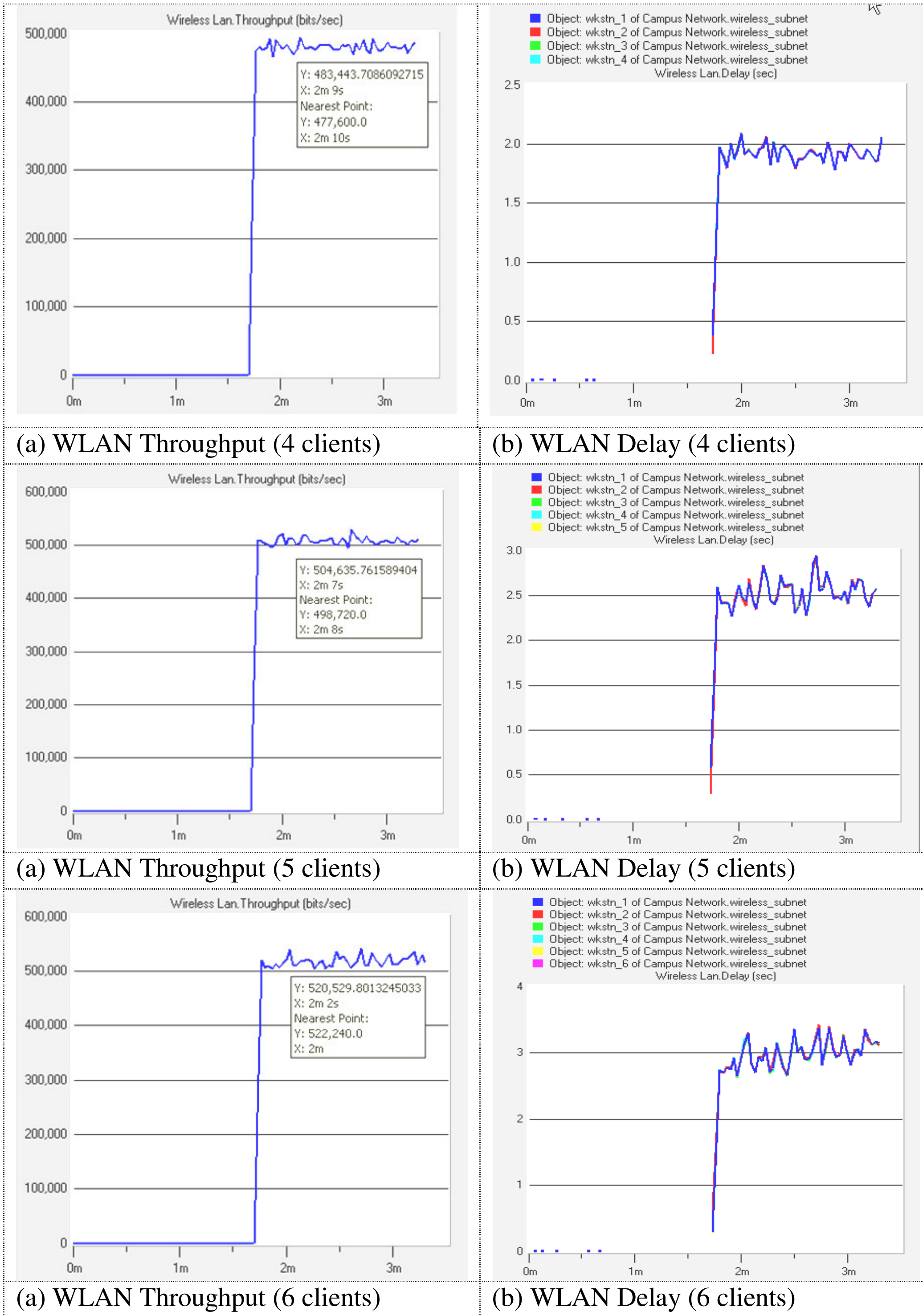
Figure 5.20 shows that WLAN delay reaches to an unacceptable level for a guaranteed voice quality. Delay measured for all wireless workstations is about 1000 msec which degrades quality of service.



(a) WLAN Delay (3 clients)	(b) WLAN Media Access Delay (3 clients)

Figure 5.20 : WLAN delay for G.711 with 3 wireless clients

Figure 5.21 (a)-(f) shows that WLAN Throughput is not increasing anymore when the number of clients reach 5. This means the network elements have reached their maximum utilization in the network.



