2003-09

Analysis of voice quality problems of Voice Over Internet Protocol (VoIP)

Tasyumruk, Lutfullah

Monterey, California. Naval Postgraduate School

NAVAL
POSTGRADUATE
SCHOOL

MONTEREY, CALIFORNIA

THESIS

ANALYSIS OF VOICE QUALITY PROBLEMS OF VOICE OVER INTERNET PROTOCOL (VOIP)

by

Lutfullah Tasyumruk

September 2003

Thesis Advisor:	Bert Lundy
Thesis Second Reader:	R. Scott Coté

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** September 2003	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE: Analysis of Voice Quality Problems of Voice Over Internet Protocol (VoIP)			**5. FUNDING NUMBERS**
6. AUTHOR(S) Lutfullah Tasyumruk			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE**

13. ABSTRACT

After its introduction in mid 90s, Voice Over Internet Protocol (VoIP) or IP telephony has drawn much attention. The prospect of cost savings on long distance and international toll calls, the global presence of Internet Protocol (IP), and the trend to converge data networks with voice networks have made VoIP one of the fastest growing telecom sectors. Additionally, the emergence of 3rd Generation (3G) cellular technology which offers high bandwidth will result in the convergence of the Internet and the cellular networks which will further stimulate the growth of VoIP. However, VoIP faces many problems mainly because of the nature of IP networks which were built to transport non-real-time data unlike voice.

This thesis analyzes factors affecting the voice quality of VoIP. These factors are delay, jitter, packet loss, link errors, echo and Voice Activity Detection (VAD). Further, implementation suggestions to lessen the effects of these factors are presented and finally, these suggestions are analyzed.

14. SUBJECT TERMS Voice over Internet Protocol, VoIP, IP Telephony, Internet Telephony, H.323, Voice Quality			**15. NUMBER OF PAGES** 95
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UL

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

ANALYSIS OF VOICE QUALITY PROBLEMS OF VOICE OVER INTERNET PROTOCOL (VOIP)

Lutfullah Tasyumruk
First Lieutenant, Turkish Army
B.S., Turkish Military Academy, 1997

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

from the

NAVAL POSTGRADUATE SCHOOL
September 2003

Author: Lutfullah Tasyumruk

Approved by: Bert Lundy
Thesis Advisor

R. Scott Coté
Second Reader

Peter Denning
Chairman, Department of Computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

After its introduction in mid 1990s, Voice over Internet Protocol (VoIP) or IP telephony has drawn much attention. The prospect of cost savings on long distance and international toll calls, the global presence of Internet Protocol (IP), and the trend to converge data networks with voice networks have made VoIP one of the fastest growing telecom sectors. Additionally, the emergence of 3rd Generation (3G) cellular technology which offers high bandwidth will result in the convergence of the Internet and the cellular networks which will further stimulate the growth of VoIP. However, VoIP faces many problems mainly because of the nature of IP networks which were built to transport non-real-time data unlike voice.

This thesis analyzes factors affecting the voice quality of VoIP. These factors are delay, jitter, packet loss, link errors, echo and Voice Activity Detection (VAD). Further, implementation suggestions to lessen the effects of these factors are presented and finally, these suggestions are analyzed.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I would like to express my gratitude to Professor Bert Lundy for his guidance, insights, and patience throughout the course of my thesis. Additionally, I am very thankful to R. Scott Coté for his invaluable contribution to my thesis.

I am also deeply grateful to my parents for having provided me with their best and to Turkish Army for giving me the opportunity to pursue a master's degree at NPS.

Further, I want to thank my beloved wife Gulay for her endless support and understanding during my studies. Without her support, it would not be possible to finish this thesis. Lastly, I would like to thank Idil, my little baby girl, for behaving well during my studies.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Voice over IP (VoIP), also known as IP telephony, packet telephony, Internet telephony, and voice conferencing, is the transmission of voice packets over packet-switched networks such as the Internet. After the introduction of first IP telephony software by VocalTec Inc. in early 1995 Ref [22], IP Telephony has drawn much attention. The global presence of IP, the trend to converge data networks with voice networks, and the prospect of savings for both individuals and companies made VoIP one of the fastest growing telecom sectors; however, there are a lot of challenges and problems to overcome.

Public- Switched Telephony Networks (PSTN) or traditional voice networks were built to provide service for voice applications which require low delay, low delay variance (jitter) and a constant bandwidth and they have been able to provide good toll quality for more than a century. Consequently, users have become accustomed to high service and voice quality provided by the PSTN and they expect similar quality from IP Telephony. But this is a very challenging task to achieve due to some factors which will be analyzed in this thesis.

Two major factors which affect voice quality in IP networks are *delay* and *jitter*. IP networks, unlike PSTNs, were built to support data. Data and voice transmissions over IP networks have very different requirements; data is not sensitive to delay and jitter, while voice is. For example, it is not a problem for an e-mail to arrive in two minutes to its destination, but a delay of more than a fraction of a second is not acceptable for voice communication. In fact, high quality voice communication cannot tolerate a delay of more than about one-tenth of a second.

Another factor affecting voice quality is that IP networks are subject to *packet loss* due to congestion which might happen especially during peak times when available bandwidth is saturated. Packet loss is not an issue for data packets which generally use Transmission Control Protocol (TCP). TCP provides reliability by retransmitting dropped packets. However, voice communication, which requires real-time transmission of packets, cannot use TCP's retransmission mechanism because of the delay.

Other factors affecting voice quality are *echo* and *compression*. To the user, echo is like hearing one's own voice from the telephone headset. Compression is a technique which reduces bandwidth usage by compressing voice signals.

A. SCOPE OF THIS THESIS

This thesis examines the factors affecting voice quality of VoIP and presents six specific implementation suggestions to lessen the effects of these factors. Further, six VoIP configurations are tested to demonstrate the effects of the suggestions. Lastly, the results of the tested configurations are analyzed. However, a complete solution for a successful VoIP implementation requires a wider consideration which includes the interoperability, security, Quality of Service (QoS) and transmission media issues. Those issues are beyond the scope of this thesis.

B. ORGANIZATION

This thesis is composed of six chapters. Chapter II presents an overview of VoIP, explaining the factors affecting the growth of VoIP and the main protocols used by VoIP. Chapter III discusses the factors affecting voice quality. Chapter IV explains the methods used to measure voice quality. Chapter V presents tests and analyzes implementation propositions. Chapter VI, the conclusion, mentions areas for possible further study.

II. VOIP OVERVIEW

This chapter presents the factors affecting the growth of VoIP and the main protocols used by VoIP. These are H.323, SIP, RTP, TCP, UDP and of course IP.

A. FACTORS AFFECTING THE GROWTH OF VOIP

The convergence of voice and data over the same network has long been a promising goal of both service providers and enterprises. The realization of this concept has the potential to bring tremendous cost savings and better control networks. Table 1 summarizes the factors affecting VoIP.

Factors Affecting The Growth of VoIP
1. Cost savings on long distance and international calls
2. Cost savings for businesses on intra business calls
3. Enhanced control over networks
4. User mobility
5. More efficient use of network capacity
6. New applications made possible
7. Introduction of 3rd Generation (3G) wireless technology

Table 1. Factors affecting the growth of VoIP.

The major driving force behind IP telephony for the corporate world is the savings in long distance and international toll calls. In the case of international calls, there are a lot of regulatory charges for the calls, but this costly situation does not apply to the unregulated packet switched networks such as internets which make VoIP a very inexpensive way of making voice calls. VoIP is also promising to reduce costs within a building or campus. Additionally, convergence of two separate networks will combine separate staffs (one for voice and one for data) into a single staff, simplifying the management of network and probably reducing the number of network staff. Another advantage of VoIP is that it does not require system reconfiguration when the employees

change their offices. An employee can have the same number wherever his or her IP phone plugs in anywhere on the network. This enhanced mobility for IP phone users is possible since IP phones are referenced via Ethernet MAC addresses unlike traditional phones which are connected to the specific ports on PBXs.

Another benefit of VoIP is that it is very bandwidth-efficient. The traditional circuit-switched calls reserve the entire dedicated bandwidth for the duration of the calls whether there is speech or not. Considering that nearly half of a typical human conversation is silence, much bandwidth is wasted transmitting silence. This makes circuit-switched calls very bandwidth-inefficient. VoIP packets, however, share the available bandwidth with other packets. Furthermore, traditional voice over circuit-switched networks uses Pulse Code Modulation (PCM) coding which consumes a bandwidth of 64 kbps while VoIP applications may consume as little as 18 kbps. Finally, with a technique called *Silence Suppression*, VoIP calls can save an additional bandwidth up to 50 percent. Silence suppression, also called Voice Activity Detection, is explained in Chapter III.

Yet another driving factor for VoIP is that combined voice and data offers new and enhanced ways of communication such as Unified Messaging and teleconferencing. *Unified messaging* is a solution which provides a single point of access to email, fax, and voice messages by using email, or the telephone. These new capabilities may increase productivity and effectiveness in the workplace.

Lastly, the introduction of 3G wireless technology which offers data rates up to 2 Mbps will be another stimulus for VoIP. The data rate of 3G wireless is almost 37 times faster than the speed of 56K modem and more importantly, it also offers the benefits of mobility. Additionally, 3G technology combines high speed mobile access with IP-based services such as VoIP. VoIP over 3G promises cost reductions in operational and capital expenditures for service providers as a result of using packet-switching instead of circuit switching.

All those driving factors contribute to the fast growth of IP telephony. Figure 1 illustrates the fact that 47% of all telephone minutes of use worldwide will be over IP networks by 2005. Already, some long-distance telecom companies are using the

Internet to route their calls to avoid paying local access fees to the local telecom companies. According to the Wall Street Journal (29 July 2003), AT&T, a major long-distance telecommunications company, routes some of its own calls through the Internet. According to current Federal Communications Commission (FCC) rules, calls routed through the Internet, instead of traditional phone lines, are exempt from access fees.

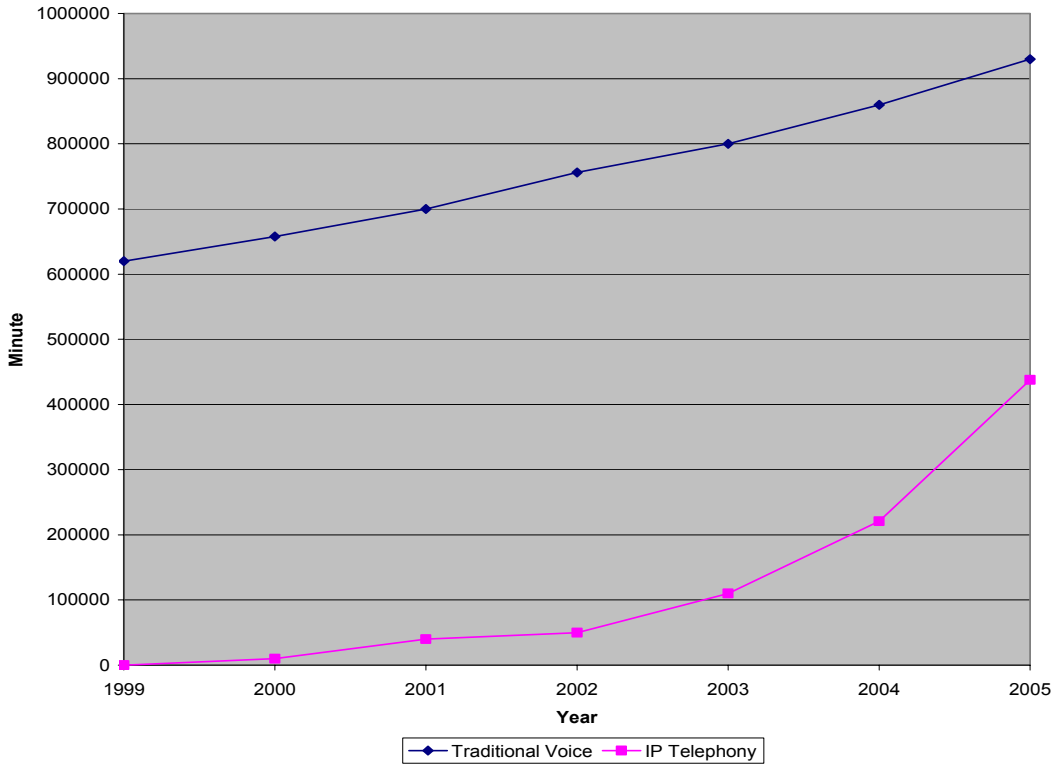

Figure 1. Worldwide traditional voice and IP Telephony minutes of use (After Ref[16])

B. VOIP TRANSPORT AND NETWORK LAYER PROTOCOLS

In order to better understand the VoIP signaling protocols H.323 and Session Initiation Protocol (SIP), it is necessary to have a basic understanding of the network and transport layer protocols that voice packets flow through. These protocols are Internet Protocol (IP), Transport Control Protocol (TCP), and User Datagram Protocol (UDP). Detailed information about header fields of these protocols will not be covered since it is not necessary to understand the VoIP signaling protocols.

In the current VoIP architecture, the analog voice is first digitized and optionally compressed to be sent over the IP network. After this phase, digitized voice is segmented and encapsulated with Real-Time Transport Protocol (RTP), UDP, and then IP headers to

be passed down to the link layer. The packet format of a voice packet is presented in Table 2. The total bytes of voice payload and link header fields differ depending on the codec and type of link layer.

Link Header	**IP Header**	**UDP Header**	**RTP Header**	**Voice Payload**
Variable size depending on link layer protocol	20 bytes	8 bytes	12 bytes	Variable size depending on codec

Table 2. Voice Packet

1. Internet Protocol (IP)

IP is the de facto standard for the transmission of data over the Internet and private networks. IP is a connectionless, best-effort protocol, that is, it does not require any prior call setup to exchange data between two hosts and it does not guarantee the delivery of packets to their destinations. Packets may be lost, duplicated, or arrive out of order. IP does not have any retransmission or error recovery mechanisms to fix such problems. It leaves these kinds of controls to TCP which will be explained further in paragraph 2. These low level characteristics of IP make it a fairly robust protocol for transmission of voice packets.

Even though IP is not the most efficient protocol for voice transmission, it is used for voice transmission since no other protocol has such a universal presence. However, IP has some drawbacks for real-time data transmission. For example, IP networks were built to support data transmission. But data and voice transmissions over IP networks have very different requirements; data is not sensitive to delay and jitter, while voice is. It is not a problem for an e-mail to arrive in two minutes to its destination, but a delay of more than 400 milliseconds is not acceptable for a VoIP application (Ref [2]).

2. TCP and UDP

TCP and UDP are both transport layer protocols within the OSI model; however, they have different characteristics.

TCP is a connection oriented and reliable protocol which compensates for the connectionless and unreliable IP. TCP provides reliability through the use of flow control, error checking, retransmissions, and sequencing. However it introduces delay as a result of its retransmission and congestion control mechanisms. The fact that TCP is reliable makes it a suitable protocol for non-real-time data transmissions since data is loss-sensitive and delay-tolerant. On the other hand, the fact that voice is loss-tolerant and delay-sensitive makes TCP an unacceptable mechanism for voice communications. It is never desirable to receive an e-mail corrupted; however, depending on the codec being used, VoIP applications can still function with up to 5 % packet loss (Ref [9]). Additionally, VoIP, which requires real-time transmission of packets as mentioned above, cannot use TCP's retransmission mechanism because a late voice packet equals a lost packet for a VoIP application. Therefore, the transport layer in VoIP stack uses UDP. VoIP uses TCP only for signaling messages to ensure the reliability of the setup of a call.

In contrast to TCP, UDP is a connectionless best-effort protocol with no retransmission mechanism and therefore a very good mechanism for multimedia communications. UDP is generally used when the reliability of TCP is not required. UDP is a simple protocol compared to TCP. UDP header has only four fields which are source port, destination port, length, and checksum. The Length specifies the total length of UDP header and data. The optional checksum field provides packet integrity checking. Actually, what UDP does is just to add a port addressing to the capability of IP in order for the packets to find the intended upper-layer application.

To summarize, there are two mechanisms to consider: TCP/IP and UDP/IP. TCP/IP is used if reliability is more important than delay and UDP/IP is used if delay is not acceptable and retransmission of lost packets is not required.

C. H.323

The International Telecommunication Union (ITU) *H.323* Recommendation, approved in 1996, is a peer-to-peer protocol suite which provides a foundation for voice and video communications over packet-based networks such as Ethernet LANs and the Internet. H.323 is the most widely used standard for VoIP.

After the first version was approved in 1996, VoIP began making progress and new requirements emerged accordingly. As a result version 2 was approved in January 1998. Version 2 introduced a new method called *Fast Connect* or *Fast Start* which is a new way of call setup which bypasses some previous steps, making the connection faster. Version 3, which made a few improvements, was approved in September 1999. Version 4, approved in November 2000, contains improvements in reliability, scalability, and flexibility. Version 5, approved in July 2003, is the latest version. More on the enhancements of the versions can be found at reference [19].

H.323 is in fact part of the H.32X family specified by ITU-T as shown in Table 3. H.323 assumes that the transmission medium is a LAN without guaranteed Quality of Service (QoS). Other recommendations in the H.32X family address other types of networks such as ISDN and PSTN as presented in Table 3. The interoperability between H.323 and these standards is made possible via *gateway*. Gateway is one of the basic components of an H.323 network which provides interoperability between dissimilar networks. Figure 2 shows this interrelationship.

