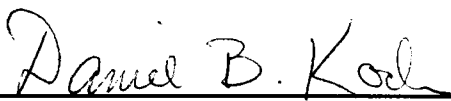
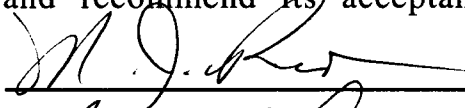
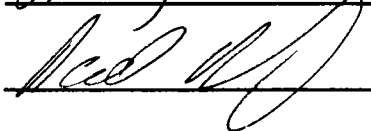


To the Graduate Council:

I am submitting herewith a thesis written by Raymond Mounib El-Bitar entitled "Desktop Videoconferencing: a Network Study." I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Electrical Engineering.


Dr. Daniel B. Koch, Major Professor

We have read this thesis
and recommend its acceptance:

Accepted for the Council:


Associate Vice Chancellor
and Dean of The Graduate School

Desktop Videoconferencing: A Network Study

A Thesis presented
for the
Master of Science Degree

The University of Tennessee, Knoxville

Raymond M. El-Bitar
May 1997

Acknowledgments

The author would like to express his appreciation to Dr. Daniel Koch for serving as his major professor. The author is indebted to him for his help, patience and guidance. He also wishes to thank Dr. Michael Roberts and Dr. Paul Crilly for their time and participation on the thesis committee.

The author would also like to thank Mr. Ed Mahon and Mr. Joe Gibson in the department of Network Services for their kindness and support. Finally, he wishes to express a special thanks to his beloved family for their infinite love, patience and encouragement.

Abstract

This thesis deals with desktop videoconference systems and their related issues. Specifically, it attempts to resolve compatibility issues among various systems currently in use by investigating available communication standards and makes recommendations for future deployment. In order to carry out these goals, a review of the available literature and some experimentation was conducted.

The literature review describes different communications switching techniques, acquiring and processing images, and transport protocols over various types of physical networks. This work also investigates and experiments with some of the available videoconferencing systems that fit into a university campus environment. Through a structured approach to deployment, desktop videoconferencing offers improved collaboration among faculty, staff, and students.

Table of Contents

1	Introduction	1
1.1	Problem Statement	3
1.2	Motivation	4
1.3	Overview of the Thesis	5
2	Background	7
2.1	Cross Platform Operation	7
2.1.1	ITU-T Recommendation H.320	8
2.1.2	ITU-T Recommendation T.120	9
2.1.3	ITU-T Recommendation H.324	10
2.2	Circuit vs. Packet Switching Network	11
2.2.1	Circuit Switching	12
2.2.2	Packet Switching	16
2.2.3	Circuit or Packet Switching	19
2.2	Desktop Videoconferencing Products	20
3	Desktop Video Systems	25
3.1	Image Acquisition and Encoding	26
3.1.1	CCD Cameras	27
3.1.2	NTSC Encoding	29
3.2	Compression Standards	32
3.2.1	Audio Compression	32
3.2.2	JPEG Still-Image Coding Standard	35
3.2.3	MPEG Video Compression standard	39
3.2.4	ITU-T Recommendation H.261	45

4	Transport Protocols	4 9
4.1	Ethernet	49
4.1.1	Ethernet Standard	50
4.1.2	Frame Structure	53
4.1.3	Fast Ethernet	54
4.2	ATM	55
4.2.1	ATM Technology	56
4.2.2	Adaptation Layer Support	59
4.3	ISDN	61
4.3.1	ISDN Architecture and Standard	62
4.3.2	ISDN Protocol	63
4.3.3	Broadband-ISDN	66
4.4	ADSL	67
5	Cross Platform Issues	6 9
5.1	Multicast Backbone (MBone)	70
5.2	QuickTime Conferencing (QTC)	72
5.3	Intel ProShare	75
6	Field Trials and Results	7 8
6.1	First Topology	79
6.2	Second Topology	83
6.3	Third Topology	87
7	Conclusions and Recommendations	9 1
	References	9 3
	Vita	9 8