(a) WLAN Throughput (4 clients)
(b) WLAN Delay (4 clients)
(a) WLAN Throughput (5 clients)
(b) WLAN Delay (5 clients)
(a) WLAN Throughput (6 clients)
(b) WLAN Delay (6 clients)

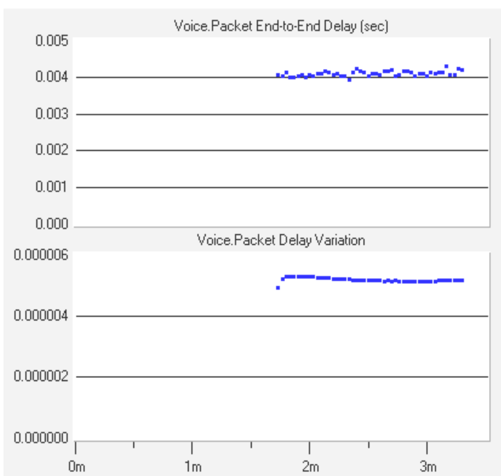
Figure 5.21 : WLAN throughput for G.711 with 4, 5 and 6 wireless clients

We see that some packets have been dropped in the router for congestion reasons. There might be two possible causes for that;

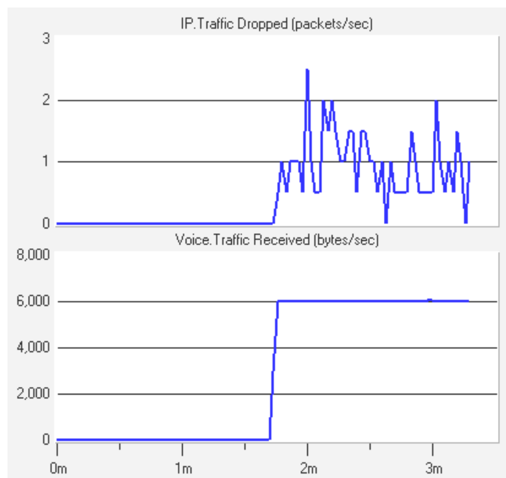
- The congestion attributes such as the maximum number of buffered packets or the queue sizes are too small.
- There is too much traffic on the interface and the link connected to this interface has a too small data rate.

Now, we analyze the same parameters with voice encoder scheme G.729. We will see that the number of wireless clients that meets the acceptable voice quality can now be increased up to 6 wireless workstations.

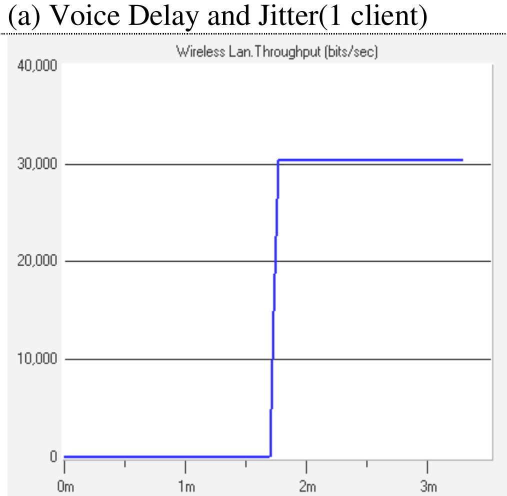
Figure 5.22 (a)-(d) shows the graphics for G.729 with 1 wireless client.



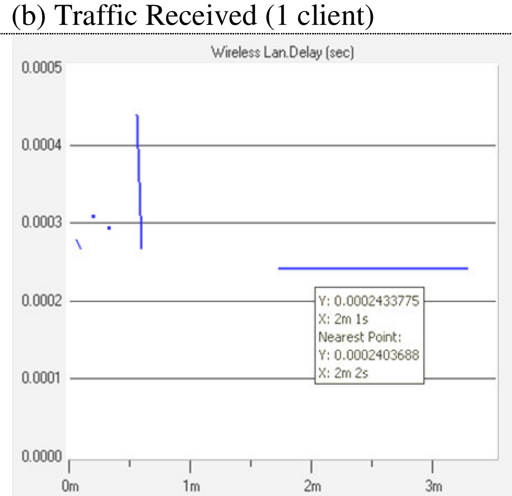
(a) Voice Delay and Jitter(1 client)