Recommendation	**Network Type**
H.320	Integrated Services Digital Network (ISDN)
H.321	Broadband ISDN (B-ISDN)
H.322	LAN with guaranteed Quality of Service (QoS)
H.323	Packet Network such as LAN
H.324	Public Switched Telephone Network (PSTN)

Table 3. H.32X Family

1. H.323 Components

The four basic components of an H.323 network are *Terminal, Gateway, Gatekeeper*, and *Multipoint Control Unit*. These components provide point-to-point and point-to-multipoint multimedia communications. A *terminal* is a user interface to make real time communications with other H.323 endpoints. A *gateway* is a bridge between an

H.323 network and a non-H.323 network. A *gatekeeper* provides address translation and access control for H.323 endpoints. A *multipoint control unit* provides the capability to make multipoint conferences for three or more terminals or gateways. Figure 2 shows these components

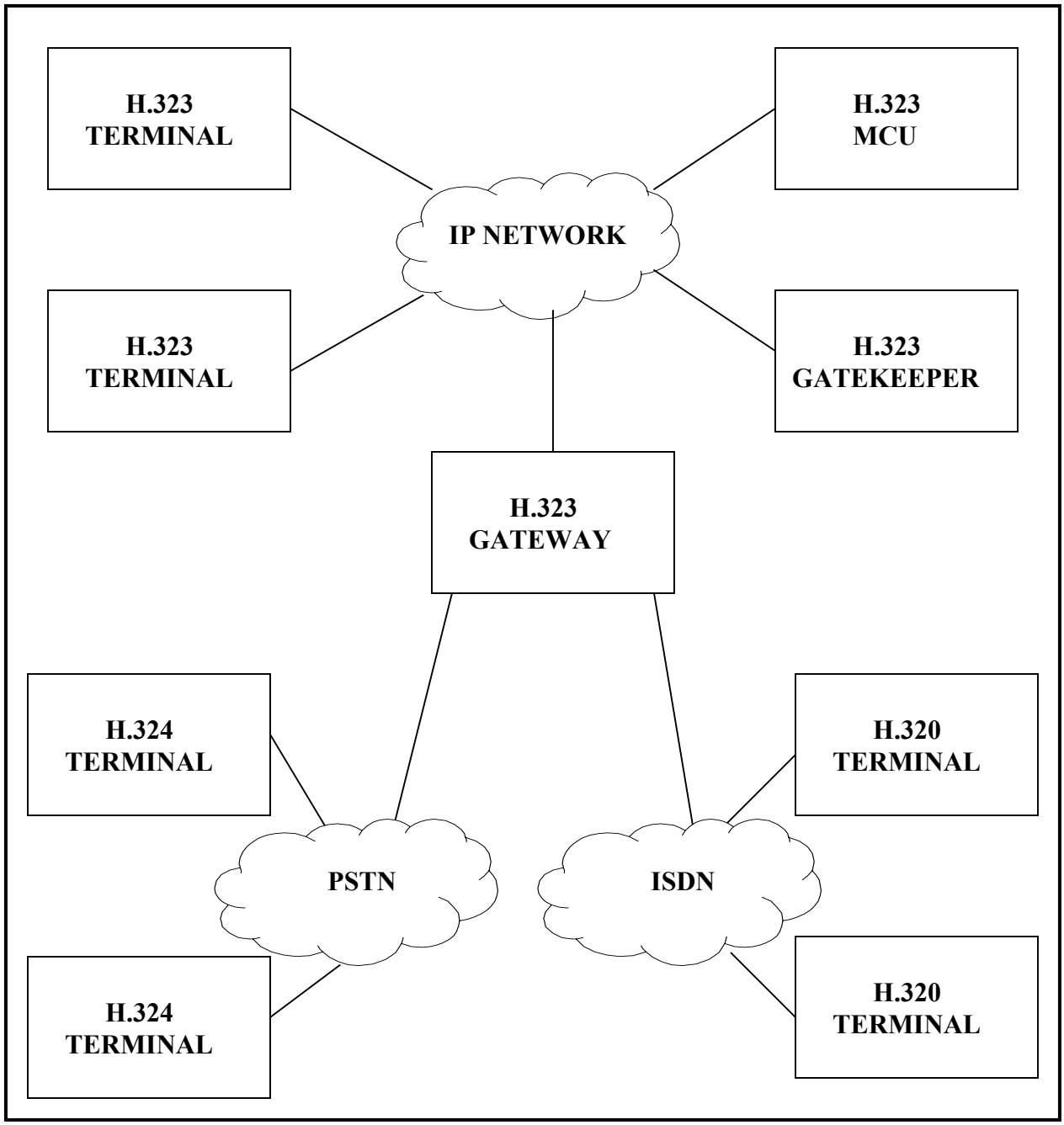

Figure 2. Components and environment of H.323

a. Terminal (TM)

An H.323 *Terminal* is an endpoint device which provides an interface for users to make real-time, two-way communications with another H.323 terminal, gateway, or multipoint control unit. While a terminal has to support audio communications, video and data are optional. A personal computer or internet telephone with the required software can be a terminal.

b. Gateway (GW)

An H.323 *Gateway* is an endpoint which provides protocol conversion between an H.323 network and a non-H.323 network. In other words, a gateway is the bridge between a packet-switched network and other types of networks on which other H.32X family standards run such as PSTN, ISDN, and B-ISDN. The voice packets coming to gateway are buffered and sent to the non-H.323 network at a constant rate. On the other hand, the voice transmissions coming from the non-H.323 network are compressed and turned into IP packets to be sent over the packet-switched network. A gateway is not necessary for two H.323 terminals to communicate on an H.323 network.

c. Multipoint Control Unit (MCU)

A *Multipoint Control Unit* is an endpoint device which makes it possible for three or more terminals or gateways to make a multipoint conference. MCU consists of two fundamental system components:

- Multipoint Controller (MC)
- Multipoint Processor (MP)

MC sends capabilities such as codec type to each endpoint participating in a conference. MP sends and receives media streams from endpoints in a conference.

d. Gatekeeper (GK)

A *Gatekeeper* provides call control, bandwidth management, and address translation for the connections between H.323 endpoints. It is an optional part which is used when there is a need for traffic analysis and billing. A gatekeeper is a logical part and may coexist with a terminal, gateway, or multipoint control unit.

2. H.323 Protocols and Specifications

H.323 specification includes several other H-Series recommendations and protocols for multimedia communication processing and controlling. H.245 Call Control, H.225 Call Signaling and H.225 Registration, Admission and Status (RAS) Channels will be explained further in this chapter; however brief descriptions of each are given below to better understand the H.323 protocol stack which is presented in Table 5:

a. H.225 Registration, Admission and Status (RAS) Channel

RAS channel is established between endpoints and a gatekeeper to provide pre-call control. RAS is used for registration, admissions, bandwidth change, status, and disengage procedures between endpoints and gatekeepers. RAS channel will be explained later in this chapter.

b. H.225 Call Signaling

H.225 Call Signaling Channel provides call control procedures for H.323 networks. Call setup, teardown, and supplementary services are all handled through the H.225 call signaling channel.

c. H.245 Call Control

H.245 Call Control Channel provides interoperability and manages the operation of all endpoints. H.245 messages include information about capabilities exchange, channel opening and closing, flow control and general commands. An endpoint establishes one H.245 call control channel for each call.

d. Audio Codec

An *audio codec* (Audio Coder and decoder) digitizes the analog voice and vice versa. All H.323 terminals must support at least one default audio codec as specified in the ITU-T G.711 recommendation. Additional codecs may be supported. Table 4 presents the most commonly used codecs and their bit rates. High-bit-rate codecs introduce much less processing delay than low-bit-rate codecs as a result of their simple algorithms which do not require much computational time. However, these high-bit-rate codecs consume much bandwidth. In other words, a high-bit-rate codec spends more bits than a low-bit-rate codec to represent the same amount of voice information at the expense of more computational time. Chapter III explores more about the codecs.

Codec	Bit Rate (kbps)
G.711	64
G.726	32
G.728	16
G.729A/B	8
G.723.1	6.3
G723.1	5.3

Table 4. Audio codecs

e. Video Codec

A video codec encodes video fed from the transmitting H.323 terminal and decodes it at the receiving H.323 terminal side. Because H.323 specifies that video support is optional, a video codec may or may not be present in an H.323 terminal. However, a terminal must support the video codec standard specified in the ITU-T H.261 recommendation if the terminal provides video communications.

Table 5 summarizes the protocols and ITU Recommendations used by H.323 which uses IP/UDP encapsulation for audio and video packets and IP/TCP for signaling and control packets.

<table>
<tr><td colspan="3">System Control User Interface</td><td>Audio</td><td>Video</td></tr>
<tr><td>H.245
Call
Control</td><td>H.225
Call
Signaling</td><td>H.225
RAS
Channel</td><td>G.711
G.722
G.723.1
G.728
G.729</td><td>H.261
H.262
H.263</td></tr>
<tr><td colspan="3" rowspan="2"></td><td colspan="2">RTP/RTCP</td></tr>
<tr><td colspan="2">TCP</td><td colspan="2">UDP</td></tr>
<tr><td colspan="5">IP</td></tr>
</table>

Table 5. The H.323 protocol stack

3. H.225 RAS Channel

A RAS channel is established between endpoints and a gatekeeper to provide pre-call control. H.225 RAS channel uses the unreliable but more efficient IP/UDP encapsulation. In networks with a gatekeeper, RAS channel is the first step to perform registration, admission, status, bandwidth changes, and disengaging procedures between endpoints and gatekeepers. However, a network may be without a gatekeeper, and then the endpoints skip the RAS channel messages and begin the next step; the exchange of H.225 call signaling messages which is the next topic after H.225 RAS channel.

An endpoint first finds a gatekeeper with *gatekeeper discovery* process, and then registers with this gatekeeper via *registration* process to inform the gatekeeper about its transport and alias addresses. Additionally, Endpoints and gatekeepers, which have an alias address for an endpoint, use *Location* messages to determine contact information of the endpoint. Also, gatekeepers may use *status information* messages to learn whether an endpoint is up and running or down. An endpoint uses *admission* messages to ask for permission from the gatekeeper to initiate or receive a call and *bandwidth* messages to

update the changing bandwidth requirements during a call. Lastly, *disengage* messages are used to inform gatekeepers about the end of a call. All of these types of RAS channel messages summarized in this paragraph are explained in the following paragraphs beginning with *gatekeeper discovery* below.

a. Gatekeeper Discovery

When an endpoint logs on a network which has a gatekeeper, it registers with that gatekeeper using the logical RAS channel. If there is more than one gatekeeper on the network, then each endpoint manually or automatically determines which gatekeeper to register with, this process is called *Gatekeeper Discovery*. In the manual process, endpoints register directly with their predefined and statically configured gatekeeper. This method has administrative overhead, since configuration file at the endpoint has to be changed to change to another gatekeeper. In the case of automatic process, an endpoint finds a gatekeeper over the network through an IP multicast message. The following messages below are used for this automatic registration process:

- GRQ (Gatekeeper Request)
- GCF (Gatekeeper Confirmation)
- GRJ (Gatekeeper Reject)

An endpoint sends a multicast GRQ message to find a gatekeeper. The well-known UDP port number for GRQ messages is 1718. Gatekeepers which can provide the requested function responds with a GCF message containing the transport address of the gatekeeper's RAS channel. The *transport address* consists of the network address and the TSAP identifier. The transport address is defined by the network protocol in use, for example, the transport address for a TCP/UDP/IP network is composed of IP address and TCP or UDP port number. Those gatekeepers which reject the requested function respond with a GRJ message. In the case of more than one GCF response back to the endpoint, endpoint chooses one of the gatekeepers willing to serve. This automatic process is known as *Auto Discovery* as presented in Figure 3.

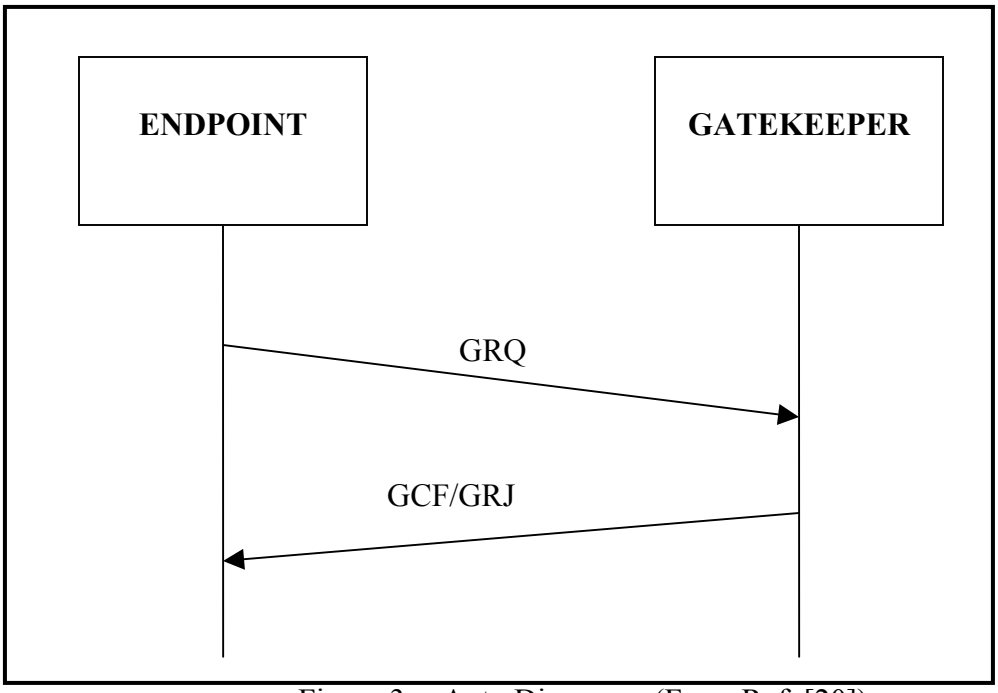

Figure 3. Auto Discovery (From Ref. [20])

b. Registration

After the gatekeeper discovery process, the endpoints in a network must register with their gatekeepers prior to any call attempt. Endpoints inform their gatekeepers of their transport addresses and alias addresses. *Alias addresses* provide an alternate way of addressing the endpoints. These addresses may include telephone numbers, alphanumeric strings, and e-mail-address-like addresses. The following messages are used for registration process:

- RRQ (Registration Request)
- RCF (Registration Confirmation)
- RRJ (Registration Reject)
- URQ (Unregister Request)
- UCF (Unregister Confirm)
- URJ (Unregister Reject)

Figure 4 presents the registration process between an endpoint and a gatekeeper. First, the endpoint sends a RRQ message to the transport address of

gatekeeper's RAS channel, then gatekeeper responds with either RCF or RRJ. Later, the cancellation of registration can be initiated by either the endpoint or the gatekeeper. The initiating part sends a URQ message which is responded with an UCF or URJ message.

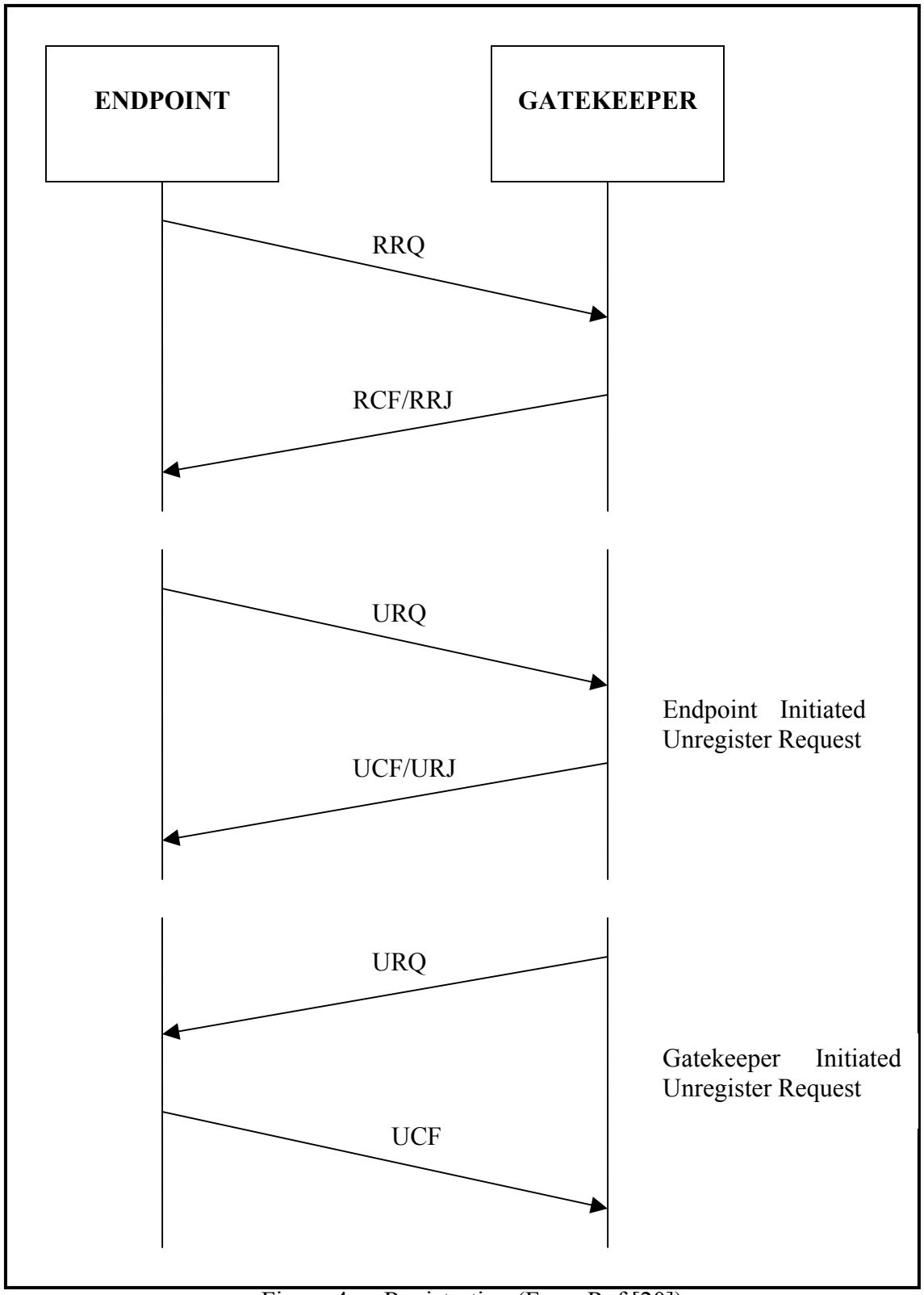

Figure 4. Registration (From Ref.[20])

c. Location

Endpoints and gatekeepers, which have an alias address for an endpoint, use *Location* messages to determine contact information of the endpoint. The contact information includes the Call Signaling Channel and RAS Channel addresses. Endpoints send location requests to their local gatekeepers and gatekeepers send location messages to other gatekeepers. The following messages are used for this process:

- Location Request (LRQ)
- Location Confirmation (LCF)
- Location Reject (LRJ)

An endpoint or a gatekeeper sends a LRQ message to request the contact information of a gatekeeper or an endpoint. A gatekeeper sends a LCF message which contains the call signaling channel and RAS channel addresses of the gatekeeper or the requested endpoint. When the gatekeeper routes the call, it provides its own contact information. When endpoints directly call each other, the gatekeeper provides the requested endpoint's contact information. A LRJ message is sent by a gatekeeper to indicate that the requested endpoint is not registered or the gatekeeper does not have available sources. Figure 5 presents endpoint location process.

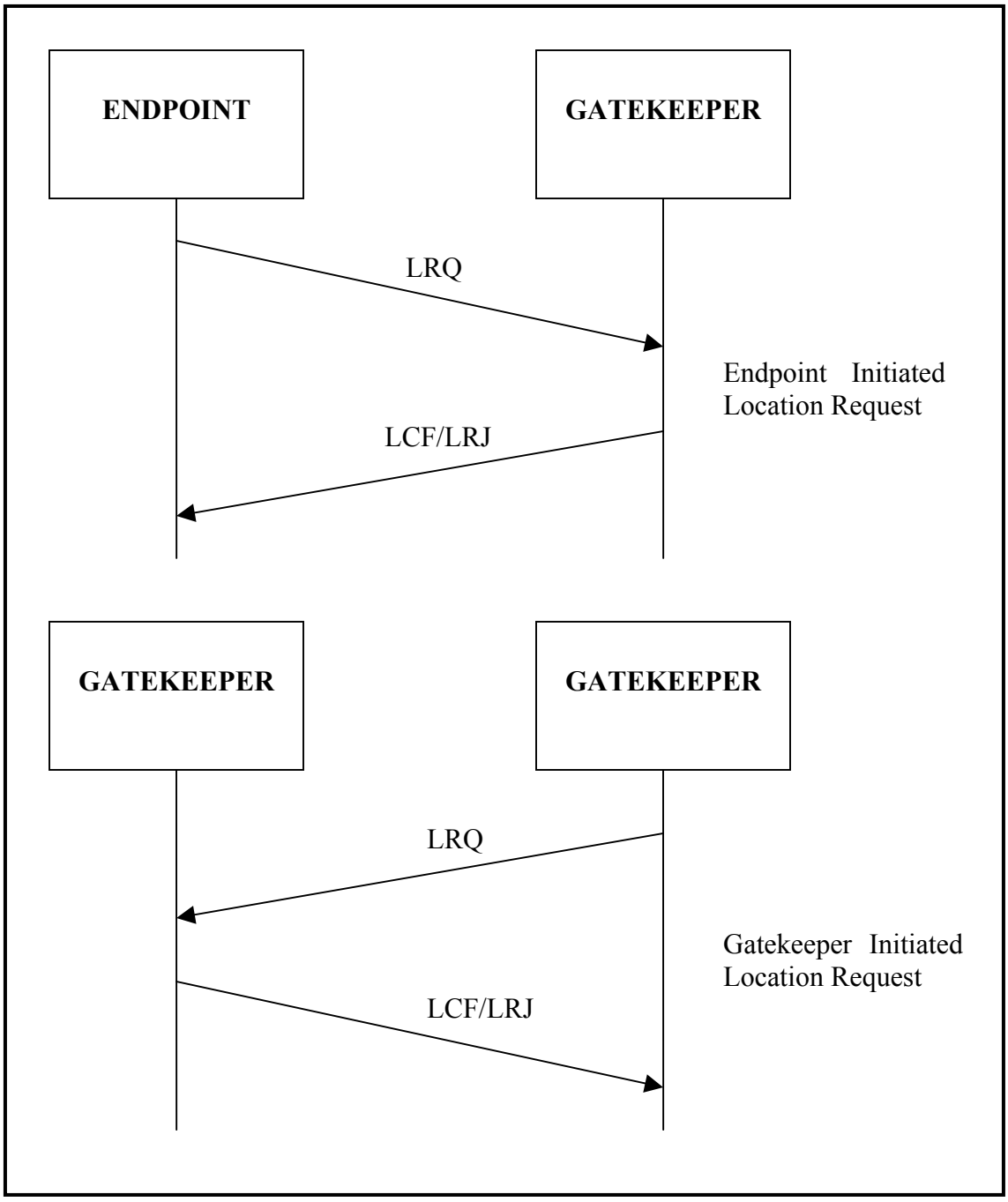

Figure 5. Endpoint Location Process

d. Admissions and Bandwidth

Endpoints use admission messages to ask for permission from the gatekeeper to initiate or receive a call. The following messages are used for admission:

- ARQ (Admission Request)

- ACF (Admission Confirmation)
- ARJ (Admission Rejected)

An endpoint sends an ARQ message with the requested bandwidth to the gatekeeper to initiate or receive a call. The gatekeeper in turn may give permission with an ACF message including the allowed bandwidth (which may be less than originally requested) and the IP address of the terminating gatekeeper or gateway or it may reject the admission request with an ARJ message because the endpoint is not registered with it or for other policy reasons. An ARJ message includes reasons for rejection and optionally other gatekeepers to query.