List of Tables

2.1	H.320 series of recommendations	8
2.2	T.120 series of recommendations	10
2.3	H.324 series of recommendations	11
2.4	Current and proposed TDM carrier standards	14
2.5	List of desktop videoconferencing products from various vendors	21
3.1	Layers and their function in MPEG-1	41
3.2	Parameters Bounds for MPEG-2 MP@ML	44
3.3	Header content of H.261 bit stream layer	46
4.1	Bit rate hierarchy of fiber optic channels	66
5.1	Video and audio codecs supported by QTC LAN and ISDN kits	73
6.1	Bandwidth measurements for video compressors with μ -law audio compressor	80
6.2	Bandwidth measurements for audio compressors with H.261 video compressor	82
6.3	Bandwidth measurements for audio compressors with no video transmission	82
6.4	Bandwidth measurements for video compressors with μ -law audio compressor	84
6.5	Bandwidth measurements for audio compressors with H.261 video compressor	85
6.6	Bandwidth measurements for audio compressors with no video transmission	86
6.7	Bandwidth measurements for video compressors with μ -law audio compressor	88
6.8	Bandwidth measurements for audio compressors with H.261 video compressor	89
6.9	Bandwidth measurements for audio compressors with no video transmission	89

List of Figures

3.1	Basic components of a desktop videoconferencing system	26
3.2	Schematics of CCD imagery	28
3.3	Typical NTSC encoder	30
3.4	JPEG video codec	36
3.5	Zig-zag scanning	38
3.6	Motion-Compensated Interpolation using Bidirectional prediction	41
3.7	Schematic block diagram of an MPEG-1 decoder	42
3.8	Alternate scan for MPEG-2 coding	44
3.9	Layered structure of H.261 bit stream	46
4.1	Ethernet frame structure	53
4.2	ATM cell format	58
4.3	ISDN interface between switch and terminal equipments	62
6.1	Schematic representation of the first topology	79
6.2	Plots of the average bandwidth of Table 6.1	81
6.3	Plots of the average bandwidth of Table 6.2	82
6.4	Plots of the average bandwidth of Table 6.3	83
6.5	Schematic representation of the second topology	84
6.6	Plots of the average bandwidth of Table 6.4	85
6.7	Plots of the average bandwidth of Table 6.5	85
6.8	Plots of the average bandwidth of Table 6.6	86
6.9	Schematic representation of the third topology	87
6.10	Plots of the average bandwidth of Table 6.7	88
6.11	Plots of the average bandwidth of Table 6.8	89
6.12	Plots of the average bandwidth of Table 6.9	90

Chapter 1

Introduction

Videoconferences are becoming increasingly frequent on both the Internet and over dedicated digital communication lines like ISDN, and are generating much research interest [1]. Availability of software tools enable real-time audio and video channels as well as whiteboards that allow two or more people to collaborate easily. The utilization of desktop computers and low cost hardware makes desktop videoconferencing a cost effective solution in comparison to room videoconferencing.

A few years ago, transmitting video across the Internet via twisted pair wire was considered impossible. However, with the advancement of video compression standards and the evolution of digital networks, real-time desktop videoconferencing is now available. The key point in transmitting video streams is bandwidth availability. MBone (Multicast Backbone), for example is a virtual network that provides a multicasting facility to the Internet, thus providing a bandwidth-efficient solution where one packet of video

can reach many workstations on a network. On the other hand, compression techniques can reach up to a 100:1 compression ratio where a video stream normally consists of from 10 to 40 Mbps (Mega bits per second) and can be transmitted, for instance, over several ISDN (Integrated Service Digital Network) channels with a bandwidth of 64 kbps (kilo bits per second) each [2,3].

Due to the fact that there are different computer platforms available in the market as well as different products for each platform, the market is evolving towards conformance to the emerging videoconferencing standards. These standards, such as H.320 for "Narrow-Band Visual Telephone Systems and Terminal Equipments" and T.120 for "Transmission Protocols for Multimedia Data", allow for interoperability among different products and platforms across different physical networks. By supporting H.320, Apple for instance, is assuring all QTC (QuickTime Conferencing)-savvy application users cross-platform communications between Macintoshes, PCs, UNIX and room-based conferencing systems [4].

Desktop videoconferencing applications built for use with today's networking infrastructure, including ISDN on Wide Area Networks (WANs) and Ethernet on Local Area Networks (LANs), will scale to an emerging network infrastructure more appropriate for real-time video streams, such as Fast Ethernet, ATM (Asynchronous Transfer Mode), and ADSL (Asymmetric Digital Subscriber Line), when they are more widely deployed.