(b) Traffic Received (1 client)



(c) WLAN Throughput (1 client)

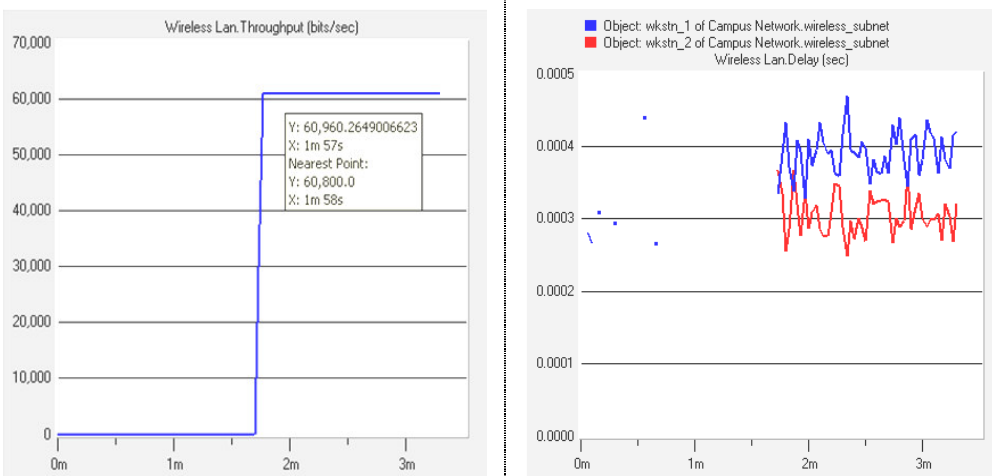


(d) WLAN Delay (1 client)

Figure 5.22 : WLAN througput and delay for G.729 with 1 wireless client

As seen in Figure 5.22 (c), WLAN Throughput is about 30 Kbps for 1 wireless client, while WLAN Delay is 0.2 msec which is in acceptable level.

Figure 5.23 shows WLAN Throughput and delay for 2 wireless clients.

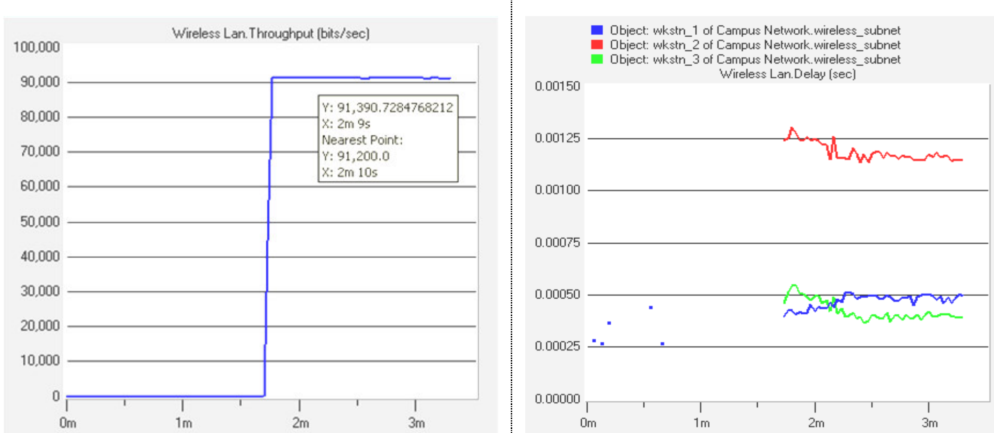


(a) WLAN Throughput (2 clients)	(b) WLAN Delay (2 clients)

Figure 5.23 : WLAN througput and delay for G.729 with 2 wireless clients

WLAN Throughput is 60 Kbps for 2 wireless clients. WLAN delay is about 0.4 msec for wkstn_1 and 0.3 msec for wkstn_2.

Figure 5.24 shows the graphics for G.729 codec with 3 wireless clients.