The bandwidth control is first done through the admission process as stated in the previous paragraph. However, the bandwidth requirements may change during the call. The following messages are used for bandwidth control:

- BRQ (Bandwidth Request)
- BCF (Bandwidth Confirmation)
- BRJ (Bandwidth Rejected)

A BRQ message is sent when there is a change in the need for the initially allowed bandwidth in the ACF message. The change may occur for instance as a result of changing from voice communications to video communications. Upon receiving the BRQ message, the gatekeeper may either confirm with a BCF message or reject with a BRJ message.

Figure 6 presents both admission and bandwidth messages between endpoints and gatekeepers.

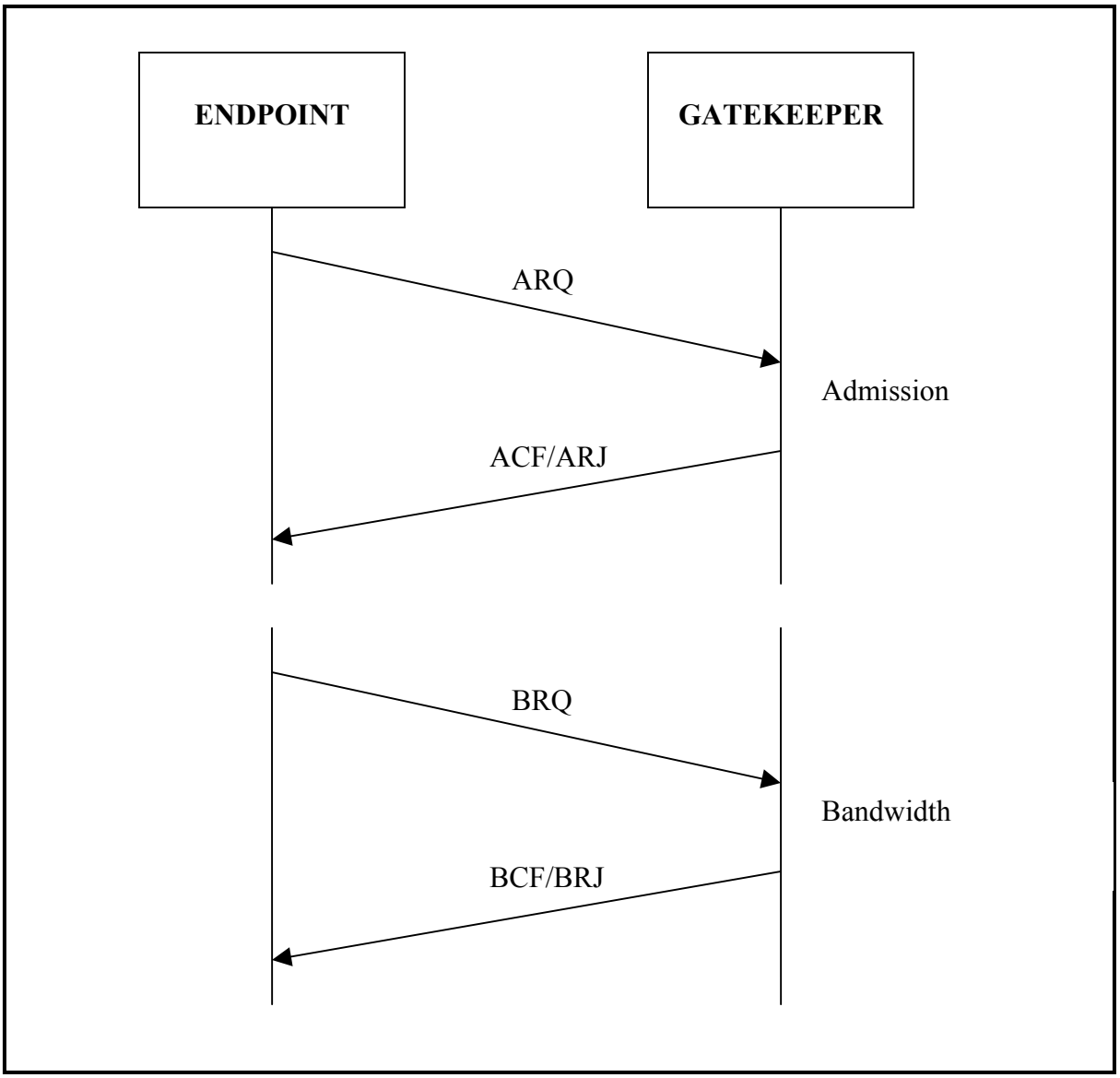

Figure 6. Admission and Bandwidth Processes

e. Status Information

Gatekeepers use the RAS *status information* messages to learn whether an endpoint is up and running or down. The following messages are used for status:

- IRQ (Information Request)
- IRR (Information Request Response)
- IACK (Information Acknowledged)
- INAK (Information not Acknowledged)

There may be two different scenarios with status information messages. In the first scenario, a gatekeeper sends an IRQ message to an endpoint, and the endpoint responds with an IRR message in response. Secondly, during the ACF, the gatekeeper requests the endpoint to periodically send IRR messages during the call. The gatekeeper acknowledges those periodic messages with either IACK or INAK. The exchange of status messages is shown in Figure 7.

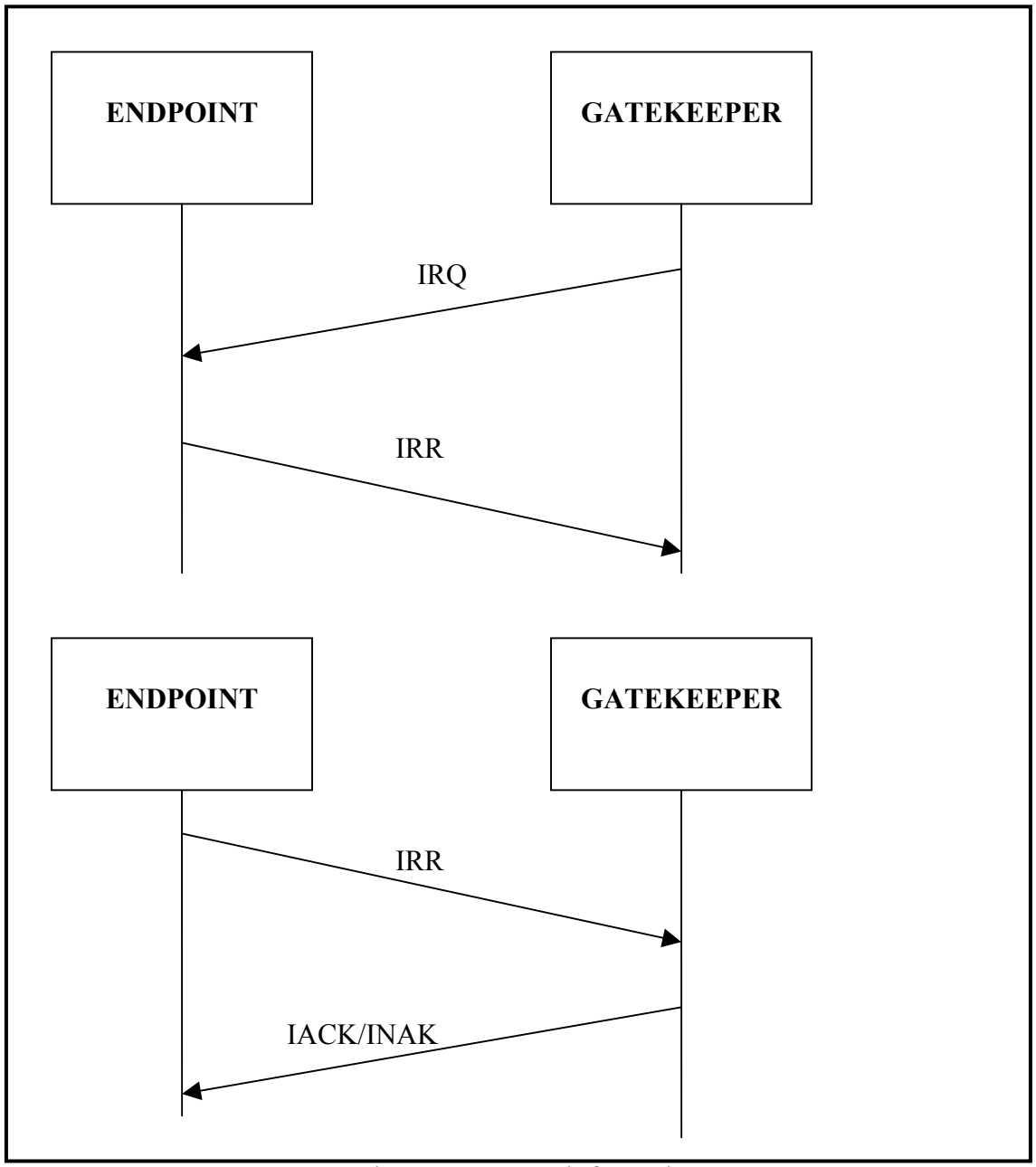

Figure 7. Status information

f. Disengage

The call disconnection information is crucial when the H.225 call control channels do not route through the gatekeeper. Endpoints must inform their gatekeepers with a disconnect message. With the help of this information, a gatekeeper knows that a previously admitted call is over. Further, the gatekeeper updates its record of available bandwidth in the network. The following messages are used for the disengage process:

- DRQ (Disengage Request)
- DCF (Disengage Confirmation)
- DRJ (Disengage Rejected)

A DRQ message is sent to the gatekeeper to indicate a disconnection, the gatekeeper acknowledges the DRQ message with a DCF message or rejects it with a DRJ message. Rejection may happen due to a DRQ message from an unregistered endpoint.

4. H.225 Call Signaling

The ITU Recommendation H.225 provides call control procedures for H.323 networks. Call setup, teardown, and supplementary services are all handled through the H.225 call signaling channel. H.225 creates a reliable call control channel on TCP port 1720.

H.225 uses and supports ITU-T Recommendations Q.931 and Q.932. Q.931 is a signaling protocol to establish and terminate calls. This protocol provides traditional telephone functionality to the H.323 calls such as dial tone and ringing. Q.932 provides for supplementary services. Table 6 presents the Q.931 and Q.932 messages used for call signaling.

Q.931/Q.932 Messages	Description
SETUP	Used for initiating a call
CALL PROCEEDING	Indicates that call establishment procedures began
ALERTING	Indicates that the called party was signaled (ringing)
CONNECT	Indicates that the called party accepted the call
RELEASE COMPLETE	Indicates that the call is being released
FACILITY	Indicates whether a call should be direct between endpoints or routed through a gatekeeper
STATUS	Used for RAS status information messages

Table 6. Q.931 and Q.932 Messages

After the H.225 RAS channel messages are exchanged, the endpoints get registered with their gatekeepers. In order to make a call, the next step is to exchange H.225 call signaling messages. Call signaling messages may be passed in two different ways. The first method is *direct endpoint call signaling* which takes place directly between endpoints. The second method is *gatekeeper routed call signaling* which takes place between endpoints through routers. Additionally, a network may not have a gatekeeper; in this case, the exchange of H.225 call signaling messages becomes the first step to establish a call as presented in Figure 8 since RAS channel messages are exchanged with gatekeepers.

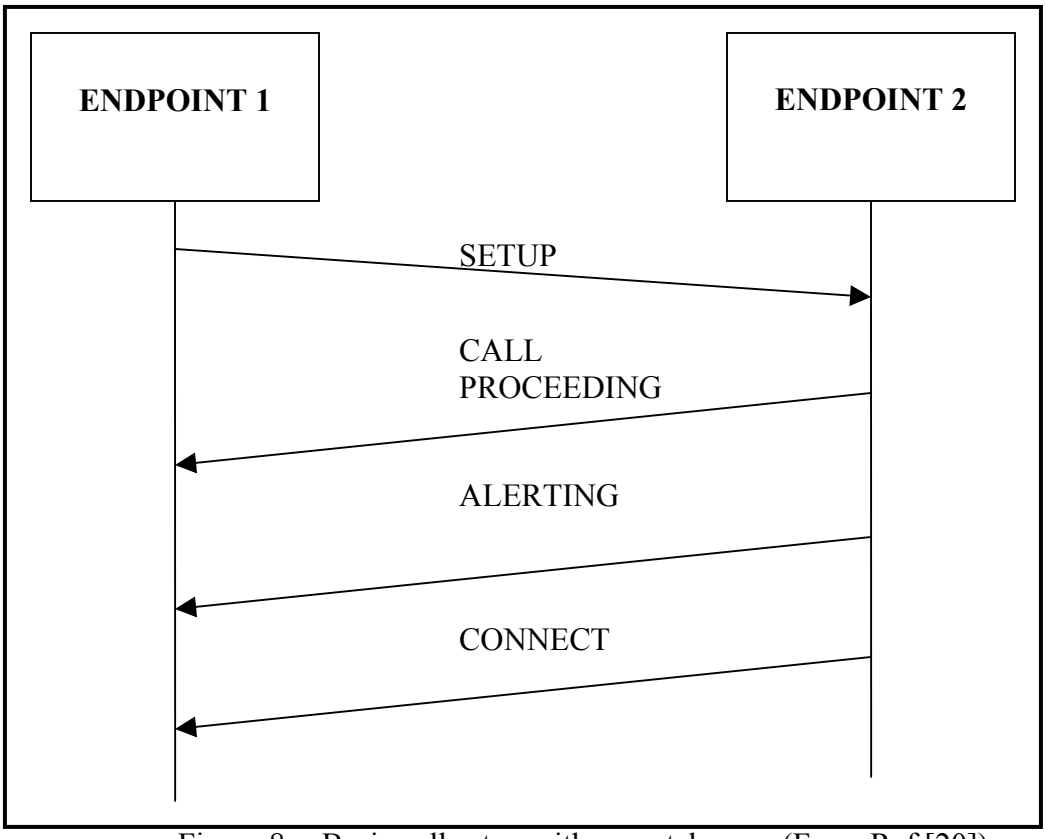

Figure 8. Basic call setup with no gatekeeper (From Ref.[20])

In Figure 8, endpoint 1 sends SETUP message to establish a call with endpoint 2. Upon receiving the SETUP message, Endpoint 2 sends a CALL PROCEEDING message to indicate that call establishment procedures began, and ALERTING to indicate that it is being actively signaled finally CONNECT to indicate that it accepted the call.

For networks with a gatekeeper, the endpoints must first request admission from the gatekeeper via ARQ message and the permission is granted via ACF message as explained earlier in H.225 RAS channel signaling part. The ACF message, which gives permission to an endpoint to initiate a call, also indicates whether the call should be direct or routed through a gatekeeper. After this step, an endpoint may initiate a call via H.225 SETUP message.

If the ACF message, sent by the gatekeeper, requires that the call be direct between endpoints, then call signaling messages are directly routed between endpoints as presented in Figure 9 . The endpoints interact with gatekeeper only through RAS channel

to ask for admission to the network, and then the H.225 call control signaling messages are exchanged between the endpoints without gatekeeper intervention.

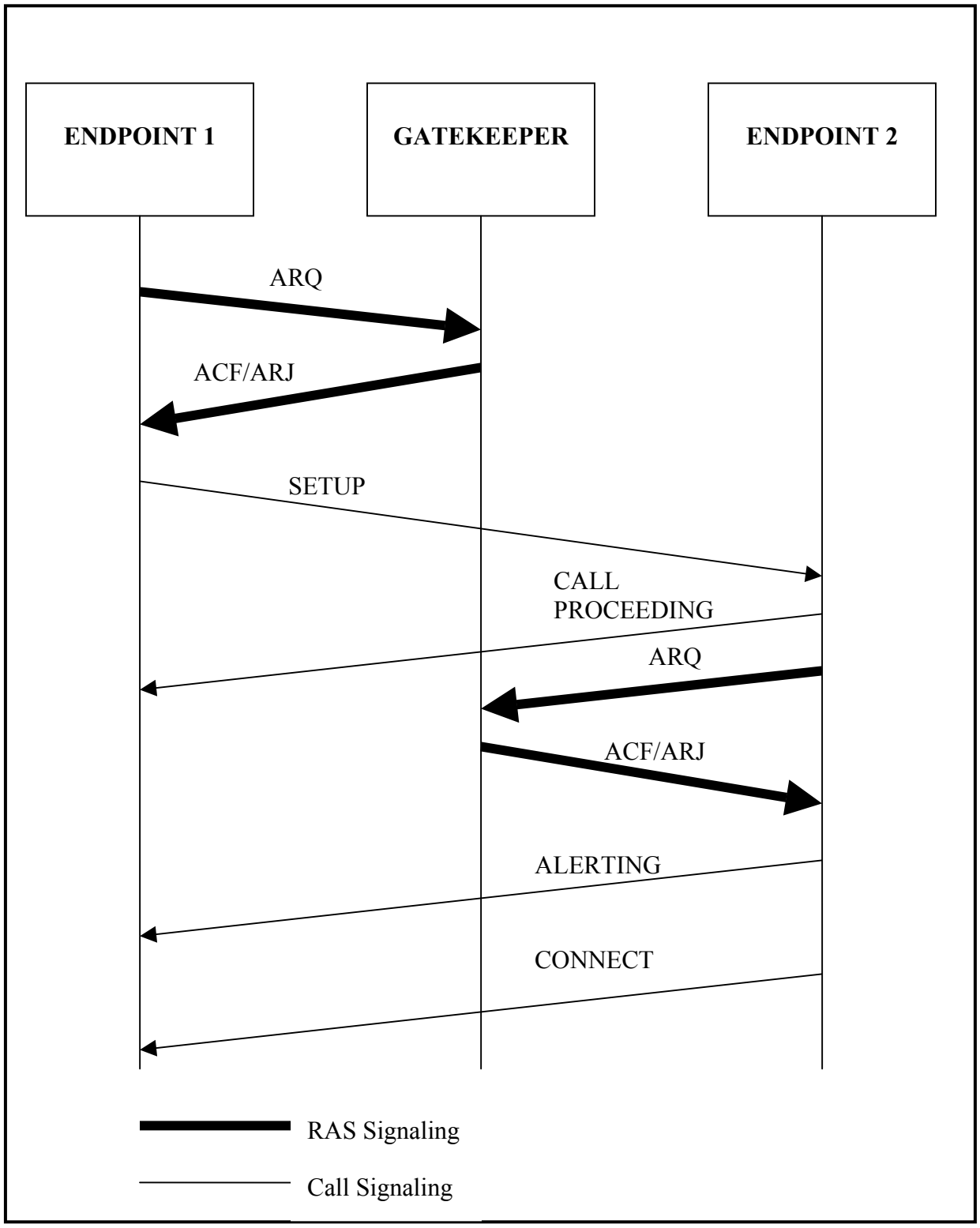

Figure 9. Direct Endpoint Call Signaling (From Ref.[20])