Collaboration through videoconferencing can be grouped into two different categories. First, when a conference is being carried out between two people, the type of connection is termed *point-to-point*.

Collaboration is achieved through a bidirectional flow of information. Second, a *multipoint* connection is defined when a connection has been established between more than two people and they share data among all users. The physical network over which the connection is being made could be *circuit-switched* (dedicated path) or *packet-switched* (shared path). The subjects discussed so far will be presented in more detail in the following chapters.

1.1 Problem Statement

In the past, desktop media conferencing and networked systems relied on proprietary software. Solutions took a long time to develop and were incompatible between platforms. In many cases, costly solutions were hard to manage and run on networks, and users who might access these solutions found it difficult to share media objects with other connected computers, even on the same platform. Personal computer users continue to demand the ability to work with multimedia elements in a seamless manner over local and wide area networks.

This thesis project deals with desktop videoconferencing systems on the University of Tennessee campus. Specifically, it will attempt to resolve compatibility issues among the various systems currently in use and will make recommendations for future deployment. In order to carry out these goals, a review of the literature available and some experimentation will need to be conducted. Through a structured approach, desktop videoconferencing offers improved collaboration among faculty, staff, and students

1.2 Motivation

As the price of desktop video conference systems continue to fall, it is expected that more people will adopt their use for professional collaboration. Many manufactures have begun to include video conference systems in their computers. A case point is Apple's latest generation of Power Macintosh computers which come standard with video input/output ports and QuickTime Conference software already installed. On the other hand, the worldwide digital network is rapidly expanding, and the Internet is evolving to support real-time media applications through international standards for audio/video compression and communications protocols.

Collaboration using a desktop video conference system will occur both within the UTK campus network as well as between UTK and external Internet sites. This collaboration could be conducted over a *circuit-switched* network using a dedicated *point-to-point* ISDN link, or over the Internet with a *packet-switched* network via TCP/IP or X.25 standard which can be bandwidth intensive. Thought must be given as to how this new traffic will be handled. For example, fixed bandwidth can be assigned to each user with Simple Network Management Protocol (SNMP) tools by a network administrator. Also, other network elements such as routers need to be configured properly to support communication protocols such as H.320, MovieTalk, and MBone. Video compression/decompression (codec) standards such as H.261, JPEG, and MPEG also need to be reviewed.

1.3 Overview of the Thesis

The first stage of this project will be a review of the various desktop video systems and the issues related to them. This will include the issues of cross-platform operation, circuit vs. packet video, transport protocols, and bit rate video coding algorithms.

Once a basic knowledge of the desktop video conference systems related issues is obtained, the second stage will be to inventory the kinds of systems currently in use on the UTK campus and catalog their specifications. Efforts will be made to determine to what extent these systems can be compatible with one another. Based on this experience, recommendations to the user community for the future purchase of such systems will be written and disseminated.

Finally, infrastructure issues will be explored to address the increased traffic anticipated and to see to what extent packet and circuit switched video data can be made compatible.

This project will result in valuable information involving desktop video conferencing which offers improved collaboration between professionals. Through the evaluation of current systems, forecasts can be made for future acquisitions. This will ensure a carefully planned implementation and reduce the risk of future problems. This project is an attempt to ensure a maximal level of performance by implementing international standards so that what happened with proprietary groupware solutions is not repeated in the videoconferencing arena.

Enabling technologies for desktop videoconferencing are

discussed, including the compression of audio and video data and the transmission of this data over various communication channels. Features of currently available desktop videoconferencing systems and emerging interoperability standards are addressed. The following chapter will discuss cross platform operation and circuit versus packet switching, and present an overview of various products. Chapter 3 deals with desktop video systems, and reviews different image acquisition standards as well as various compression methods. In Chapter 4, transport protocols will be presented. Consideration will be given to Ethernet, ISDN, ATM, and ADSL. In Chapter 5, cross platform issues will be addressed, based on different products. Chapter 6 will present the testing results of the products presented in Chapter 5. Finally, the concluding chapter summarizes this research and makes recommendations to the user community.