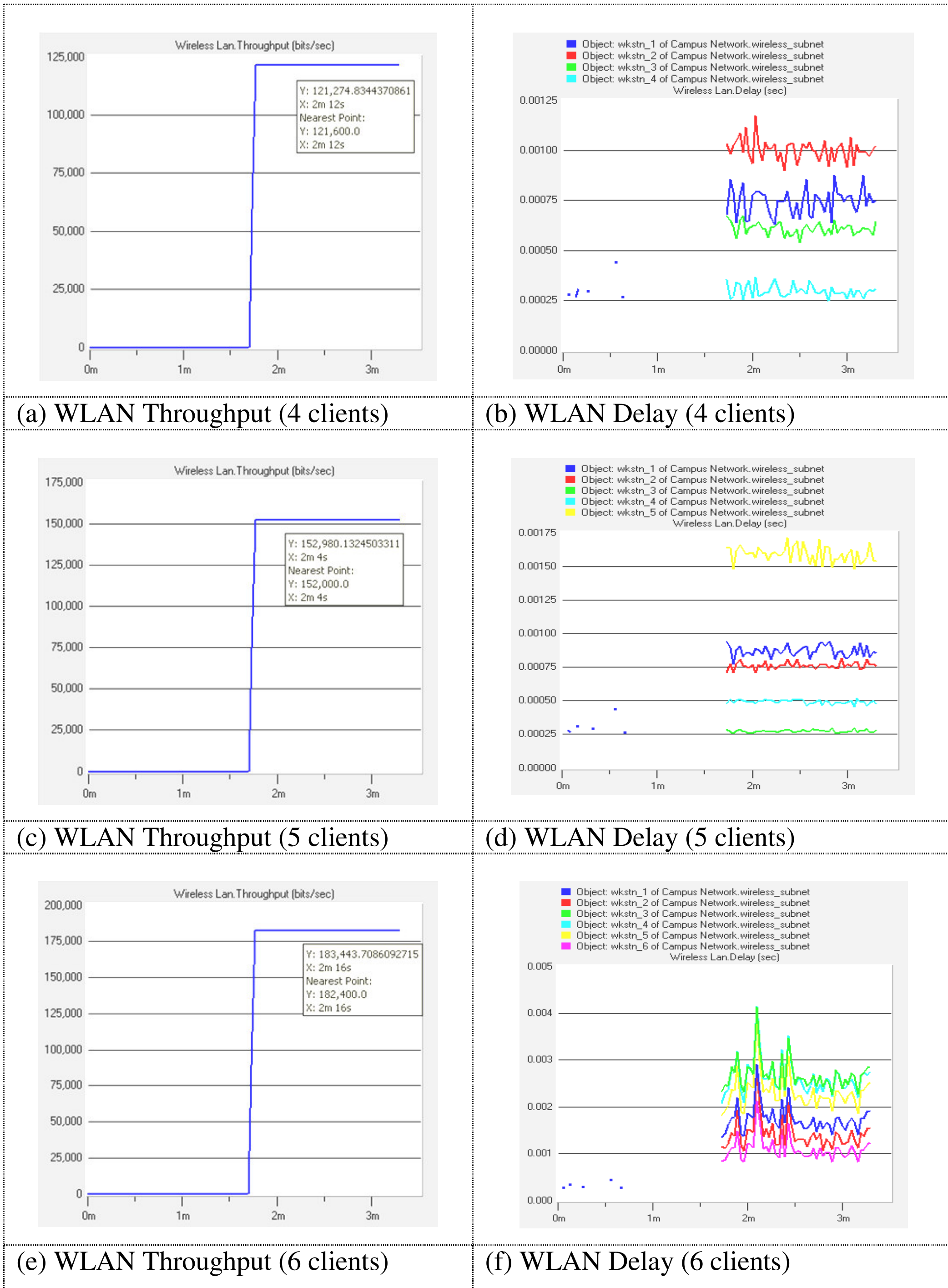


(a) WLAN Throughput (3 clients)	(b) WLAN Delay (3 clients)

Figure 5.24 : WLAN throughput and delay for G.729 with 3 wireless clients

WLAN Throughput is 90 Kbps for 3 wireless clients. Delay is 0.5 msec for wkstn_1, 1.15 msec for wkstn_2, 0.4 msec for wkstn_3.

Figure 5.25 (a)-(f) shows the graphics for 4, 5, 6 wireless clients, respectively.