If the ACF message, sent by the gatekeeper, requires that the call be routed through itself, then H.225 call signaling messages are routed through the gatekeeper as presented in Figure 10. This kind of call is known as *Gatekeeper Routed Call Signaling*. The RAS messages again flow between the endpoints and the gatekeeper

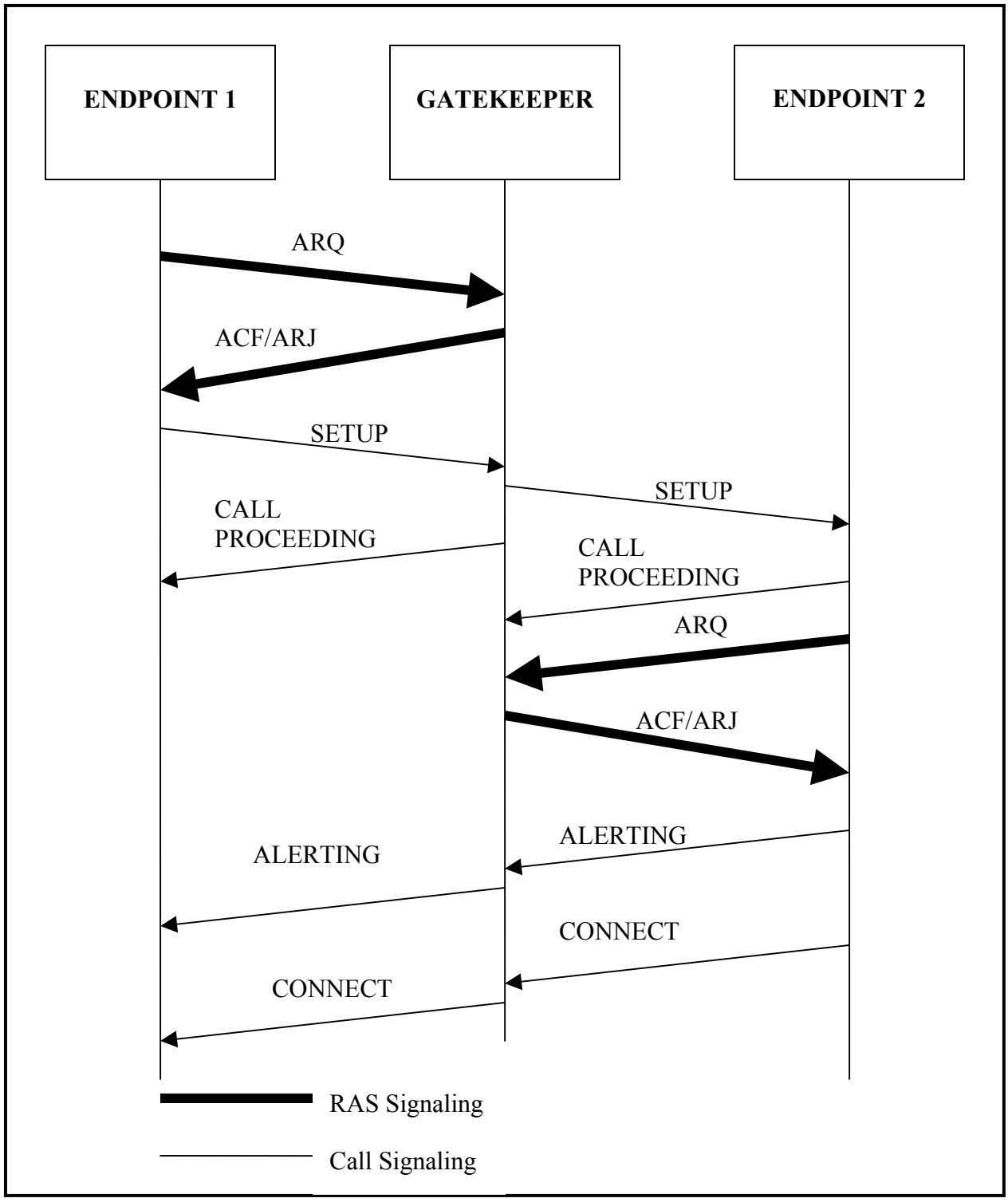

Figure 10. Gatekeeper Routed Call Signaling (From Ref.[20])

Lastly, Figure 11 presents a scenario where both endpoints are registered with different gatekeepers. Both gatekeepers route the call control signaling messages. First, endpoint 1 exchanges RAS messages with its gatekeeper to gain admission to the network. Then, endpoint 1 sends SETUP message to its gatekeeper to establish a call with endpoint 2. Gatekeeper 1 then sends it to Gatekeeper 2 which has the called endpoint 2 as a registered endpoint. Then, Gatekeeper 2 sends the SETUP message originated from endpoint 1 to endpoint 2. Upon receiving the SETUP message, endpoint 2 sends a CALL PROCEEDING message to indicate that call establishment procedures began and begins exchanging RAS messages with Gatekeeper 2 to gain admission to the network. After receiving the admission, endpoint 2 sends ALERTING to indicate that it is being actively signaled and CONNECT to indicate that it accepted the call.

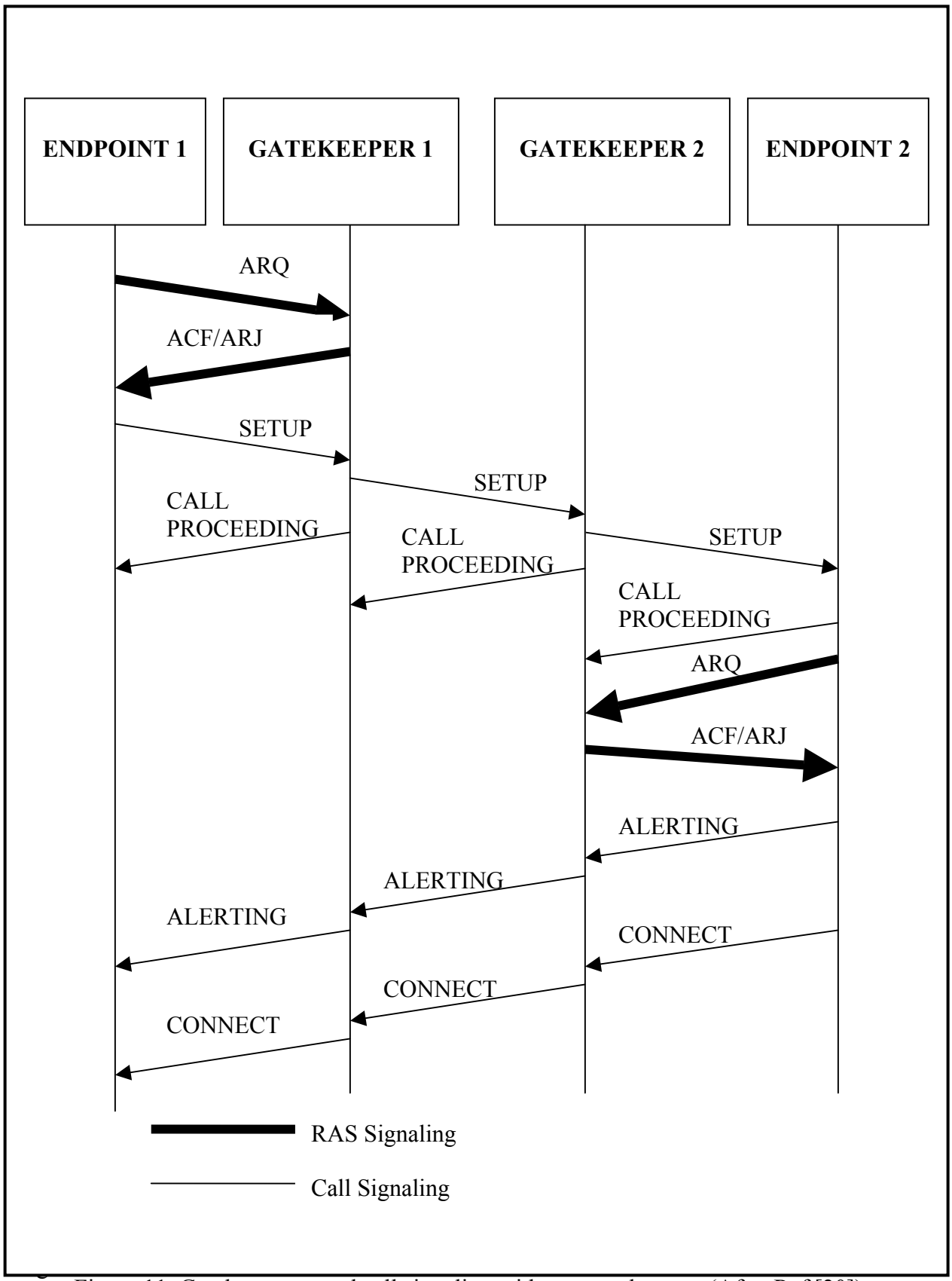

Figure 11. Gatekeeper routed call signaling with two gatekeepers (After Ref.[20])

5. H.245 Call Control

After the connection has been established as described above via H.225 call signaling messages, it is time to use H.245 Call Control Channel which provides interoperability and manages the operation of all endpoints. This is the last step to start the exchange of actual voice packets. H.245 messages include information about capabilities exchange, channel opening and closing, flow control and general commands. An endpoint establishes one H.245 call control channel for each call. Figure 11 presents a typical H.245 messages flow between two endpoints.

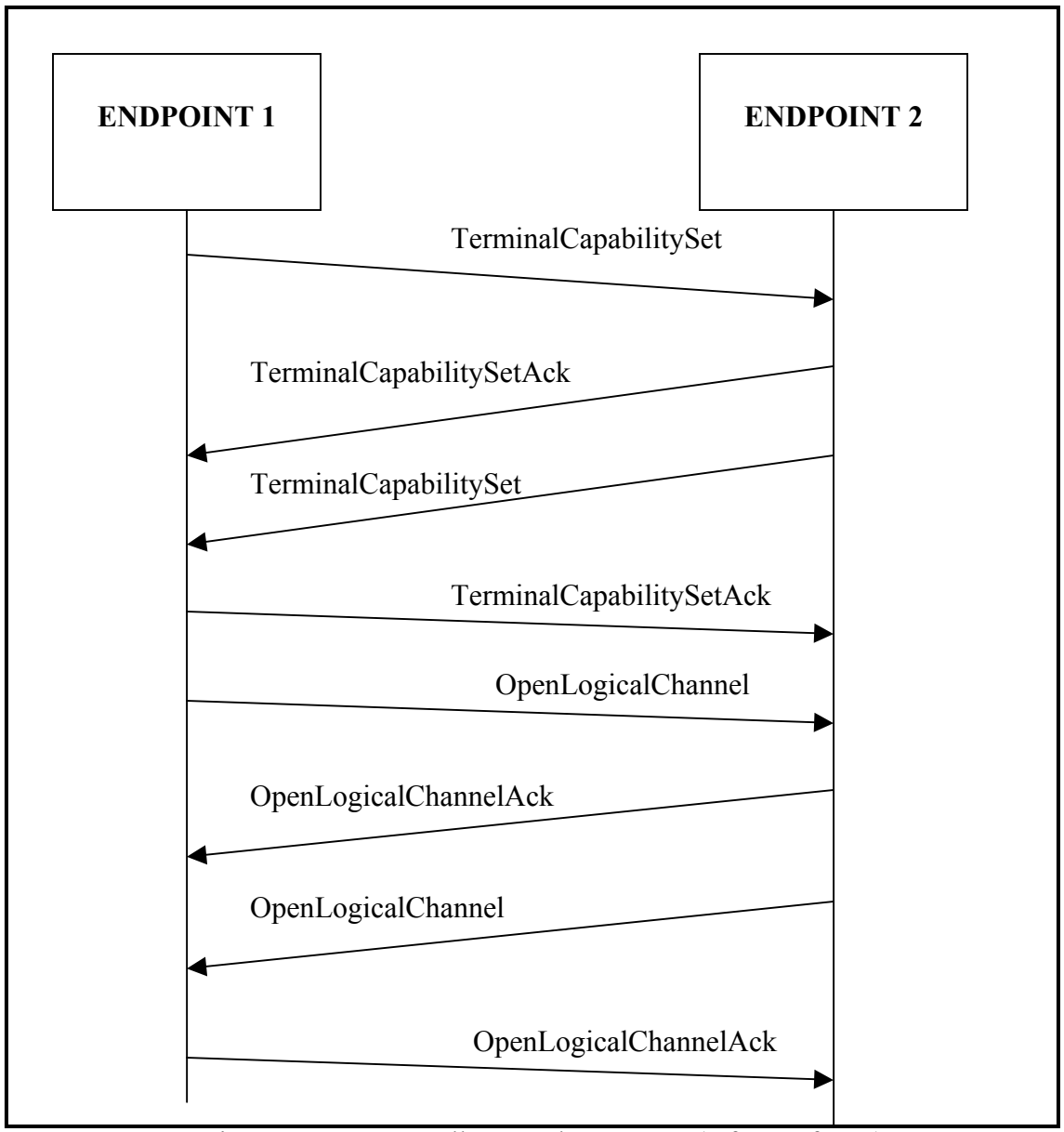

Figure 12. H.245 Call Control Messages (After Ref.[20])

Assuming that the connection between two endpoints has been established with H.225 messages before, endpoint 1 sends an H.245 TerminalCapabilitySet message to endpoint 2 to exchange capabilities. Endpoint 2 acknowledges endpoint 1's capabilities with a TerminalCapabilitySetAck message. Endpoint 2 also sends its capabilities to be acknowledged by endpoint 1. Then, endpoint 1 opens a media channel with endpoint 2 by sending an OpenLogicalChannel message to endpoint 2 which in turn acknowledges the message with an OpenLogicalChannelAck message including the Real-Time Transport Protocol (RTP) transport address to be used for sending multimedia packets. Then endpoint 2 opens a media channel with endpoint 1 by sending an OpenLogicalChannel message to be acknowledged by endpoint 1.

6. RTP and RTCP

After the exchange of H.245 messages, the endpoints are ready for the exchange of actual voice packets. *Real-Time Transport Protocol* (RTP) provides for this capability by running over UDP/IP.

RTP, specified in RFC 1889, provides a foundation for transmitting data with real time characteristics such as voice and video over packet networks. As it was previously mentioned, real-time packets use the connectionless, best-effort UDP on top of IP due to the delay-sensitive nature of voice traffic; however, UDP does not provide sequencing and time stamping services needed to reassemble a traffic stream at the receiving end. RTP fills this gap with a 12-byte additional header overhead to a typical voice packet (see Table 2). RTP typically runs over UDP to make use of its multiplexing and checksum services.

The *Real-Time Transport Control Protocol* (RTCP), also specified in RFC 1889, is a counterpart of RTP which provides for monitoring network and application performance. For example, RTCP informs the sender about the number of packets detected as lost. RTCP packets are sent periodically by a receiver to a sender and consume very little bandwidth. The maximum frequency suggested by RFC 1889 is one RTCP packet every five seconds. Finally, RTP does not have to run with RTCP. RTCP may be used when there is a need for monitoring the network performance and problem areas in the network and applications. Figure 12 presents the flow of RTP and RTCP media streams between two endpoints.

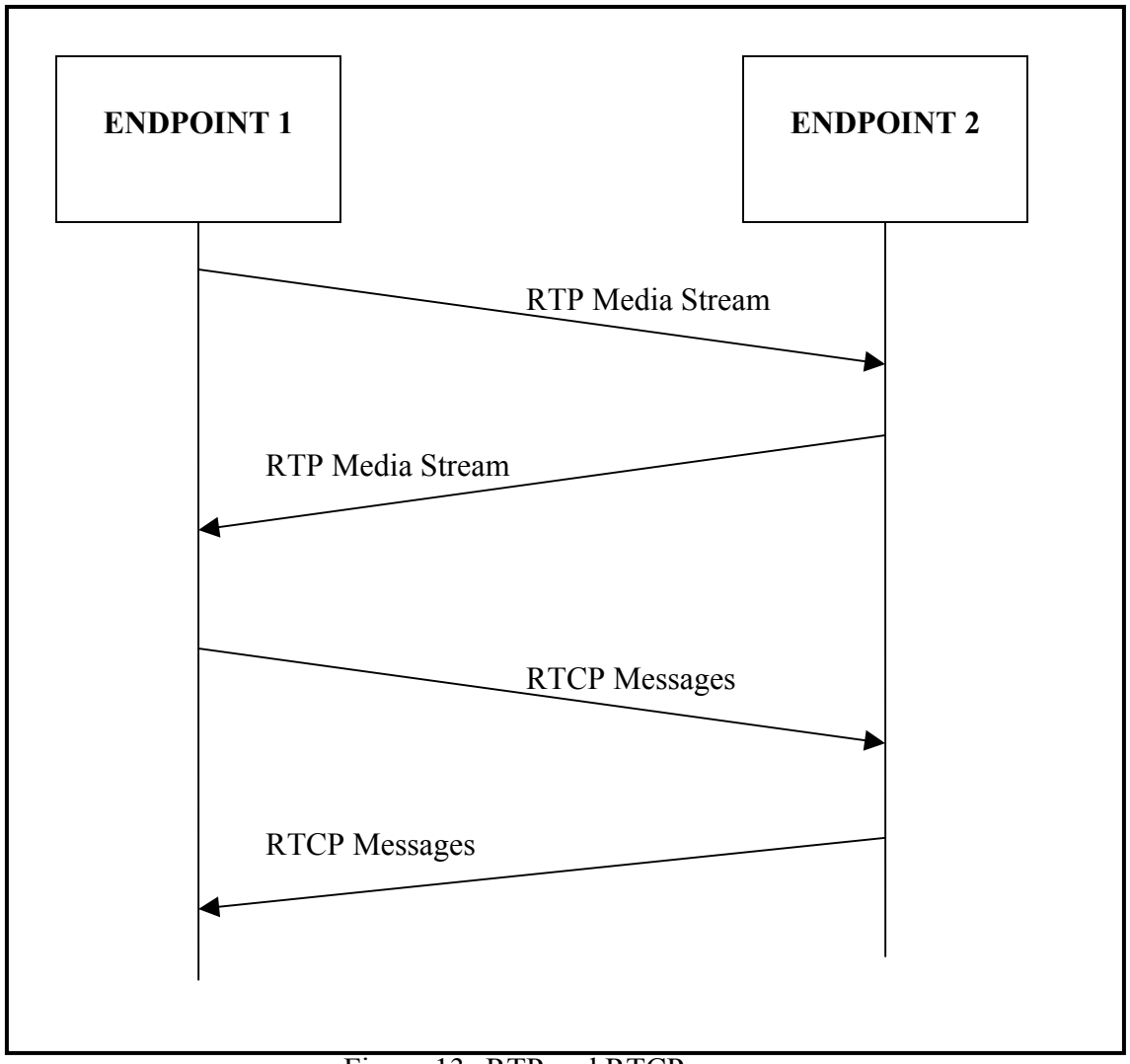

Figure 13. RTP and RTCP messages

This section covered H.323, the most widely used VoIP protocol in detail. In particular, the main components of an H.323 network and underlying protocols within H.323 were explained. The next section presents a brief overview of Session Initiation Protocol (SIP) which is a very lightweight and less-detailed protocol compared to H.323.