Chapter 2

Background

This chapter will address the issue of cross-platform operation and interoperability among different platforms. Different standards have been adopted by different manufacturers. These standards will be discussed in detail as well as their influence on cross-platform operation. The issue of circuit-switched and packet-switched videoconferencing will also be addressed. Due to the commercial use of desktop videoconferencing, many products are available in the market from different vendors. The last section of this chapter will review various products and compare their specifications.

2.1 Cross-Platform Operation

There are three major platforms for desktop videoconferencing products. These are, UNIX-based workstations, Intel-based personal computers (PCs), and Macintosh from Apple Computer. Recently, several organizations pushed for the standardization of videoconferencing so that different products from different vendors

with different platforms would be interoperable. The ITU (International Telecommunication Union) is perhaps the most influential body which came up with several standards which are described as a series of recommendations.

2.1.1 ITU-T Recommendation H.320

ITU-T Recommendation H.320 is titled "Narrow-Band Visual Telephone Systems and Terminal Equipment." This recommendation was approved in March of 1993. Narrow-band bit rates are defined to range from 64 kbps to 1920 kbps (64 kbps $\times$ 30). This recommendation was designed primarily for ISDN. A typical ISDN Basic Rate Interface that might be available for desktop videoconferencing consists of two 64 kbps channels (p=2) or 128 kbps total bandwidth. H.320 is actually a series of recommendations. These recommendations and their titles are shown in Table 2.1.

Table 2.1: H.320 series of recommendations.

H.261	Video codec for audiovisual services at (px64) kbps
G.711	Pulse code modulation (PCM) of voice frequencies
G.722	7 KHz audio-coding within 64 kbps
G.728	Coding of speech at 16 kbps using low-delay code excited linear prediction
H.221	Frame structure for a 64 to 1920 kbps channel in audiovisual teleservices
H.230	Frame-synchronous control and indication signals for audiovisual systems
H.242	System for establishing communication between audiovisual terminals using digital channels up to 2 Mbps

An H.261 (discussed in Chapter 3) video codec requires bandwidth in multiples of 64 kbps. There are several bandwidths defined for audio. G.711 operates at 64 kbps which is one entire ISDN B channel. This leaves only 64 kbps for video and other data, assuming ISDN BRI. G.722 operates at 48/56/64 kbps. G.728 operates at 16 Kbps. G.711 and G.728 both support frequencies up to 3 kHz. G.722 supports frequencies up to 7 kHz. H.221, H.230 and H.242 specify how calls are negotiated, how the data are multiplexed and framed, etc.

2.1.2 ITU-T Recommendation T.120

ITU-T Recommendation T.120 is titled "Transmission Protocols for Multimedia Data." This recommendation defines multipoint transport of multimedia data. T.120 enables participants to share data during a conference. This data could be whiteboard data or another type of data such as a binary file.

T.120 is also actually a series of recommendations. These recommendations and their titles are shown in Table 2.2. T.123 defines connections for various types of networks (including ISDN, and LANs). T.122 and T.125 define multipoint communication. T.124 provides the mechanism to set up and manage conferences. T.127 defines file exchange. T.126 defines viewing and annotating shared images and application sharing. Some of these recommendations have been formally approved by the ITU-T while others are in the draft stage and expected to be approved soon.

Table 2.2: T.120 series of recommendations.

T.121	Generic application template
T.122	Multipoint communication service for audiographics conference applications
T.123	Protocol stacks for audiographic and audiovisual teleconference applications
T.124	Generic conference control
T.125	Multipoint communication service protocol specification
T.126	Multipoint still image and annotation protocol
T.127	Multipoint binary file transfer
T.128	Audiovisual control for multipoint multimedia systems

2.1.3 ITU-T Recommendation H.324

ITU-T Recommendation H.324 is titled "Multimedia Terminal for Low Bit Rate Visual Telephone Services over the GSTN." The H.324 series of recommendations (formerly called H.32P) defines real-time audio, video and data transfer over V.34 modems on the GSTN (Global Standard Telephone Network). Work on these documents is still in progress. H.324 recommendations and their titles are shown in Table 2.3.

A V.34 modem has a total bandwidth of 28.8 kbps. G.723 has two audio bandwidth modes: 5.3 kbps and 6.4 kbps. This leaves either 23.5 Kbps or 22.4 kbps for video and overhead. G.723 also has a silence suppression mode so that the audio bandwidth can be used for other data when no audio is being transmitted. H323, on the other hand, is a series of recommendations still under development which aim to standardize video conferencing over LANs.