(a) WLAN Throughput (4 clients) (b) WLAN Delay (4 clients)

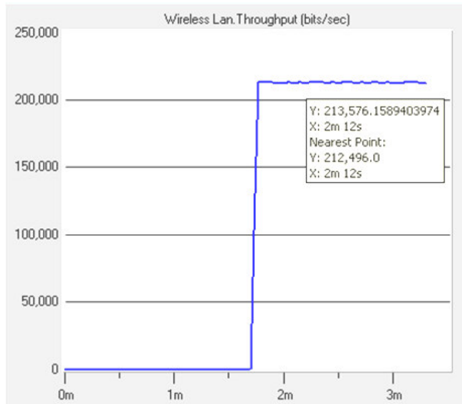
(c) WLAN Throughput (5 clients) (d) WLAN Delay (5 clients)

(e) WLAN Throughput (6 clients) (f) WLAN Delay (6 clients)

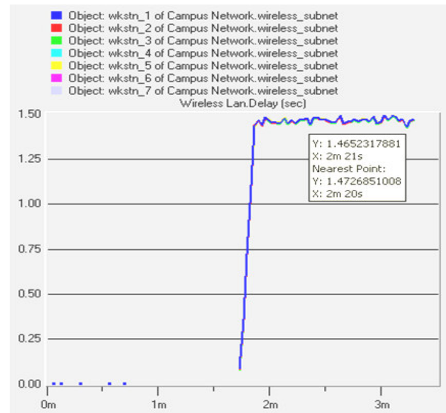
Figure 5.25 : WLAN throughput and delay for G.729 with 4, 5 and 6 clients

WLAN Throughput is 121, 150 and 180 Kbps for 4, 5 and 6 wireless clients respectively. Delay is still acceptable with 6 wireless clients.

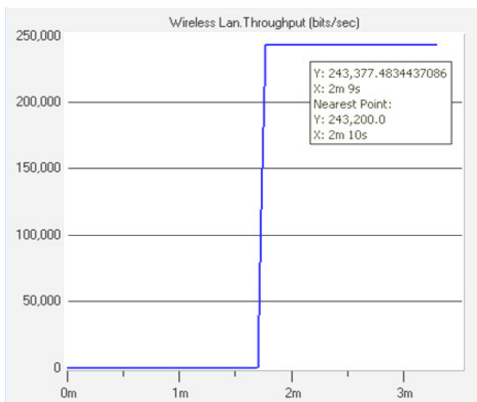
Figure 5.26 shows the graphics for the network with 7, 8 and 9 wireless clients.



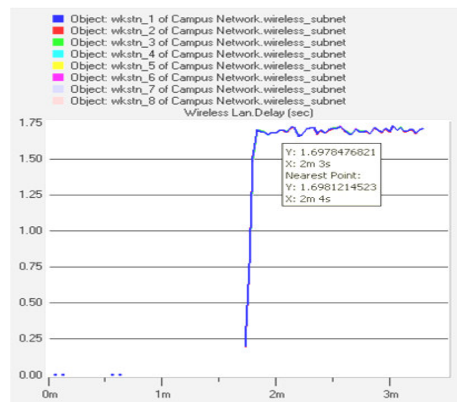
(a) WLAN Throughput (7 clients)



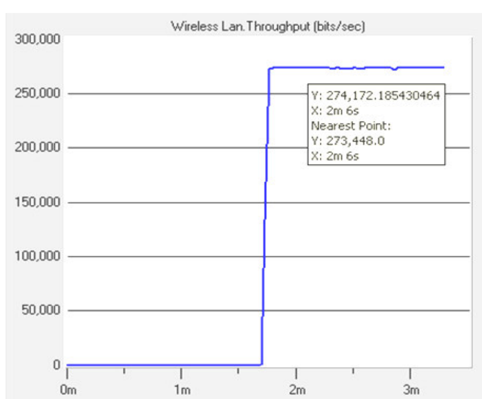
(b) WLAN Delay (7 clients)



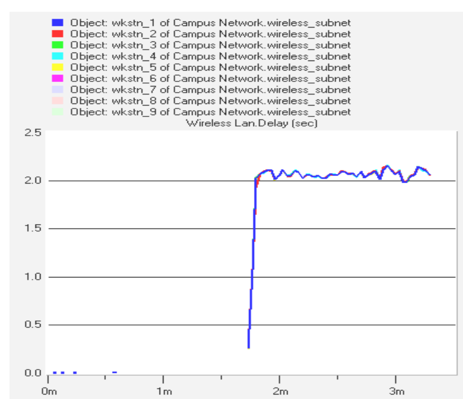
(c) WLAN Throughput (8 clients)



(d) WLAN Delay (8 clients)



(e) WLAN Throughput (9 clients)



(f) WLAN Delay (9 clients)

Figure 5.26 : WLAN throughput and delay for G.729 with 7, 8 and 9 clients

The summary of our simulation results is shown in Table 5.4.