D. SIP

SIP, developed by Internet Engineering Task Force (IETF) and described in RFC 2543, is a text-based protocol similar to HTTP (Hyper Text Transfer Protocol). It supports multimedia communications such as voice and video conferencing. It is different from H.323 in that it is less complex and more flexible. SIP does not send many signaling and control messages over the network which is one of the most criticized drawbacks of

H.323. SIP, as a result of its simplicity and practicality, has received much attention and support from major companies like Microsoft, Cisco, Nortel and Lucent during the last several years. For example, Microsoft has deployed SIP in Microsoft XP, Pocket PC and MSN Messenger. However, we currently live in an H.323 world.

SIP is part of IETF's multimedia protocols. These protocols are Session Description Protocol (SDP), Session Advertising Protocol (SAP), Real-Time Protocol (RTP) Real-Time Control Protocol (RTCP) and Real-Time Streaming Protocol (RTSP). SDP, specified in RFC 2327, provides session and flow control for multimedia sessions. SAP is used for advertising multicast conferences. RTP and RTCP, as explained before, provide real-time delivery of data. Real-Time Stream Protocol (RTSP), specified in RFC 2326, provides on-demand delivery of real-time data. The details of these protocols are beyond the scope of this thesis.

SIP uses the following messages called as *methods* for signaling. Sample scenarios to see the exchange of these messages will be presented further in this chapter.

Method	**Description**
REGISTER	Registers the user with a SIP server.
INVITE	Initiates a call.
ACK	Acknowledges acceptance of a call.
BYE	Ends a call.
CANCEL	Ends a call not yet connected.
OPTIONS	Queries a server about its capabilities

Table 7. SIP Methods

A SIP system includes five major components: *User Agents, Proxy Servers*, *Redirect Servers*, *Registrar Servers,* and *Location Servers*. *User agents* are end-user devices just like the terminals in an H.323 network. A user agent can be a PC or an IP phone. A user agent is both a *User Agent Client* which initiates a SIP request and a *User*

Agent Server which responds to the SIP request. *Registrar servers* accept REGISTER messages from the user agents. *Proxy servers* accept SIP session requests from User agents and make requests on behalf of the clients. *Redirect servers* provide clients with the address information of the desired party in order for the clients to make a direct call. *Location servers* contain user information which provides redirect and proxy servers with the possible locations of the called party.

SIP servers operate in two modes: *proxy* and *redirect*. In the proxy call mode, proxy servers function as relay points between user agents and servers which make SIP requests on behalf of other user agents. Figure 14 presents a simplified SIP call in proxy mode. This is analogous to H.323 gatekeeper-routed call (see Figure 10). In the redirect call mode, redirect servers provide calling user agents with the network address (IP address) of the called parties and leave the job of establishing the call to the calling user agent as shown in Figure 15. This is analogous to H.323 direct endpoint calling (see Figure 9).

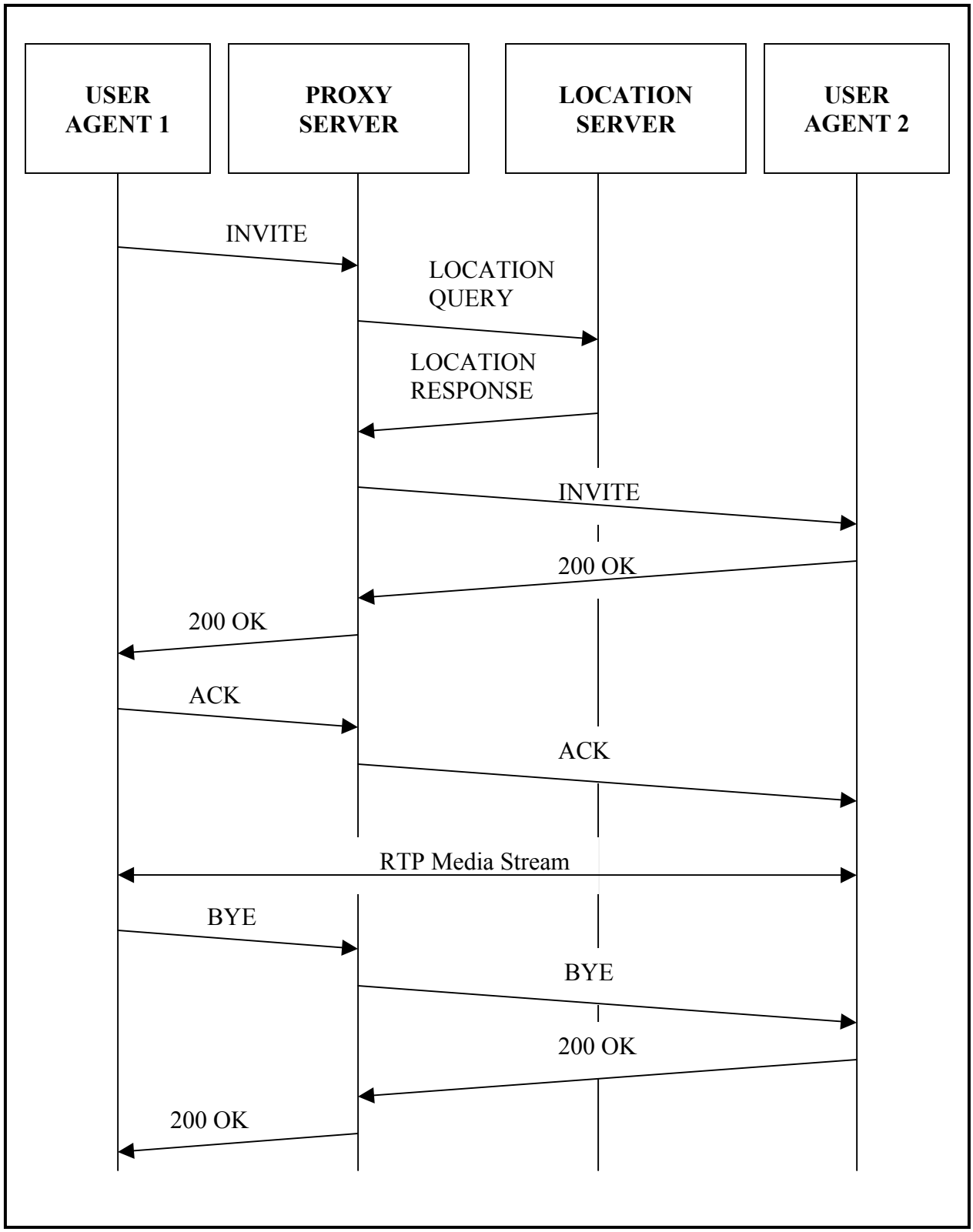

Figure 14. SIP Call Using Proxy Server (After Ref.[21])

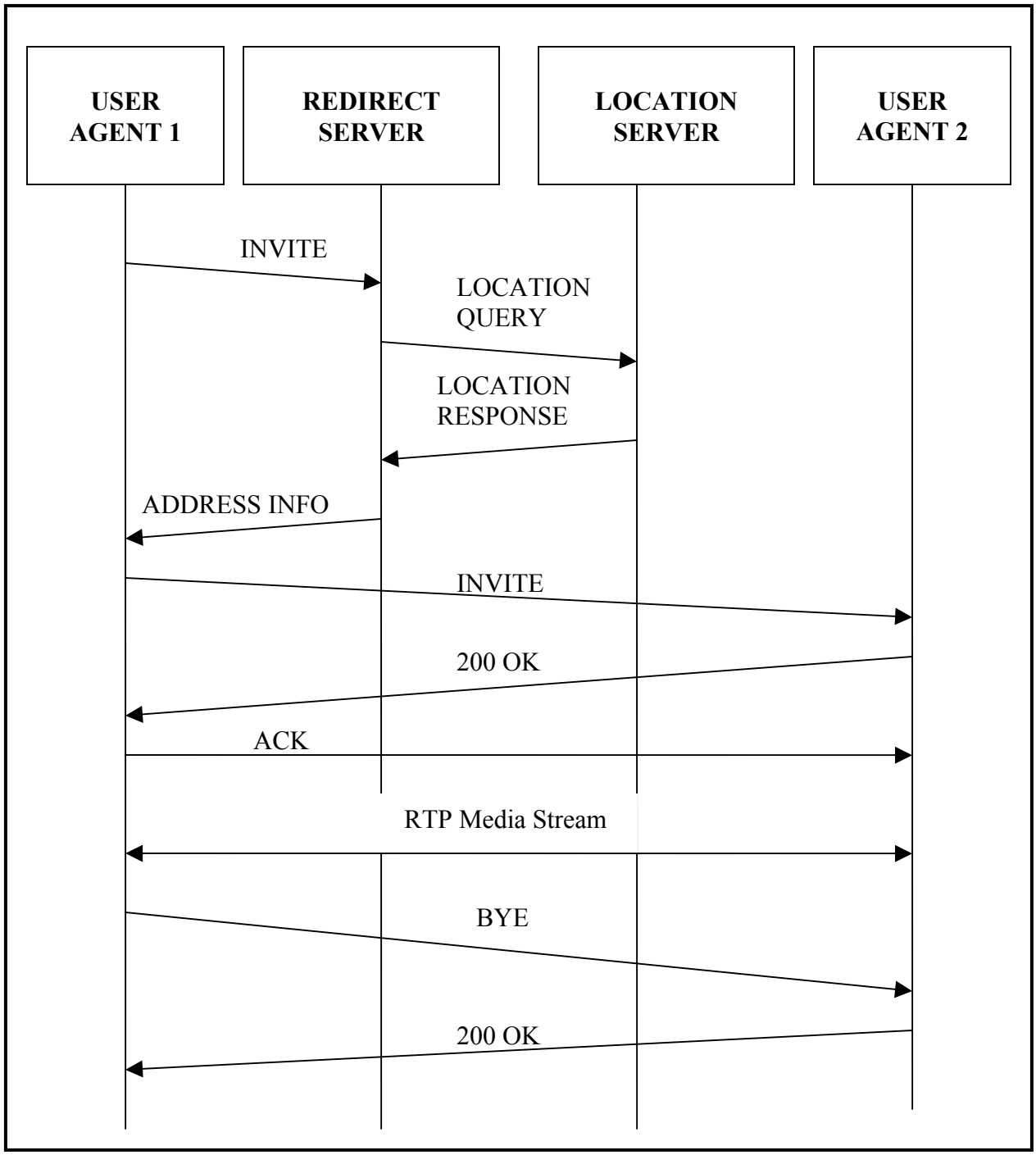

Figure 15. SIP Call Using Redirect Server (After Ref.[21])

THIS PAGE INTENTIONALLY LEFT BLANK

III. FACTORS AFFECTING VOICE QUALITY

Previous chapters presented an overview of VoIP and the factors affecting its growth. This chapter will explore the main factors affecting voice quality: *delay, jitter, packet loss, link errors, echo,* and *voice activity detection.* Having an understanding of these factors is important to realize why VoIP is not growing faster than it is today.

A. DELAY

Delay is the time interval between the instant that the talker speaks and the listener hears. *This is the most important factor in determining the voice quality for VoIP***.** One major problem that delay can cause is *Speech Overlap.* In a regular conversation when a talker finishes speaking, he or she waits for the listener to speak. If response does not arrive within a suitable time, then talker begins speaking again just before he or she gets the delayed response which collides with the talker's speech. Discerning the collision, both parties cease to speak for a while and begin speaking again only to stop when they hear others speaking and so on. The delay threshold for a conversation not to suffer speech overlap is 250 ms [Ref3]. However, a conversation with a delay of up to 400 ms is acceptable provided that both parties are aware of the impact of delay.

Another type of problem caused by delay is one-way delay of more than 25 ms, making echo annoying to the speaker. Therefore, echo cancellation is required for better voice quality when one-way delay is larger than 25 ms. [Ref. 2]. Table 8 specifies delay guidelines of ITU-T Recommendation G.114 for an adequately controlled echo. Delay can also cause packet loss, if delay variance is large. Packet loss due to delay variance (jitter) will be explained later in this chapter.

DELAY	COMMENTS
0-150 ms	Acceptable for most applications
150-400 ms	Acceptable, provided that the impact of delay is known.
Above 400 ms	Unacceptable

Table 8. ITU-T Recommendation G.114 for Delay Specification (After Ref. [2])

Overall delay consists of components such as propagation delay, codec processing, packetization, serialization, routing, queuing and jitter buffer delay.

1. Propagation Delay

Propagation delay is the time required for a signal to travel from one end to another as it passes through wired or wireless media. This delay is proportional to the speed of light which is 300,000,000 meters per second through a vacuum and approximately 2/3 of that speed through copper and fiber.

2. Codec Processing Delay

A voice *codec* (Coder and Decoder) digitizes the analog voice and vice versa. *Codec processing delay* occurs when a codec codes, compresses, decompresses and decodes the voice signals. *Coding* is the digitization of the analog audio signals. *Compressing*, which is used by some codecs, is the process of using an algorithm that reduces the need for space and the bandwidth required to store and transmit the data. *Decoding* and *Decompressing* are just the reverse processes of coding and compressing respectively. Generally codecs with low bit rates such as G723.1 and G.729 introduce more delay as a result of the complex algorithms they use for compression. Compressing voice signals reduces bandwidth requirements at the expense of higher computational time which means more delay and thus degraded voice quality. A tradeoff between bandwidth and voice quality has to be considered based on the available computational power, desired voice quality and network bandwidth requirements.

Table 9 presents characteristics of several voice codecs sorted by bit rate. High-bit-rate codecs introduce much less processing delay than low-bit-rate codecs as a result of their simple algorithms which do not require much computational time. However, these high-bit-rate codecs consume much bandwidth. In typical VoIP applications, the DSP (Digital Signal Processor) generates a frame every 10 ms. Two of these frames are then placed in one voice packet. Table 9 presents the codec processing delays for different codecs to handle a voice packet of 20 ms as explained above. Finally, the *lookahead time* is required by some codecs' algorithms to look ahead into the next voice frame while processing the current one. Chapter V will present what to consider when choosing a codec.

Codec	Bit Rate (kbps)	Codec Processing Delay (ms)	Lookahead Time (ms)
G.711	64	0.75	0
G.726	32	1	0
G.728	16	3 to 5	0
G.729A	8	10	5
G.723.1	6.3	30	7.5
G723.1	5.3	30	7.5

Table 9. Characteristics of Several Voice Codecs (After Ref. [5] and [7])

3. Packetization Delay

Packetization delay is the time it takes to fill the packet payload with the digital and probably compressed voice samples. This delay depends on the number of codec frames placed into one single packet. The more codec frames are placed into one packet, the more packetization delay since the first codec frame will have to wait for the other frames to be generated and put in the same packet. In short, packetization delay increases, as the packet gets larger. The motivation behind choosing large packets is the goal of using bandwidth efficiently. Using small packets increases the number of packets to be sent which means more overhead caused by the header information and more possible collisions. In this situation, there is a tradeoff between bandwidth utilization efficiency and packetization delay and thus voice quality. The header information of a voice packet makes up a total of 40 bytes for every packet as shown on Table 10.

RTP ➔12 Bytes	**UDP** ➔ 8 Bytes	**IP**➔ 20 Bytes

Table 10. Voice Packet Header Overhead

Table 11 illustrates overhead and bandwidth efficiency rates for two codec types: G.723.1 with 5.3 kb/s coding speed and G729A with 8 kb/s coding speed. Calculations are made both with one frame per packet and two frames per packet for two codecs. The payload size for G.723.1 and G729A are 20 bytes and 10 bytes respectively. As it is seen on Table 11, G.723.1 with two frames is the most bandwidth efficient one with 50 % efficiency. There are two reasons for VoIP to be bandwidth-inefficient.

- The voice payload has to be small in order to reduce packetization delay.
- VoIP has to use RTP on top of UDP since it is a real time standard. This necessity causes an extra 12 Bytes overhead. This problem can be lessened with Compressed Real Time Protocol (CRTP).

Codec	Frames per packet	IP packet size (bytes)	UDP packet size (bytes)	Payload size (bytes)	Overhead	Bandwidth efficiency
G.723.1 5.3 kb/s	1	60	40	20	66.7%	33.3%
	2	80	60	40	50.0%	50.0%
G.729A 8 kb/s	1	50	30	10	80.0%	20.0%
	2	60	40	20	66.7%	33.3%

Table 11. Bandwidth Efficiency (From Ref.[4])

4. Serialization Delay

Serialization delay is the time it takes to place bits on the link. This delay is proportional to the speed of the link. So, the higher the link speed, the less time it takes to place the bit. In other words, the higher the speed, the less serialization delay. Table 12 presents serialization delay values for a 64-byte packet under different link speeds.

Packet Size (byte)	**Link Speed**				
	64 kbps	**256 kbps**	**512 kbps**	**1 Mbps**	**10 Mbps**
64	8 ms	2 ms	1 ms	0.5 ms	0.05 ms

Table 12. Serialization Delay (After [Ref. 6])

5. Routing and Queuing Delay

Routing delay is the time it takes for a network element to forward a packet. That is, the time between the instant a packet arrives at and exits from a network element such as a router. Routing time depends on the architecture, configuration, performance, and load of network devices. Hardware-based architectures are faster than software-based ones. *Queuing delay* occurs when there are more packets than the network device can process at an instant, causing the packets to be placed in a queue. Giving priority to voice packets may lessen this delay; however, a voice packet has to wait its turn if there are other voice packets ahead in a queue.

6. Jitter Buffer Delay

Jitter is the variation in arrival rate of voice packets as explained in the next section. In order to compensate for this variation, a jitter buffer has to be introduced at the receiving end to create a constant playout of packets. This buffer unfortunately adds to the overall delay.

It is very important not to set jitter buffers too low or too high. Setting this buffer too low might cause loss of packets since packets arriving outside the jitter buffer delay window will be discarded. Setting it too high causes unnecessary delay. As a result, jitter buffer should be set very carefully, as it would have a negative effect as explained above.

B. JITTER

Jitter is the variation in arrival rate of voice packets at the destination. This is due to the fact that packets will reach their destinations by a number of different routes which introduce varying delays to the packets passing through.

As mentioned in Jitter Buffer Delay discussion above, a jitter buffer smoothes the jitter problem by playing out the voice frames at a constant rate. In order for jitter buffer not to introduce unnecessary delay or packet loss, it should be adaptive, that is, it should monitor timestamps of arriving packets and adjust the buffer size accordingly. This kind of adaptive buffer size will minimize itself in low-delay environments. In environments with varying delay such as the Internet, jitter buffer adapts itself to increasing or decreasing delay variations.

C. PACKET LOSS

Packet loss is common in both private and public IP networks. Packet loss happens when the network is congested with too much traffic or bandwidth is overrun and when the network quality is poor, that is, the network has unsteady network components or underpowered equipments. Those reasons mentioned above are valid for both voice and data packets. However, voice packets unfortunately have one more factor to be discarded; voice packets are discarded when they arrive at their destinations too late to be useful by exceeding the jitter buffer's time limit.

Packet loss is not an issue for data packets which use Transmission Control Protocol (TCP). TCP provides reliability by detecting and retransmitting dropped packets. However, VoIP, which requires real-time transmission of packets, cannot use TCP's retransmission mechanism because a late packet equals a lost packet for a VoIP application and TCP introduces unacceptable delay for voice packets. The next section on *link errors* presents bearable packet loss ratios for different codecs.

D. LINK ERRORS

Link errors affect VoIP calls in several ways:

1. TCP packets carrying signaling messages between gateways might get corrupted and no call establishment would be possible.
2. IP headers might get corrupted which causes IP packets to be dropped at the routers.
3. Voice payloads might get corrupted which causes the quality of voice to degrade.

There are two types of link errors: random bit errors and burst errors. *Random bit errors* occur non-periodically and are measured in terms of bit error rate (BER). *Burst errors* occur when a series of adjacent bits are corrupted. Since random bit errors are more distributed than burst errors, a voice frame is not badly affected by random bit errors. In other words, random bit errors are less destructive than burst errors for voice packets. Table 13 illustrates packet and frame loss percentages for codecs G.723.1 and G729. Voice frame loss values are higher than IP packet loss values. This is because an IP packet reaches its destination with its intact header but damaged voice payload. Bit

Error Rate (BER) of up to 10^{-5} is acceptable. At BER 10^{-4}, the packet loss and frame loss ratios exceed 5% which is a threshold for bearable voice conversion. At BER 10^{-3}, the connections are down [Ref 8].