Table 2.3: H.324 series of recommendations

H.263	Video coding at rates less than 64 kbps
G.723	Speech coder for multimedia telecommunications transmitting at 5.3/6.3 kbps
H.245	Multimedia system control
H.223	Multiplexing protocol for low bit rate multimedia terminals

2.2 Circuit vs. Packet Switching Network

Information including audio, video, and data can be transmitted through various network channels. The common types of networks are based on either circuit or packet switching. Circuit switching implies that a dedicated communication path has to be established between two end-points through one network, or in some cases, many different networks and kept open for the duration of the session. Bandwidth is allocated for the exclusive use by this session. In packet switching however, different users share the same channel. Data sent by one user can be transmitted over different channels. The data stream is divided into small pieces called packets. These packets are sent individually through a network, and depending on network conditions, may take different routes and arrive at their destination at different times and out-of-order. However, no dedicated bandwidth channel is set up prior to transmission as with circuit switching.

The choice of using *circuit-switched* or *packet-switched* networks depends on the type of application. A voice call, for instance, is very sensitive to delay so it is usually not practical to send it in packets over a network. On the other hand, it is very practical to send data in

packets over a network since a path is not in use all the time and data can tolerate delay if its transmission is successful. The next three sections discuss circuit and packet switching in more detail and their implementation in desktop videoconferencing.

2.2.1 Circuit Switching

Circuit-switched networks were originally designed to carry voice traffic, which is sensitive to ordering and to delay [5]. This traffic is carried through a dedicated path where transmission and resources remain dedicated to the call for its duration. The bandwidth is reserved for this specific call regardless of whether the call uses all of it or not. Switching in old networks relied on analog devices based on space-division multiplexing. Presently, most circuit-switched networks deploy digital switching based on FDM (frequency-division multiplexing) or TDM (time-division multiplexing) [6]. The most familiar circuit-switched networks are the global telephone systems. Within the telephone networks however, the gradual conversion from analog to digital transmission and switching technologies has enabled them to handle various types of traffic, including voice, data, and video.

Communication via circuit switching involves three phases. First, establishing a connection. This done by allocating a free channel through various nodes and requesting an end-to-end circuit connection. Routing a connection is based on availability of channels, as well as minimum delay and cost. Second, once a circuit has been established, traffic can be transmitted . Generally, the connection is full duplex, and data may be transmitted in both directions

simultaneously. Finally, to end a session, the connection can be terminated by either of the two users. Usually, the connection is *point-to-point* and it requires multi-conferencing units (MCU) to accommodate *multipoint* conferences. The establishment of such circuits takes place by means of signal multiplexing, routing, and controlling.

Multiplexing

The most common multiplexing method is carried out through FDM or synchronous TDM. As wide-area telecommunication networks evolve toward ISDN, synchronous TDM techniques are becoming dominant. ISDN and B-ISDN (discussed in Chapter 4) are based on multiplexing data through several channels [6,8].

The basis of the TDM hierarchy is DS-1 (Digital Signal). DS-1 is considered one frame with a duration of 125 μ sec and a length of 193 bits. The basic element of DS-1 is a DS-0 channel with a bandwidth of 64 kbps. DS-1 consists of 24 DS-0 channels in North America and Japan with a capacity of 1.544 Mbps, and 30 channels in Europe with a capacity of 2.048 Mbps [6,8].

Synchronization is very important in TDM since each channel per frame is designated to a different user. This is done by assigning a number for each channel to a corresponding user which determines its transmission time relative to the beginning of the frame. Previous systems utilized some of the channel capacity for control and signaling, while recent systems use a separate channel for this task. For example, channel 24 of each frame is used for control and signaling and the remaining 23 channels are used to transmit data.

To achieve bandwidth higher than 1.544 Mbps, several frames

could be interleaved together. For example, DS-1C is made up of 2 DS-1 channels with a bandwidth of 3.152 Mbps, and DS-2 consists of 4 DS-1 channels with a bandwidth of 6.312 Mbps. Notice that additional capacity is added to each higher level for framing and control bits. Table 2.1 lists the current and proposed hierarchy.

Table 2.4: Current and proposed TDM carrier standards.