Table 5.4: Summary of the simulation results

	WLAN Throughput (Kbps)	WLAN Delay (msec)										
	Access Point	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11
G.711	120	2.3										
	240	10	5									
	360	1000	1000	1000								
	480	2000	2000	2000	2000							
	504	2500	2500	2500	2500	2500						
	520	3000	3000	3000	3000	3000	3000					
G.729	30	0.2										
	60	0.4	0.3									
	90	0.5	1.15	0.4								
	121	0.75	1	0.6	0.3							
	150	0.8	0.75	0.27	0.5	1.6						
	180	2.3	2.2	3.3	3.2	3.1	1.6					
	210	1465	1465	1465	1465	1465	1465	1465				
	243	1697	1697	1697	1697	1697	1697	1697	1697			
	274	2069	2069	2069	2069	2069	2069	2069	2069	2069		
	304	2781	2781	2781	2781	2781	2781	2781	2781	2781	2781	
	304	2811	2811	2811	2811	2811	2811	2811	2811	2811	2811	2811
	332	4000	4000	4000	4000	4000	4000	4000	4000	4000	4000	4000

As seen in Table 5.4, WLAN Delay is becoming unacceptable after 3 wireless clients with G.711, while it is unacceptale after 7 wireless clients with G.729. Delay should be less than 150 msec for an acceptable voice quality. With G.711 encoder scheme, delay reaches to 1000 msec for each of the 3 wireless clients, which may lead to degredation in voice quality and WLAN thoughput is becoming stable after 5 wireless clients. This is probably due to the network elements reach their maximum capacity. With G.729 encoder sceheme, delay reaches to 1465 msec for each of the 7 wireless clients, which is quite higher than 150 msec. WLAN throughput is becoming stable after 11 wireless clients.

6. CONCLUSION AND RECOMMENDATIONS

This study evaluated the performance of VoIP over Local Area Networks through three different scenarios using OPNET simulation tool. This study also measured various real time communication parameters, such as packet end-to-end delay, delay variation (jitter), traffic received (bytes/sec) and IP packets dropped as well as the wireless LAN parameters (WLAN delay, WLAN media access delay and WLAN throughput). By using the OPNET simulation tool statistical and graphical analysis, one can determine how network elements can be deployed and how the parameters should be defined such as voice encoder schemes, queuing mechanisms, etc. Researchers and network designers can decide the optimum service parameters by analyzing OPNET simulation results and they can rearrange them in order to achieve the acceptable service quality. Due to the trade of between bandwidth requirements and quality of service, the optimum service parameters must be selected by using OPNET features.

In this thesis, we simulated a real communication network that associates non real-time data traffic with real-time voice and video traffic, which all sharing the network resources simultaneously. Since real time communication, such as voice traffic is more susceptible to the changing network conditions rather than non-real time data traffic, we use different queuing mechanisms such as FIFO, Priority Queuing and Weighted Fair Queuing. By giving different priority levels to the existing data, voice and video traffic with QoS Definition Object, we are able to prioritize the voice packets, thus improving the overall service performance in terms of voice packet end-to-end delay and delay variation (jitter). It can be seen that voice end-to-end delay is 2.6 msec for FIFO, while it is approximately 1.18 msec for PQ and WFQ. Delay variation is 0.0106 msec for FIFO, while it is 0.0035 msec for PQ and WFQ. Thus, Priority Queuing is more effective compared to FIFO queuing scheme.

It is also seen that voice encoder scheme has an undeniable impact on the performance of the system. G.711 delivers optimal voice quality (toll) but requires the most bandwidth. It is also the most resilient of packet loss and delay-tolerant, with least processing delay. G.729 is intended for low bandwidth connections, such as WAN and provides best balance of voice quality and bandwidth savings.G.729 requires less bandwidth than G.711 and delivers near toll quality of voice. It also uses silence suppression or Voice Activity Detection (VAD) in the transmission of voice communications that can further reduce bandwidth requirements. G.723.1 is intended for low bandwidth connections when voice quality is not a requirement. It provides most voice channels at the sacrifice of voice quality but has most processing delay due to the compression algorithm complexity. According to the simulation results, voice end-to-end delay is 4.2 msec, 6.2 msec and 7 msec for G.711, G.729 and G.723.1, respectively. Delay variation (jitter) is reaching 0.005 msec, 0.05 msec and 0.07 msec by changing the voice encoder scheme as G.711, G.729 and G.723.2, respectively.