	G.723.1		G.729	
BER	IP packet loss	Voice frame loss	IP packet loss	Voice frame loss
10^{-7}	0.009%	0.089%	0.006%	0.100%
10^{-6}	0.065%	0.168%	0.052%	0.145%
10^{-5}	0.699%	1.427%	0.558%	0.653%
10^{-4}	6.860%	9.168%	5.363%	5.453%
10^{-3}	Connections down			

Table 13. Packet and frame loss by random errors for G.723.1 and G.729. (From [Ref.8])

Table 14 presents packet and frame loss percentages by burst errors for codecs G.723.1 and G729. The referenced paper uses three parameters to characterize the burst errors:

- Burst length: The length of the burst error in bits. (set at 40 and 60)
- Burst density: The error ratio during the burst. (set at 0.2)
- Gap length: The average time interval between error bursts (set at 1, 10, 100, 1000 milliseconds)

A comparison of packet and voice frame loss values of burst and random errors from Tables 13 and 14 proves the fact that burst errors are more destructive than random bit errors. As Table 14 shows the connections are down or severely damaged in terms of voice quality at BER 10^{-5} while it is possible to have an acceptable or better level of voice quality at the same BER rate.

Burst error (x, y, z)[1]	Equivalent BER	G.723.1		G.729	
		IP packet loss	Voice frame loss	IP packet loss	Voice frame loss
(40, 0.2,1000)	2.60×10^{-7}	0.044%	0.116%	0.026%	0.240%
(40,0.2,100)	2.80×10^{-6}	0.495%	0.592%	0.426%	0.635%
(40,0.2,10)	2.62×10^{-5}	4.527%	4.610%	1.892%	2.106%
(40,0.2,1)	1.30×10^{-4}	Connections down		Connections down	
(60,0.2,1000)	3.85×10^{-7}	0.069%	0.147%	0.034%	0.253%
(60,0.2,100)	3.70×10^{-6}	0.775%	0.830%	0.139%	0.354%
(60,0.2,10)	3.76×10^{-5}	Connections down		3.917%	4.126%
(60,0.2,1)	1.86×10^{-4}	Connections down		Connections down	

[1](x, y, z) – x: average error burst length (bits), y: error density within a burst, z: average error-free period lenght in ms.

Table 14. Packet and frame loss by burst errors for G.723.1 and G.729. (From [Ref.8])

E. ECHO

Echo is an electrical reflection of a signal through the network. Echo is generally the result of a mismatch in impedance between the 4-wire network switch connection and the two-wire local loop. Figure 16 illustrates talker and listener echo. Talker echo occurs when the originator of the speech hears his or her own voice which is reflected from the receiving end. If the reflected echo signal reflects again then it becomes listener echo which is less common than talker echo. *EL* denotes *echo loss* which will be discussed later in this section. Echo can be both useful and detrimental based on the duration of the delay experienced by the echo signal. Echo with a delay of 16- 20 ms is called *sidetone* [Ref.1]. Sidetone is desirable since hearing one's own voice in the handset while one is talking is comforting and reassuring. However, this kind of echo with a delay of more than 25 ms can cause interruptions and distractions in the conversation if it is not handled properly.

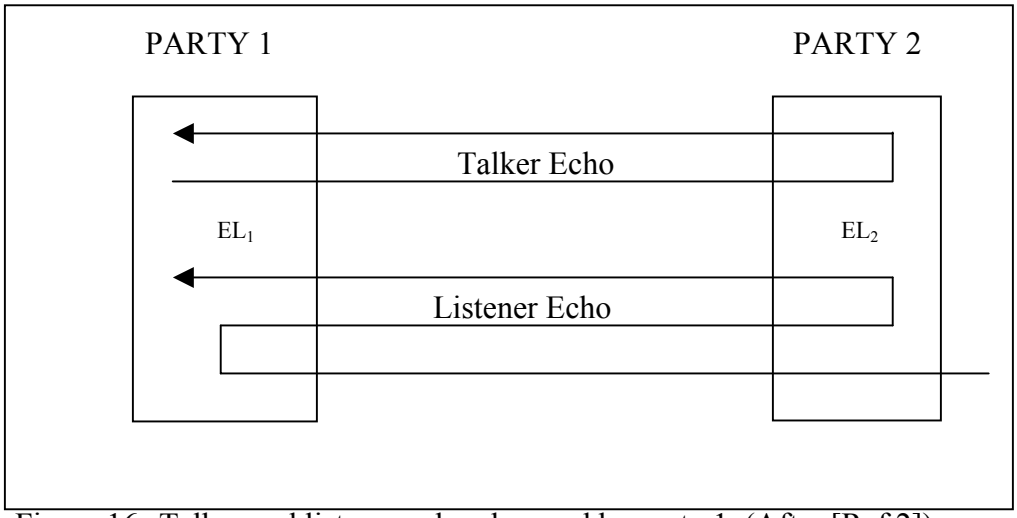

Figure 16. Talker and listener echo observed by party 1. (After [Ref.2])

The circuit-switched networks such as PSTNs handle echo problem with either using echo cancellers or reducing the power level of signals to minimize the signal level of the reflected signals. In packet-switched voice networks, echo cancellers are placed into Digital Signal Processors (DSP).

Echo cancellers work by keeping an inverse pattern of voice signals passing through them for a certain amount of time and then combining this inverse pattern with the echo signal bouncing back from the receiving end.

Table 15 presents tolerable mouth-ear-delays below which traditional (circuit-switched) quality is obtained for different codecs with packet loss values of 0.1, 1 and 5% and echo loss values of 51 dB and infinity. *EL* (Echo loss) is the ratio of the power of voice signal and the reflected echo signal. So, the bigger EL value, the smaller echo problem in the network. Hence, if EL is infinity, then there is no echo. PLC denotes *Packet Loss Concealment* which is a technique to compensate for lost packets with various methods. VAD stands for Voice Activity Detection which helps save bandwidth by not sending packets when there is silence in a conversation. VAD, also known as silence suppression, will be explained later in this chapter. Table 15 clearly shows the negative effect of echo loss for all codec types. For example, the tolerable delay for G.723.1 increases from 192 ms to 250 ms when EL value changes from 51 dB to infinity.

	EL = 51 dB	EL = ∞
(a) Packet loss ratio = 0.1%		
G.711 @ 64 kb/s	282	373
G.711 @ 64 kb/s (+PLC)	296	397
G.729 @ 8 kb/s (+VAD)	221	284
G.723.1 @ 6.3 kb/s (+VAD)	192	250
(a) Packet loss ratio = 1%		
G.711 @ 64 kb/s		
G.711 @ 64 kb/s (+PLC)	264	345
G.729 @ 8 kb/s (+VAD)	195	253
G.723.1 @ 6.3 kb/s (+VAD)	145	221
(a) Packet loss ratio = 5%		
G.711 @ 64 kb/s		
G.711 @ 64 kb/s (+PLC)	195	253
G.729 @ 8 kb/s (+VAD)		
G.723.1 @ 6.3 kb/s (+VAD)		

Table 15. Tolerable mouth-to-ear delays (in ms) below which traditional quality is obtained for different codecs. The empty entries denote the fact that traditional quality cannot be attained. (From [Ref.9])

F. VOICE ACTIVITY DETECTION (VAD)

The last factor affecting voice quality is VAD or *Silence Suppression*. In a normal human conversation, one speaks and other listens. Additionally a normal person has some pauses and breaks in his or her speech. This means that at least 50 percent of bandwidth is wasted if the voice packet generation is not suspended during those silent periods. *Voice activity detection* (VAD) is a technique which detects silent periods in a speech and suspends voice frame generation during those periods. Table 16 presents bandwidth gain with the usage of silence suppression (VAD) for codec G.723.1. It should be noted that

there is no bandwidth gain when there is background noise. This means that VAD cannot distinguish between voice and background noise and keep on generating voice packets all during the speech. Of course, this would not be the case for every noisy background; however, it is harder to adjust the threshold of VAD to distinguish voice from background noise in noisy backgrounds like this particular situation.

Codec	Silence Suppression	Background Noise	Number of IP packets	Number of bytes	IP-level bandwidth (kb/s)	BW gain by silence suppression
G.723.1 5.3 kb/s	ON	Quiet	8047	636,989	5.7	1.88
	OFF	Quiet	15,062	1,203,289	10.7	-
	ON	Car Noise	15,053	1,202,545	10.7	1.00
	OFF	Car Noise	15,053	1,202,569	10.7	-

Table 16. Bandwidth gain by silence suppression (After [Ref. 4])

VAD monitors changes in the power of voice signals to detect the silence and speech periods in conversations. When it detects a drop in the power of the signal, VAD waits approximately 200 ms to make sure that it is not a small pause in speech, and then it ceases to generate voice packets.

VAD faces two major problems; the first problem is called *front-end clipping* which happens when VAD changes from silence-suppression mode to packet-generation mode. Parts of the beginning of speech may be lost during this transition period. The second problem, as mentioned above, is the noisy background which makes it hard for VAD to distinguish between speech and background noise.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. MEASURING VOICE QUALITY

In the previous chapter, various factors that affect voice quality were discussed. This chapter presents brief information about how to measure voice quality.

Voice quality can be measured both *subjectively* and *objectively*. Subjective measurement methods involve a number of human beings who listen and rate voice samples. Subjective speech quality measurement methods are common. However, they are time-consuming and expensive. This chapter will explain Mean Opinion Score (MOS) which is one of the most common subjective voice quality measurement methods.

Objective measurement methods are machine-based and therefore they do not deal with different interpretations of human beings like subjective methods do. Objective methods measure voice quality by computing the quantitative distortion between the sent and the received signals. *Perceptual Speech Quality Measurement* (PSQM) recommended in ITU-T P.861 and *Perceptual Evaluation of Speech Quality* (PESQ) recommended in P.862, are objective ways of measuring speech quality. PSQM assesses the quality of codecs and PESQ may be used for both voice codecs and end-to-end measurements. This chapter does not cover objective measurement methods since these do not contribute to this thesis.

E-Model, developed by the European Telecommunications Standards Institute (ETSI), predicts the subjective voice quality on a network based on the network parameters; that is, it predicts the results of subjective tests such as MOS tests. E-Model has been included into ITU-T standards as G.107. E-Model is the second topic to be explained in this chapter.

A. MEAN OPINION SCORE (MOS)

MOS, specified in ITU-T recommendation P.800, is one of the most commonly used subjective speech quality measurement methods. MOS tests are conducted with human subjects who listen to groups of speech samples and rate the quality of the transmission on a scale ranging from 1 to 5, with 1 being “bad” and 5 being “excellent” as listed in Table 17. Since voice quality is subjective, that is everybody has a unique way of evaluating voice quality, it is important to include a wide range of subjects and speech

patterns to conduct a MOS test to get more accurate results. However, having so many people in tests makes MOS tests expensive and time consuming.

MOS Score	**Listening Quality Scale**	**Listening Effort Scale**
5	Excellent	Complete relaxation possible; no effort required
4	Good	Attention necessary; no appreciable effort required
3	Fair	Moderate effort required
2	Poor	Considerable effort required
1	Bad	No meaning understood with any feasible effort

Table 17. Rating Scales for MOS (From Ref. [10])

MOS helps quantify the performance of different voice codecs. The MOS scores for each voice codec are derived by means of conducting MOS tests as explained above. Table 18 presents MOS values for various voice codecs.

Voice Codec	Bit Rate (kbps)	MOS
G.711	64	4.1
G.726	32	3.85
G.728	16	3.61
G.729A	8	3.7
G.723.1	6.3	3.9
G.723.1	5.3	3.65

Table 18. ITU-T Codec MOS Scoring (After Ref.[11])

B. E-MODEL

E-Model specified in ITU-T Recommendation G.107 produces an R value ranging from 0 to 100. R values can be used to predict the results of subjective tests such

as MOS tests. The higher the R value, the better the voice quality is. Table 19, derived from ITU-T Recommendation G.109, presents the R values and their corresponding ratings that are related to subjective speech quality tests. Additionally Figure 17 displays the R values and their corresponding MOS values.

R-Value Range	Quality Category	User Satisfaction
$90 \leq R \leq 100$	Best	Very satisfied
$80 \leq R \leq 90$	High	Satisfied
$70 \leq R \leq 80$	Medium	Some users dissatisfied
$60 \leq R \leq 70$	Low	Many users dissatisfied
$50 \leq R \leq 60$	Poor	Nearly all users dissatisfied

Table 19. R-Values Translated to Categorical Speech Quality Estimations (From Ref.[13])

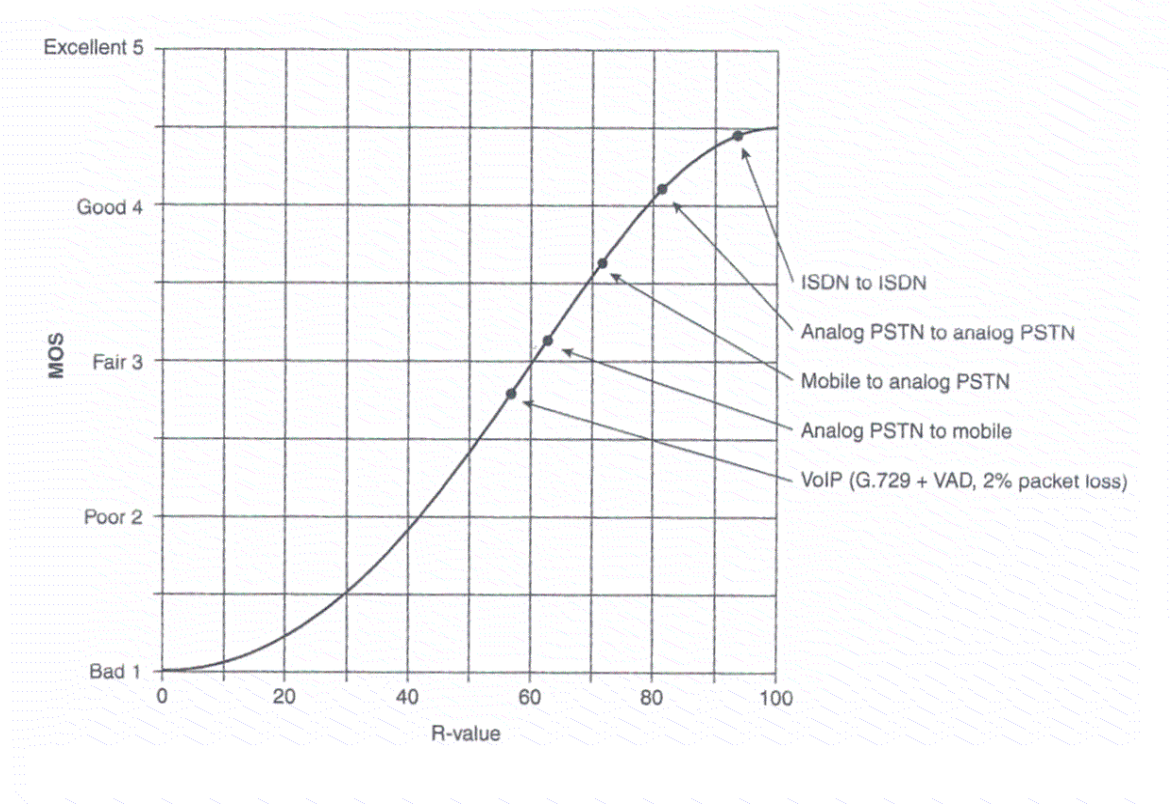

Figure 17. Relationship between R-Values and MOS Scores (From Ref.[14])

The equation for the R value is: $R = R_o - I_s - I_d - I_e + A$. The components of the equation are presented in Table 20.

Impairment	Caused By...
R_o	Noise floor (room noise at receive and send sides, circuit noise)
I_s	Simultaneous factors (excessive loudness, sidetone, quantizing distortion)
I_d	Delayed factors (talker echo, listener echo, absolute delay)
I_e	Special equipment (non-waveform low bit-rate codecs, packet network effects and so on)
A	Advantage, or Expectation factor: reduced expectation of end-users, because of another benefit
R	Calculated Planning Impairment Factor: indicates the expected level of user satisfaction

Table 20. Calculation of the R-Value, as Used in the E-Model (From Ref.[12])

R_o involves effects of background noise and circuit noise. I_s involves factors that occur simultaneously with the voice such as quantizing distortion and sidetone. I_d includes delayed factors which happen after the voice such as echo. I_e includes distortion impairment caused by use of special equipment such as low-bit-rate codecs. For example, each codec has an I_e value which increases as packet loss ratio increases as shown in Figure 18. The expectation factor, A, is the reduced expectation as a result of having an advantageous situation, for example, a wireless user accepts lower quality voice because of the convenience.

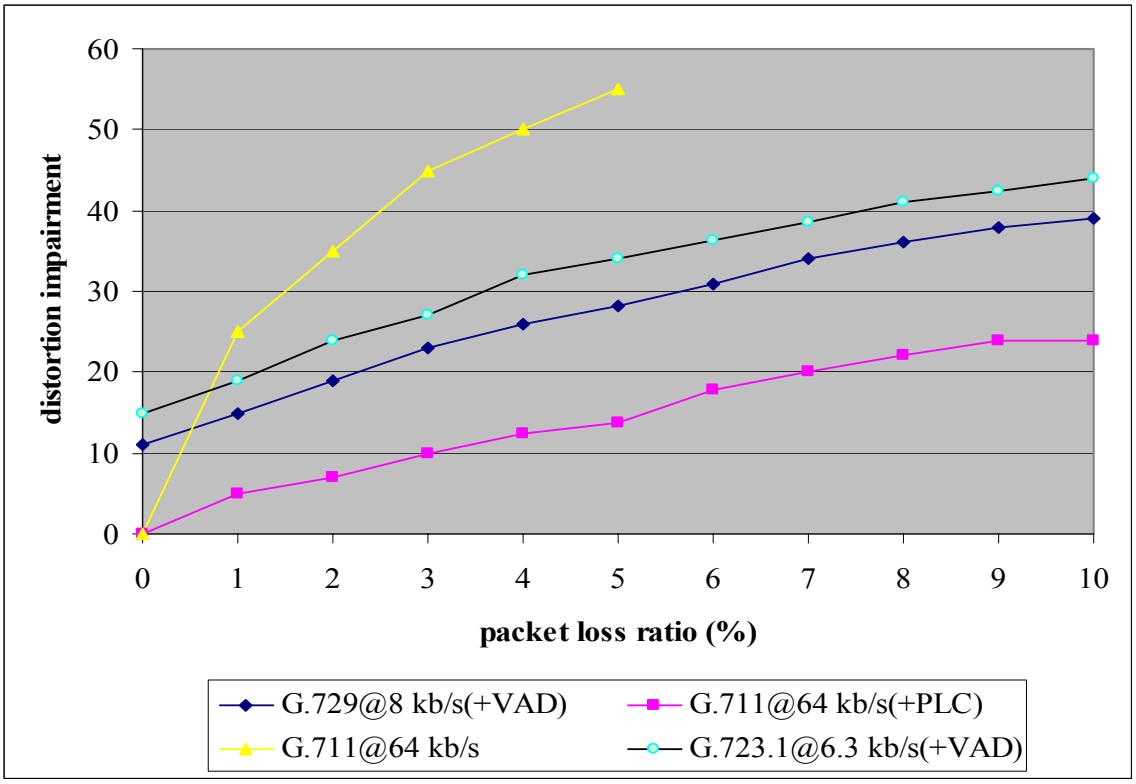

Figure 18. The distortion impairment term I_e as a function of the packet loss ratio for different codecs. (After Ref. [9])

This chapter explained MOS, one of the most common subjective voice quality measurement methods and E-Model which is helpful to predict the results of subjective tests such as MOS tests. Next chapter presents, tests, and finally analyzes six VoIP network implementation suggestions.