We finally implement a wireless network to the existing wireline network and analyze the service quality in terms of end-to-end delay, delay variation and individual WLAN statistic such as delay, media access delay and throuhput. The number of VoIP clients has significant impact on VoIP performance for both wired and wireless LAN. Especially for wireless LAN, the impact of increasing the number of wireless nodes will be more than impact of the increasing the number of wired nodes. High number of wireless clients implemented to the network make the network elements (routers, switches, servers) reach their maximum capacity in terms of handling the existing voice, video and data traffic. According to the simulation results, we see that WLAN delay is becoming unacceptable after 3 wireless clients when using G.711 as a voice encoder scheme and IEEE 802.11b as a wireless medium. This value becomes unacceptable after 7 wireless clients when using G.729.

Also the parameter WLAN throughput is becoming stable after 5 and 11 wireless clients for G.711 and G.729, respectively.

This study only considered IEEE 802.11b wireless technology. IEEE 802.11g and IEEE 802.11n are suggested as future research. This study considered voice, video and data traffic. In future studies, more recent wireless technologies can be considered.

REFERENCES

Adda, M., Peart, A., Rochfort, L., 2006. Voice over IP on Wireless Networks Information and Communication Technologies, *International Conference on Telecom Technology and Applications '06*. Vol. **2**, 17-19 May.

Al-Sayyed, R., Pattinson, C. and Dcre, T., 2007. *VoIP and Database Traffic Co-existence over IEEE 802.11b WLAN with Redundancy*. World Academy of Science, Engineering and Technology 26.

Angrisani, L., Napolitano, A., Sona, A., 2005. VoIP over IEEE 802.11 wireless networks: Experimental analysis of interference effects. Electromagnetic Compatibility - EMC Europe, 2008 *International Symposium on Electromagnetic Compatibility*, 8-12 September.

Ast, J.D., 2007. The effect of dynamic voice codec selection for active calls on voice quality, *Master of Science*, Wichita State University, Kansas City, United States of America.

Baratvand, M., Tabandeh, M., Behboodi, A. and Ahmadi, A.F., 2008. *Jitter-Buffer Management for VoIP over Wireless LAN in a Limited Resource Device*., 2008. Fourth *International Conference on Networking and Services* 16-21 March.

Broniecki, T.A., 2009. A detailed analysis: VoIP comparisons of H.323 standard and SIP protocol in an 802.11g environment, *Master of Science*, University of Nebraska, Omaha, United States of America.

Chao H.C., Chu Y.M., and Tsuei T.G., 2001. *Codec Schemes Selection for Wireless Voice over IP (VoIP).* Springer-Verlag Berlin Heidelberg.

Das, S.K., Lee, E., Basu, K. and Sen, S.K., 2003. Performance optimization of VoIP calls over wireless links using H.323 protocol. *Computers*, IEEE, Vol. **52**, Issue 6.

Feigin, J., Pahlavan, K., 1999. Measurement of characteristics of voice over IP in a wireless LAN environment. IEEE *International Workshop: Mobile Multimedia Communications*, 15-17 November.

Gnassou, A., Frigon, J.F., Sanso, B., 2008. Impact of Wireless Channel on VoIP QoS and Admission Regions in IEEE 802.11g WLANs. *IEEE International Conference on Wireless and Mobile Computing, Networking and Communications*, 12-14 October.

Gowda, S.B, 2008. Simulation of voice over Internet protocol service performance when overlaid with background traffic over a local area network, *Master of Science*, California State University, Long Beach, United States of America.

Hirannaiah, R.M., Jasti, A. and Pendse, R., 2007. Influence of Codecs on Adaptive Jitter Buffer Algorithm. *66th IEEE, Vehicular Technology Conference* September 30.

Jiang, Y., 2006. Performance evaluation of SIP based voice conferencing over the IEEE 802.11 wireless networks, *Master of Science*, Simon Fraser University, Burnaby, British, Columbia, Canada.

Jianming Z. and Dacheng Y., 2006. Quality of VoIP in wireless communication networks. *Wireless Communications and Networking Conference,* IEEE Vol. **1**, 3-6 April.

Jin Y. and Ioannis K.G., 1999. *Wireless VoIP: Opportunities and Challenges*. Universal Mobile Telecommunication System Advanced Development, Lucent Technologies, Swindon, United Kingdom.

Liu, M., 2006. Quality of Service improvement schemes for real-time wireless VoIP, *Doctor of Philosophy*, The University of Arizona, Tucson Arizona, United States of America.

McNeill, K.M., Liu, M. and Rodriguez, J.J., 2006. An Adaptive Jitter Buffer Play-Out Scheme to Improve VoIP Quality in Wireless Networks, *Military Communications Conference*, 23-25 October.

Narbutt, M. and Davis, M., 2006. Gauging VoIP call quality from 802.11 WLAN resource usage, *World of Wireless, Mobile and Multimedia Networks International Symposium*,10-12 September.

Perkins, C., Hodson, O. and Hardman, V., 1998. *A survey of packet loss recovery techniques for streaming audio*, Network, IEEE Vol. **12**, pg 42-43.

Ramjee, R., Kurose, J., Towsley, D., and Schulzrinne, H., 1994. Adaptive playout mechanisms for packetized audio applications in wide-area networks. *13th IEEE INFOCOM Networking for Global Communications* Vol. **2**, 12-16 June.

Razvi M.D., 2007. Adaptive Packet Loss Concealment Mechanism For Wireless Voice over IP, *Innovative Algorithms and Techniques in Automation, Industrial Electronics and Telecommunications*, pg 30-33.

RFC-2198, 1997. RTP Payload for Redundant Audio Data, *IEEE,* USA.

RFC-3550, 2003. RTP: A Transport Protocol for Real-Time Applications *IEEE,* USA.

RFC-3261, 2002. SIP: Session Initiation Protocol *IEEE*, USA.

RFC-3263, 2002. Session Initiation Protocol (SIP): Locating SIP Servers *IEEE*, USA.

Riaz, A. and Chan, H.A., 2003. Wireless VoIP phone architecture and hardware requirements. *Industrial Electronics Society 29th* IEEE Vol. **3**, 2-6 November.

Sang K.K., Dong H.L. and Kyu O.L., 2009. The resource management method for providing VoIP service with guaranteed QoS in BcN. Advanced Communication Technology, *International Conference on Advanced Communication Technology*, Vol. **1**, 15-18 February.

Tianyi C. and Barry C., 2006. Error Concealment for Voice over Wireless LAN as applied to Converged Enterprise Networks, *Master of Science*, School of Computer Science, The University of Manchester, Manchester, United Kingdom.

Url-1 < *http://www.opnet.com*>, accessed at 15.11.2009.

Wei, W., Soung, C.L. and Li, V.O.K., 2005. Solutions to performance problems in VoIP over a 802.11 wireless LAN. *Vehicular Technology*, IEEE, Vol. **54**, pg 79-81.

Wilson, V., 2008. Analysis and simulation of Voice over Internet Protocol on a campus sized computer network, *Master of Science*, California State University, Long Beach, United States of America.

Xin G.W., Geyong M., Mellor, J.E., 2004. Improving VoIP application's performance over WLAN using a new distributed fair MAC scheme. Advanced Information Networking and Applications, *Advanced Information Networking and Applications International Conference*, 17-18 November.

Yiannakoulias, A. and Kuehn, E., 2005. Evaluation of header compression schemes for IP-based wireless access systems, *Wireless Communications*, IEEE Vol. **12**, pg 15-16.

Zhitao L., Min Y. and Huimin X., 2008. Performance Improvement of VoIP over WLAN via Packet Segmentation Strategy, *Wireless Communications, Networking and Mobile Computing*, 12-14 October.

CURRICULUM VITAE

Candidate's full name:	Zehra Çiğdem YEŞİL
Place and date of birth:	Bornova -1983
Permanent Address:	Ağaoğlu Eltes Güneşi Sitesi Ümraniye/İSTANBUL
Universities and Colleges attended:	İstanbul Üniversitesi, Electrical and Electronics Engineering Faculty