THIS PAGE INTENTIONALLY LEFT BLANK

V. VOIP IMPLEMENTATION

The previous two chapters covered the factors affecting voice quality and the ways to measure it. As explained in Chapter III, the major factors affecting voice quality are delay, jitter, packet loss, link errors, echo, and voice activity detection (VAD). This chapter first presents VoIP design suggestions and then tests six VoIP configurations to demonstrate the affects of some of these suggestions and finally analyzes the test results of proposed configurations.

A. VOIP NETWORK DESIGN SUGGESTIONS

Before discussing the design suggestions, the attributes of two different VoIP environments in terms of manageability should be reviewed. The first environment is the *Internet* which is not managed by any authority and the second one is *private networks* such as LANs which are controlled and predictable networks.

The Internet is a collection of thousands of networks with millions of users. Everyday new users and networks are signing on to use the Internet's limited bandwidth. This heavy use of the Internet causes congestion which in turn causes delay, jitter and packet losses. In addition, because of the Internet's connectionless Internet Protocol (IP), the individual voice packets may travel over different paths which complicates the matter further. More importantly, there is no single authority which controls the Internet. All these factors make the Internet an unreliable and unpredictable means with limited bandwidth for VoIP. Because of all these reasons, there is little in the way of controlling VoIP over the Internet. Consequently, some of the implementation suggestions in this chapter are not applicable for VoIP over the Internet.

The second environment for VoIP is private networks. Private networks have more predictable and manageable bandwidth. Currently, the implementation of VoIP in the private networks is becoming more common and many corporations and institutions shifting to VoIP technology. In fact, there is no reason to fail provided that a careful pre-deployment audit of the network is done. This kind of an audit should include bandwidth consumption, packet loss, delay, and jitter values for the network. As a result of the audit,

the reasons for unacceptable results can be eliminated to make the network suitable for VoIP; for example, upgrading the old network switches or increasing network bandwidth.

The following sections include suggestions to decrease the negative impacts of factors affecting voice quality. However, it should be noted that these suggestions are only part of the solution. A complete solution for a successful VoIP implementation requires a wider consideration which includes the interoperability, security, QoS and transmission media issues. Those issues are beyond the scope of this thesis.

1. Use Compressed Real-Time Transport Protocol (CRTP)

As previously mentioned, the performance of VoIP calls depend mainly on the bandwidth therefore, every effort to save bandwidth is important. One of the most bandwidth-inefficient aspects of voice packets is the header overhead as presented in Table 21. Disregarding the link layer header, it can be seen that the IP/UDP/RTP header totals 40 bytes.

LINK Header	**IP Header**	**UDP Header**	**RTP Header**	**Voice Payload**
Variable size depending on link layer protocol	20 bytes	8 bytes	12 bytes	Variable size depending on codec

Table 21. Header overhead without CRTP

Considering that the default voice payload size of the mostly used low-bit-rate codecs is 20 bytes, it can be seen that the IP/UDP/RTP header is *twice* the size of actual voice. In other words almost *two third of the bandwidth is wasted for carrying the header information.* The solution to this inefficient situation is the use CRTP which compresses the IP/UDP/RTP header from 40 bytes to 2 or 4 bytes as seen in Table 22. This significant saving is achieved by not sending the entire IP, UDP, and RTP headers with every voice packet, only the header information that changes from one packet to the next, such as sequencing information, is sent Ref [23].

LINK Header	IP Header	UDP Header	RTP Header	Voice Payload
Variable size depending on link layer protocol	2 or 4 bytes			Variable size depending on codec

Table 22. Header size with CRTP

CRTP produces a 2 or 4 bytes header depending on whether the CRC field of UDP protocol is used or not. The CRC field is used for error checking purposes and consumes 2 bytes. Since link layer protocols usually include a CRC field for the same purpose, it is a good idea to disable the checksum field of the UDP when using CRTP. This reduces the header size from 4 bytes to 2 bytes, which results in a significant bandwidth saving for VoIP. Whether it is 2 or 4 bytes, the respective bandwidth gain is significant. Last thing to note about CRTP is that both ends need to be configured for CRTP since it is configured on a link-by-link basis.

2. Enable Voice Activity Detection (VAD)

VAD is a technique which detects silent periods in a speech and suspends voice frame generation during those periods. Since a normal human conversation is half-duplex that is, one speaks other listens, VAD may save up to 50% of bandwidth, which is a significant improvement. So, it is highly recommended to enable VAD. Chapter III discussed two problems: the *front-end clipping* which happens when VAD changes from silence-suppression mode to packet-generation mode. Parts of the beginning of speech may be lost during this transition period. The second problem is the *noisy background* which makes it hard for VAD to distinguish between speech and background noise. One more problem to note is that one may think that the connection is lost as a result of complete silence. Therefore a comfort noise should be sent during those silent periods. Silence Insertion Description (SID) voice payloads alleviate this problem by being generated during these silent periods.

Table 23 presents bandwidth gains as a result of using VAD and CRTP. First, each codec's bandwidth consumption without CRTP and VAD is presented, second CRTP is introduced and finally both CRTP and VAD are introduced. Voice activity is set

to 65%, which means conversion is happening during %65 of the duration of the call so, bandwidth is saved by not sending voice packets 35% of call time. The explanation of required bandwidth calculation will be offered further in this chapter in the test results section.

Codec	Codec Bit Rate (kbps)	Payload Size (Byte)	IP/UDP/RTP Header Size (Byte)	VAD (65%)	Required Bandwidth (kbps)
G.711	64	160	40	__	82.955
G.711	64	160	4 (CRTP)	__	68.52
G.711	64	160	2(CRTP)	Yes	45.8
G.729B	8	20	40	__	26.955
G.729B	8	20	4(CRTP)	__	12.52
G.729B	8	20	2(CRTP)	Yes	9.439
G.723.1	5.3	20	40	__	18.022
G.723.1	5.3	20	4(CRTP)	__	8.387
G.723.1	5.3	20	2(CRTP)	Yes	6.354

Table 23. Bandwidth gain with the use of CRTP and VAD with different codecs

3. Use Small Voice Payload

Since delay is very important for VoIP applications, these should generate small voice payloads in order to avoid the delay incurred when filling and sending the payload. The motivation behind choosing large voice payloads is the goal of using bandwidth efficiently; using small packets may lead to VoIP packets with too little actual voice data and too much header data which means inefficient use of bandwidth capacity. Shortly,

there is a tradeoff between bandwidth utilization efficiency and delay and thus voice quality. However, it should be noted that using CRTP significantly reduces header overhead as explained before.

The following example illustrates how serialization delay increases with the increase of voice payload size. Suppose that there are two voice payloads one of which is 160 bytes (1280 bits) and the other one is 720 bytes (5760 bits). The first payload size is the default size for codec G.711 and the second one is one of the other less optimal possible choices for codec G.711. In addition, suppose that the transmission rate of the link between the source and destination is 64 kbps. By dividing the voice payload size with the speed of the link, it is found that it takes the source 20 milliseconds to send the small payload and 90 milliseconds to send the big payload. The difference between the two times is 70 ms which is a huge delay for voice and it may easily be the difference between toll quality voice and degraded voice. As mentioned earlier in Chapter III, there are also other delay factors which make the overall delay and serialization delay difference alone in this example is 70 ms which is almost half the maximum allowable delay of 150 ms to obtain toll quality.

4. Use Adaptive Jitter Buffer

Jitter is the variation in the arrival rate of voice packets at the destination as discussed in Chapter III. Jitter happens because individual packets follow different routes which introduce varying delays. For this reason, voice packets reaching their destination cannot be played out instantly. This problem is solved with a *jitter buffer* which smoothes the jitter problem by playing out the voice frames at a constant rate after storing them for a short time.

It is very important not to set jitter buffers too low or too high. Setting this buffer too low might cause loss of packets while setting it too high causes unnecessary delay. However, the delay may vary in some network environments and an ideal jitter buffer set for one time may become too low or too high for another time due to changing delay characteristics.

In order for jitter buffer not to introduce unnecessary delay or packet loss, it should be adaptive rather than static; that is, it should monitor timestamps of arriving

packets and adjust the buffer size accordingly. This kind of adaptive buffer is very crucial in environments with varying delay such as the Internet.

5. Choose the Appropriate Codec

A *Co*der *Dec*oder (Codec) digitizes the analog voice and vice versa as discussed in Chapter III. It is very important to choose the codec wisely; otherwise a would-be successful implementation may turn out to be a failure. Table 24 presents the characteristics of several codecs.

Codec	Bit Rate (kbps)	Codec processing delay (ms)	Lookahead Time (ms)
G.711	64	0.75	0
G.726	32	1	0
G.728	16	3 to 5	0
G.729A	8	10	5
G.723.1	6.3	30	7.5
G723.1	5.3	30	7.5

Table 24. Characteristics of Several Voice Codecs (After Ref. [5] and [7])

When bandwidth capacity becomes an issue, low-bit rate codecs should be chosen, however these introduce much delay because of their complex compression algorithms. Compressing voice signals reduces bandwidth requirements at the expense of higher computational time which means more delay and thus degraded voice quality. Therefore, choosing a low-bit-rate codec might cause delay to become an issue. A good recommendation at this point would be the codecs G.726 and G.728 which are in between in terms of bandwidth consumption and processing delay.

Another important point is not to choose low-bit-rate codecs if burst errors are common in a network. Since low-bit-rate codecs' bits represent more information, burst errors may cause much voice to be lost.

To summarize, a tradeoff between bandwidth and voice quality has to be considered based on the available computational power, desired voice quality and network bandwidth requirements.

6. Use Echo Canceller

Echo is an electrical reflection of a signal through the network. To the user, it is hearing his or her own voice from the telephone headset. Echo is generally the result of a mismatch in impedance between the 4-wire network switch connection and the two-wire local loop as explained in Chapter III.

Echo can be both useful and detrimental based on the duration of the delay experienced by the echo signal. Echo with a delay of up to 20 ms is called *sidetone* which is even desirable since hearing one's own voice in the handset while one is talking is comforting and reassuring. However, echo with a delay of more than 25 ms can cause interruptions and distractions in the conversation and it must be handled with echo cancellers.

Echo cancellers work by keeping an inverse pattern of voice signals passing through them for a certain amount of time and then combining this inverse pattern with the echo signal bouncing back from the receiving end. Additionally, they do not require much computational time which is a very good aspect since even little delays of several milliseconds are important for voice communications. Table 15 in Chapter III illustrates the importance of echo cancellers by presenting the tolerable delays for different codecs with and without echo cancellation.

B. DESCRIPTION OF EXPERIMENT

This section describes and analyzes six different VoIP configurations. The testing software tool used is *Westplan* which was developed by Westbay Engineers Ltd. of Crawley, UK, [Ref. 18].

The configurations are made using three kinds of codecs: high-bit-rate, low-bit-rate, and lowest-bit-rate. Further, each of these codecs has two configurations: one using header compression and voice activity detection, and the other using uncompressed header and no voice activity detection. The only exception is the high-bit-rate codec which does not use voice activity detection in both configurations because the software

does not support VAD feature for codec G.711. Table 25 illustrates these configurations which are explained further in this chapter.

Configuration	Codec	Header	VAD
1	**High-bit-rate**	**Uncompressed**	**Disabled**
2	**High-bit-rate**	Compressed	**Disabled**
3	Low-bit-rate	**Uncompressed**	**Disabled**
4	Low-bit-rate	Compressed	Enabled
5	**Lowest-bit-rate**	**Uncompressed**	**Disabled**
6	**Lowest-bit-rate**	Compressed	Enabled

Table 25. Test configurations

The test assumes a company having two geographically separated locations. It is further assumed that 100 call channels are required at the busiest hour traffic between the two locations. This translates into 84 Erlangs. *Erlang* is a unit of telecommunications traffic measurement [Ref.18]. Erlang is used to describe the total traffic volume of one hour. Below is a sample Erlang calculation:

Minutes of traffic in an hour = number of calls * average duration of calls

Minutes of traffic in an hour = 1260 * 4 = 5040 minutes

Hours of traffic in an hour = 5040/60 = ***84 Erlangs***

The two locations are connected via a leased dedicated T1 line which operates at a rate of 1.544 Mbps. The link layer protocol used is Point-to-Point Protocol (PPP) which has 7 bytes of header overhead.

The codecs used in the experiment are G.711, G.729B, and G.723.1. The reason for choosing G.729B and G.723.1 is that both are low-bit rate and widely used codecs and they provide significant bandwidth savings with their compression algorithms. G.711 which operates at a rate of 64 kbps was chosen to demonstrate the bandwidth inefficiency associated with this codec which does not compress the voice. The default packet interval times used in the test for voice packets are 20 ms for G.711 and G.729B and 30 ms for G.723.1.

Figure 19 presents the blocking target and busy hour factors for the test configurations. The blocking target is set to 0.1% which means one call out of 1000 calls may be blocked. The busy hour factor is set to 17% which means that 17% of daily minutes are offered during the busiest hour of the day.

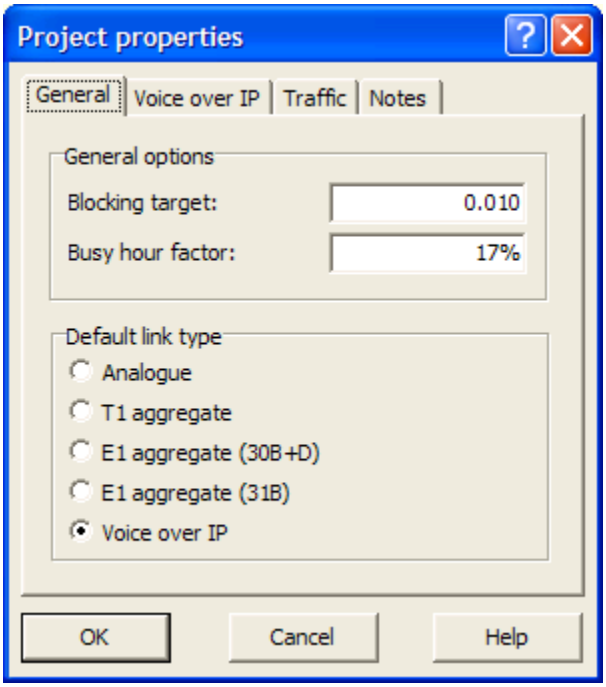

Figure 19. Blocking target and busy hour factor

Each codec was tested with two different configurations: one being bandwidth-inefficient and other being bandwidth-efficient. The first configuration uses the regular uncompressed 40 bytes RTP/UDP/IP header and RTCP as control protocol. Additionally, VAD is disabled in this configuration. A sample configuration for the codec G.711 is shown in Figure 20.

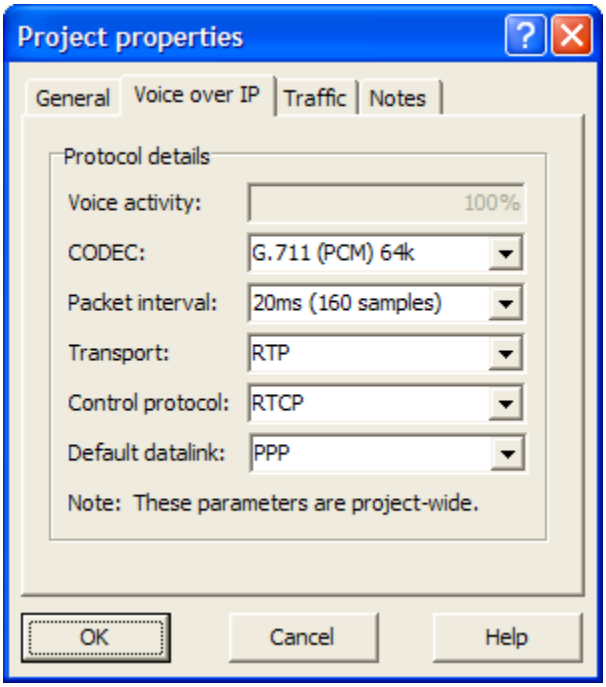

Figure 20. Bandwidth-inefficient configuration

The second configuration uses 2 bytes CRTP header and CRTCP. Second configuration sets VAD to 65% which results in an approximate 35% bandwidth saving. Figure 21 presents a sample configuration for G.729B. The only exception to this configuration is the codec G.711 which does not use VAD.

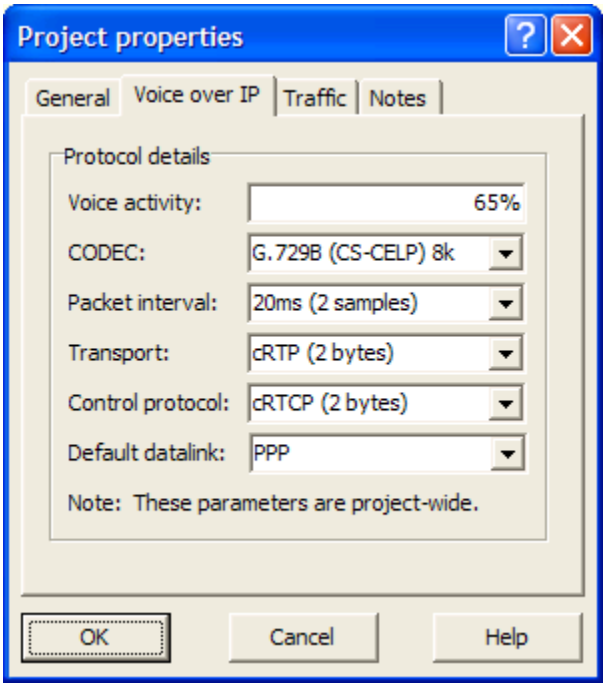

Figure 21. Bandwidth efficient configuration

As a result of applying the two configurations to the three selected codecs, six configurations came out. The summary of these six configurations to be tested is presented in Table 26.

Configuration #	Codec	Codec Bit Rate (kbps)	Transport Protocol	Control Protocol	Voice Activity
1	G.711	64	RTP	RTCP	100%
2	G.711	64	CRTP	CRTCP	100%
3	G.729B	8	RTP	RTCP	100%
4	G.729B	8	CRTP	CRTCP	65%
5	G.723.1	5.3	RTP	RTCP	100%
6	G.723.1	5.3	CRTP	CRTCP	65%

Table 26. Test configurations

C. TEST RESULTS

As a result of running the six configurations with Westplan, the following results in Table 27 were gathered. *Bandwidth per channel* is the required bandwidth for one voice channel. *Available number of channels* is the number of channels that can be accommodated with the available bandwidth capacity of 1544 kbps. *Blocking*

experienced is the percentage of calls that have been blocked due to insufficient bandwidth and *optimum bandwidth* is the overall bandwidth required to accommodate the 84 Erlangs of traffic.

Configuration #	Bandwidth per channel (kbps)	Available # of channels	Blocking experienced	Optimum bandwidth (Mbps)
1	82.960	18	78.9%	8.296
2	67.714	22	74.2%	6.772
3	26.955	57	34.3%	2.696
4	9.439	163	0%	0.944
5	18.022	85	7.5%	0.181
6	6.354	243	0%	0.636

Table 27. Test results for configurations

The Westplan makes the calculations based on assumptions and user input data. The following paragraphs explain how the calculations are made by presenting the calculations for configuration 1:

As stated previously, the blocking target for all configurations is 0.01 and busy hour factor is 17%. Available bandwidth is 1544 kbps and busy hour traffic is 84 Erlangs. This configuration's voice activity is set to 100% which means VAD is disabled.

Number of IP datagrams containing voice per second

The selected interval of 20ms and the voice activity percentage of 100% result in 50 IP datagrams per second being generated which contain active voice samples.

Size of voice payload in each datagram

The size of the voice payload in each datagram is 1280 bits (160 bytes) using the G.711 (PCM) 64k compression scheme.

Total size of each voice datagram

IP, UDP and RTP add an overhead of 320 bits (40 bytes), making the total size of each datagram 160+40 = 200 bytes (1600 bits)

Silence Insertion Description (SID) packets generation

The Voice Activity was set to 100%. Therefore, no Silence Insertion Description frames will be transmitted.

RTCP packet size and frequency

0.2 control packets are generated each second as suggested by RFC 1889 and RTCP payload size will be 496 bit (62 bytes), additionally IP and UDP headers add an overhead of 224 bits (28 bytes), making a total RTCP IP datagram size of 720 bits (90 bytes).

PPP header and frame sizes

Each PPP frame adds an overhead of 56 bits (7 bytes) to an IP datagram (including one flag). Accordingly, the PPP frame sizes are:

PPP frames carrying Voice IP packets = 1600 + 56 = 1656 bits (207 bytes).

PPP frames carrying RTCP IP packets = 720 + 56 = 776 bits (97 bytes).

Bandwidth required per voice channel

By multiplying the frequency of the packets with their size, the bandwidth occupied by each type of frame is calculated as follows:

Voice bandwidth = 1656 * 50 = 82,800 bps = 82.8 kbps.

SID bandwidth = 0 bps.

RTCP bandwidth = 776 * 0.2 = 155 bps = 0.155 kbps

Total bandwidth per voice channel = 82.8 + 0.155 = **82.955** kbps

D. ANALYSIS OF TEST RESULTS

This section analyzes the test results presented in Table 27. Figure 22 illustrates the required bandwidth per voice channel for all six configurations. The first thing to note is that the Configurations 1 and 2 using the codec G.711, which does not compress, even exceed the 64 kbps bandwidth consumed when using the circuit-switched networks. In other words, those two configurations are even worse than circuit-switched calls in terms of bandwidth consumption.

Another important thing is the significant bandwidth savings gained with the use of low-bit-rate codecs even without the use of CRTP and VAD. Configuration 3 and 5 consume 26.955 and 18.022 kbps respectively. The addition of CRTP and VAD capabilities results in bandwidth consumptions of 9.439 and 6.354 kbps for configurations 4 and 6 respectively. Of course, these tremendous bandwidth savings are achieved at the expense of more computational time which means more delay and less intelligent representation of voice. However, as long as the delay is less than 150 ms and the links are error free, one of these bandwidth-efficient solutions may fit to the needs of a particular network.

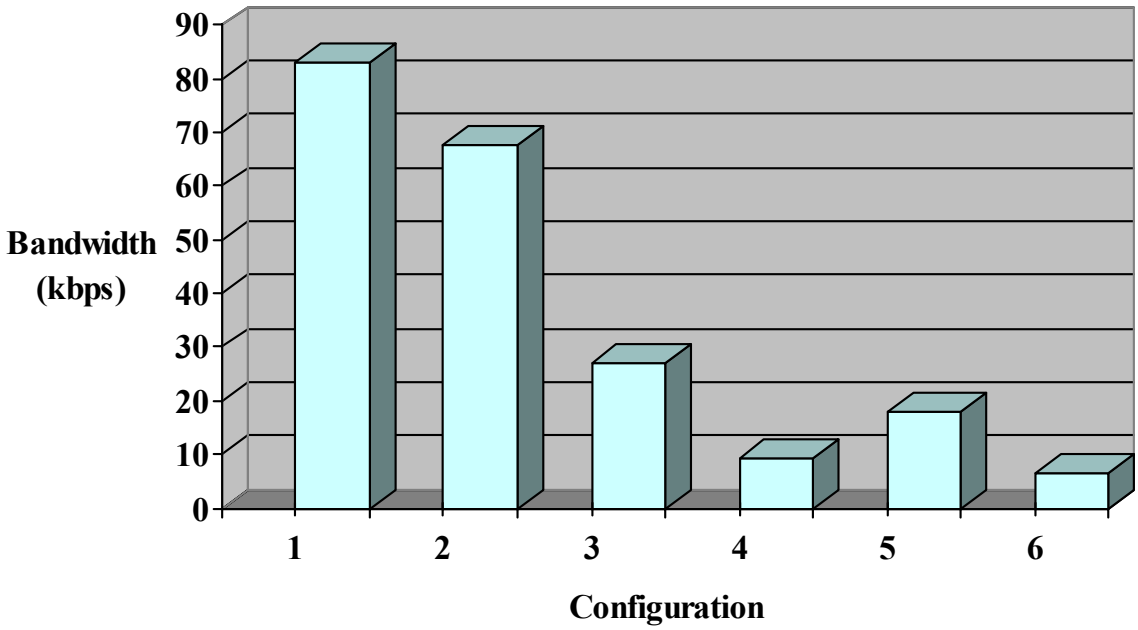

Figure 22. Total bandwidth consumed per voice channel in Kbps

Figure 23 presents numbers of voice channels available for each configuration. It is necessary to note that the number available channels were calculated assuming that all bandwidth capacity is to be used for voice communication only. However, in a typical network, some portion of available bandwidth should be considered spared for network management trafficking.

The bold line in Figure 23 indicates the required 100 channels for the company. Again, it is very clear that the codec G.711 with or without CRTP (Conf 1 and 2) does

not meet the needs of the company. Conf 1 and 2 provide 18 and 22 channels respectively. Even circuit-switched networks provide 24 channels with a T1 line.

Another important thing is that even low-bit-rate codecs cannot satisfy the needs of the company when not using CRTP and VAD (Conf 3 and 5). However, the introduction of CRTP and VAD (Conf 4 and 5) significantly increases the number of channels. For example, codec G.723.1's available channels jump from 85 to 243 (Conf 5 and 6) with the use of CRTP and VAD which means a tremendous *threefold increase*. These numbers show the fact that CRTP and VAD are very important features when implementing VoIP.

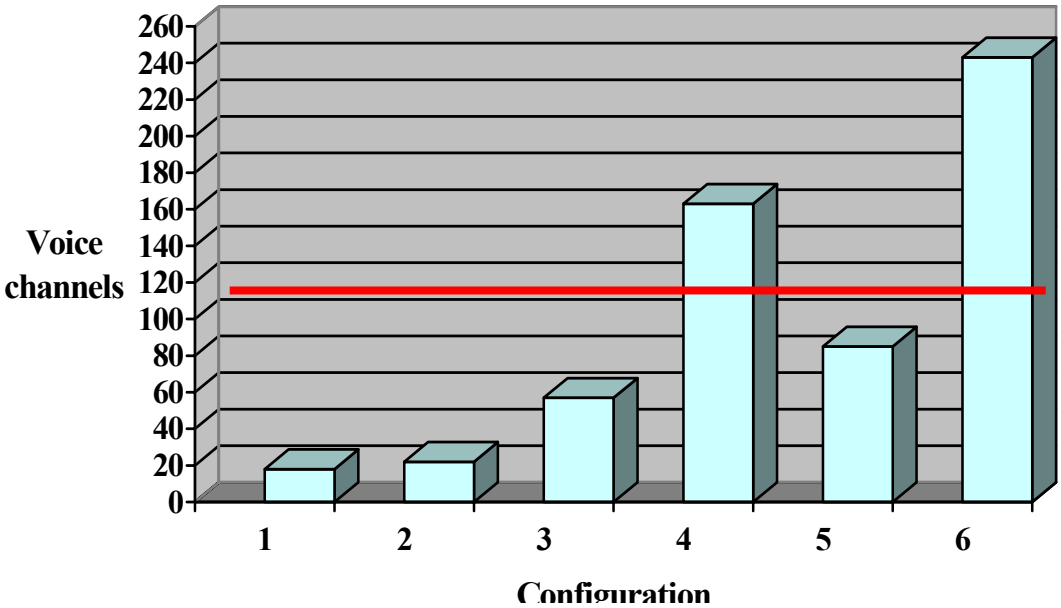

Figure 23. Total number of voice channels with the available bandwidth on T1.

Figure 24 presents the percentage of blocked calls due to insufficient network capacity. The figure shows that all configurations experience blocking except the ones which employ CRTP and VAD.

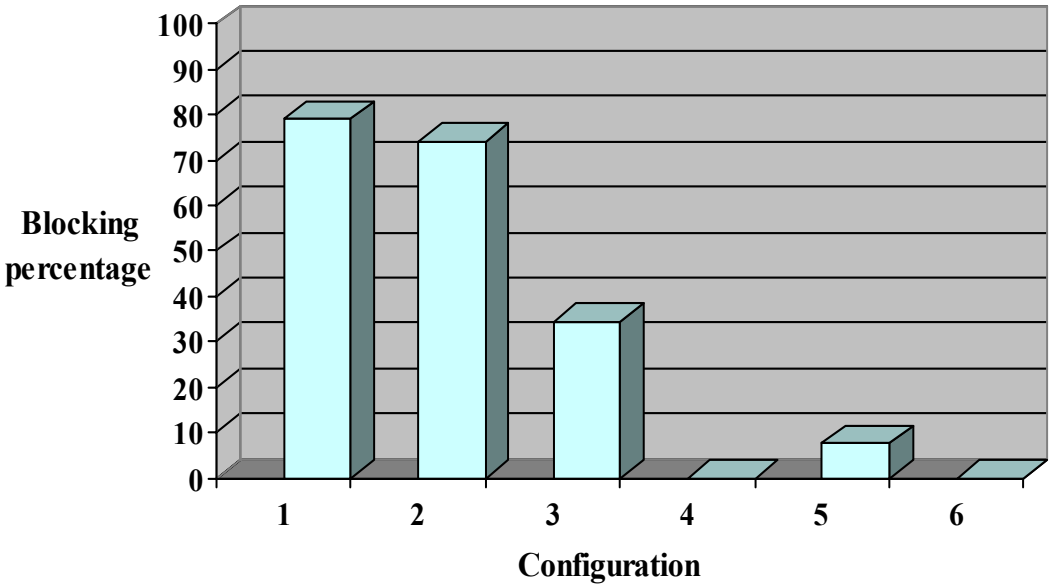

Figure 24. Percentage of blocking experienced

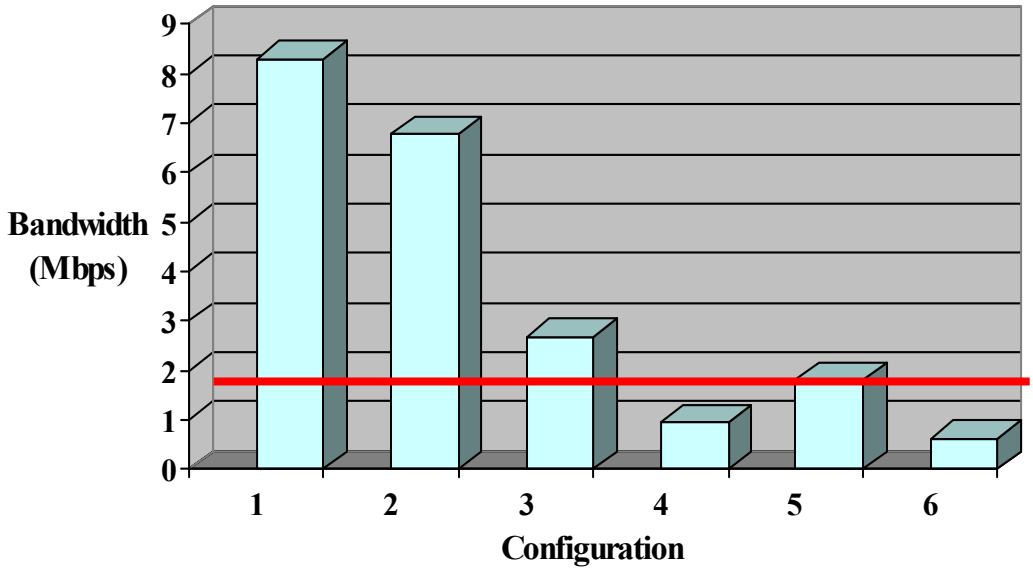

Figure 25. Optimum bandwidth for each configuration in Mbps

Figure 25 illustrates the optimum bandwidth numbers for each configuration. Configuration 1 requires 8.296 Mbps of bandwidth capacity which translates into 5.38 times more than what there is available (1.544 Mbps) as shown with a bold line in the figure. Assuming that a T1 line costs company $15,000 a year, then IT budget will have to increase $75,000 to lease five more T1 lines to meet the needs. On the other hand,

Configurations 4 and 6 need 0.944 Mbps and 0.636 Mbps respectively to meet the same demands. The current T1 line that the company has is more than enough for those configurations. They even leave bandwidth to be used for other purposes such as data traffic.

E. SUMMARY

The first part of this chapter presents six VoIP network design suggestions. These suggestions are the use of CRTP and VAD which results in significant bandwidth savings; the use of small size voice payloads, echo cancellers, and adaptive jitter; and the considerations about choosing the appropriate codec are presented.

The second part of this chapter reports tests with six different VoIP configurations to demonstrate the effects of several of these suggestions. The configurations are made using three kinds of codecs: high-bit-rate, low-bit-rate, and lowest-bit-rate. Further, each of these codecs has two configurations: one using header compression and voice activity detection, and the other using uncompressed header and no voice activity detection. In other words, one being bandwidth-efficient and the other bandwidth-inefficient.

The results of the tests show that high-bit-rate codecs are even worse than circuit-switched calls in terms of bandwidth consumption because of the additional header overhead. Another important result is the significant bandwidth savings gained with the use of low-bit-rate codecs even without the use of CRTP and VAD. The use of CRTP and VAD with low-bit-rate codecs further increased the advantage. Particularly, a *threefold increase* was achieved in the number of available voice channels with the use of CRTP and VAD.

THIS PAGE INTENTIONALLY LEFT BLANK

VI. CONCLUSION

This thesis first presents the factors affecting the growth of Internet telephony and then it explains the most widely used Internet telephony protocols in Chapter II. Later, factors affecting voice quality are analyzed in Chapter III. The following chapter explains methods to measure voice quality. Lastly, Chapter V presents six VoIP network design suggestions and reports tests with six different VoIP configurations to demonstrate the effects of several of these suggestions. These suggestions are listed below:

1. Use Compressed Real-Time Transport Protocol (CRTP).
2. Enable voice activity detection (VAD).
3. Use small voice payload.
4. Use adaptive jitter buffer.
5. Choose the appropriate codec.
6. Use echo canceller.

The results of the research and tests reveal that the performance of IP telephone calls depend heavily on the bandwidth. Therefore, various techniques which help save bandwidth should be used for successful IP telephony implementation. The results of the tests showed that up to 65% of the bandwidth can be saved with the use of these techniques. These techniques are header compression and voice activity detection.

Another result is that even low-bit-rate codecs, which may consume up to eight times less bandwidth, may not be able to meet the demand for available number of voice channels without using header compression and voice activity detection. One more interesting thing to note is that the uncompressing high-bit-rate codecs may consume more bandwidth than traditional circuit-switched voice calls.

Considering the bandwidth saving measures such as header compression, voice activity detection and using low-bit-rate codecs, one may think that it is a better idea to upgrade network bandwidth instead of struggling with these techniques which may degrade voice quality because of their specific requirements as explained in Chapters III

and V. The author believes that this is not a good idea since increased bandwidth capacity will trigger the emergence of new bandwidth-hungry applications and this will result in bandwidth to be an important issue again. Therefore, every application should consume the least possible bandwidth; however, this does not mean accepting poor quality for the sake of bandwidth saving. For example, if delay is more than acceptable for an IP telephony implementation which uses a low-bit-rate codec to save bandwidth, then switching to a higher-bit-rate codec may be considered to increase quality at the expense of bandwidth consumption.

To summarize, a tradeoff between bandwidth and voice quality has to be considered based on the available computational power, desired voice quality and network bandwidth requirements when implementing IP telephony.

A. FUTURE WORK

This thesis focused primarily on voice quality problems of VoIP. However, there are a number of related topics that need investigation; one interesting topic may be security issues of VoIP. Since VoIP uses packet switching, the same kinds of vulnerabilities faced by packet networks are valid for VoIP such as Denial of Service (DoS) Attacks. Another topic may be Quality of Service (QoS) for VoIP such as differentiated services and queuing techniques. Additionally, interoperability is a major problem. VoIP applications which implement different protocols cannot talk to each other. Moreover, even VoIP applications implementing the same protocol such as H.323 may not be able to interoperate because of proprietary differences. Lastly, VoIP via satellites, in particular, via low-earth-orbit satellites which have very little and predictable transmission delays may be an interesting topic for further study.

LIST OF REFERENCES

[1] Douskalis, Bill, *IP Telephony, The Integration of Robust VoIP Services*, pp. 231-232, Prentice-Hall, Inc.,2000

[2] Janssen, Jan and others,"*Delay Bounds for Voice Over IP Calls Transported over Satellite Access Networks*", Mobile Networks and Applications 2002, p. 80

[3] Keagy, Scott, " *Integrating Voice and Data Networks*", pp.208, Cisco Press, 2000

[4] Nguyen, Thuan and others, " *Voice over IP Service and Performance in Satellite Networks*", IEEE Communications Magazine March 2001, p.167

[5] Goode, Bur, "*Voice Over Internet Protocol (VoIP)*", Proceedings of The IEEE, Vol 90, No.9, September 2002, p.1497

[6] Dechjaroen, Chaiporn, *Performance Evaluation of Voice Over Internet Protocol*. Master's Thesis, Naval Postgraduate School, Monterey, California, December 2002

[7] Cisco Systems, "*Voice/Data Integration Technologies*", [http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito_doc/voicdata.htm], May 2003.

[8] Nguyen, Thuan and others, " *Voice over IP Service and Performance in Satellite Networks*", IEEE Communications Magazine March 2001, p.168

[9] Janssen, Jan and others," *Delay Bounds for Voice over IP Calls Transported over Satellite Access Networks*", Mobile Networks and Applications 2002, p. 81.

[10] Keagy, Scott, " *Integrating Voice and Data Networks*", p.208, Cisco Press, 2000

[11] Davidson, Jonathon and Peters, James, " *Voice over IP Fundamentals*", p.174, Cisco Press, 2000

[12] Keagy, Scott, " *Integrating Voice and Data Networks*", p.166, Cisco Press, 2000

[13] Keagy, Scott, " *Integrating Voice and Data Networks*", p.164, Cisco Press, 2000

[14] Keagy, Scott, " *Integrating Voice and Data Networks*", p.167, Cisco Press, 2000

[15] Goode, Bur, "*Voice Over Internet Protocol (VoIP)*", Proceedings of The IEEE, Vol 90, No.9, September 2002, pp.1505-1508

[16] IDC, "*How the Internet Will Affect Your Next Phone System*", white paper, [http://www.teledynamic.com/docs/misc/VOIPSt.pdf], June 2003

[17] Zahorujko,Ian and others, "*IP Convergence in Global Telecommunications-Voice Over Internet Protocol(VoIP)",* p.11, DSTO, 2000

[18] Westbay Engineers Ltd., www.erlang.com, June 2003

[19] Packetizer, Inc, "*H.323*",[http://www.packetizer.com/iptel/h323/], June 2003

[20] ITU-T Recommendation H.323, Packet-based multimedia communications systems.

[21] Miller, Mark A., P.E, "*Voice Over IP Technologies*", pp. 235-236, M&T Books, 2002.

[22] Intel Corporation, "*IP Telephony Basics*", [http://www.intel.com/network/csp/resources/white_papers/4070web.htm], June 2003

[23] Cisco Systems, "*Beyond Basic IP*", [http://www.cisco.com/warp/public/779/servpro/promotions/bbip/volume_01_issue23.html], September 2003

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Professor Burt Lundy
   Naval Postgraduate School
   Monterey, California

4. R. Scott Coté
   Naval Postgraduate School
   Monterey, California

5. Kara Kuvvetleri Komutanligi Kutuphanesi
   Yucetepe
   Ankara, Turkey

6. Kara Harp Okulu Kutuphanesi
   Bakanliklar,
   Ankara, Turkey

7. Lutfullah Tasyumruk
   Kadikoy
   Istanbul, Turkey