DS Number	North America and Japan		Europe	
	No. of Channels	Data Rate Mbps	No. of Channels	Data Rate Mbps
DS-1	24	1.544	30	2.048
DS-1C	48	3.152	120	8.448
DS-2	96	6.312	480	34.368
DS-3	672	44.736	1920	139.264
DS-4	4032	274.176	7680	565.148

Routing

An important element of network management is called routing. It consists of the decision rules used to connect calls over the network. In a large circuit-switched network, many of the circuit connections will require a path through more than one switch. The network should make a decision on how and by what means a connection can be established. By large, cost and reliability determine the overall operation of a particular network. With the availability of high-bandwidth fiber trunks and high-capacity digital switches capable of handling tens of thousands of trunks, networks can deliver reliable service at minimal cost. Several methods were adopted for routing throughout the world; however, the discussion will be limited to some of them. The reader can refer to [8-10] for

more comprehensive details.

Early networks relied on a very simple form of routing known as *Direct Routing* (DR). With DR, connection establishment is based on the rule that subscribers of the same end office are connected via this end office. The search continues up the hierarchy until subscribers of different end offices that are connected to the same regional center, are connected via this regional center. To solve the problem resulting from overload during peak hours, where a tremendous amount of traffic must be carried by the upper level switches, the network is augmented with additional switching nodes called tandem switches which interconnect adjacent end offices. This type of routing is known *Alternate Hierarchical Routing* (AHR). With this connection, traffic is always routed through the lowest available level of the network.

Dynamic routing is another and more flexible form adopted recently. In the mid 1980s, AT&T introduced a new method in the routing of long distance network. This method is known as *Dynamic Nonhierarchical Routing* (DNHR) [11]. With DNHR, there is no longer any hierarchical relation between switches. All switches in DNHR network are of the same type and can equally receive calls from or serve as a tandem switch for alternate routing. DNHR changes dynamically according to an off-line precalculated traffic forecast which is stored in order to drive switches in the way they have been designed. The advantage stems from the fact that alternate paths are limited to two links at most. Another type of dynamic routing is *Dynamic Alternative Routing* (DAR). Currently employed by British Telecom [5], DAR is an adaptive call-routing strategy that stochastically selects an alternative path, instead of preassigned ones,

when a direct path is not available.

Control Signaling

In a circuit-switched network, control signals are the means by which the network is managed and calls are established, maintained, and terminated. There are two types of control signaling. *In-channel* signaling, which uses the same channel to carry both control signals and calls and *common channel* signaling, which uses separate channels for control and calls. *In-channel* signaling can be employed in two forms. First, *in-band* signaling uses the same frequency band as the call signals. The second form is *out-of-band* signaling which uses a separate narrow signaling band within the channel. As will be shown in Chapter 4, ISDN technology employs *common channel* signaling where a separate channel is dedicated for control.

2.2.2 Packet Switching

The concept of packet switching originated as a distributed switching system for military communications [12]. The first important application was the development of the ARPANET to link together time-shared research computers in a nationwide network. Unlike circuit switching, packet switching was designed particularly for data communications rather than voice communications.

Packet switching, which operates at the network layer of the OSI (Open System Interconnection) hierarchy, is carried through various transport protocols. Packet switching is based upon dividing the data stream into small packets and transporting them through the network. Notice that the meaning of data here includes audio and

video streams. Basically, a packet is formed by concatenating the data with a header and an end-of-message flag. The header contains all the information necessary for routing the data through the network, such as source address, destination address, packet ID number, and error correction bits. The packet is stored in a buffer at each switch which determines the next node in the route based upon the information contained in the header. When the appropriate link is available, the packet is forwarded to the buffer in the next switch. Usually, an acknowledgment is returned by the receiving node if the message is received without error; otherwise, the sender retransmits the message. Since shorter packets requires less storage but more of them for a given amount of data, the exact length of a packet is a trade-off between delay and overhead. The two methods for routing packets through networks are known as *virtual circuit* (VC) and *datagram* . VC is a connection-oriented method while datagram is a connectionless service.

Virtual Circuit

The name *virtual circuit* comes from its analogy to a telephone circuit. A path, which consists of a logical connection, is first established between two users. During this establishment phase, the two users agree to set up a connection between them and decide on the quality of service that is to be associated with the connection. During data transmission, all packets follow the same route (i.e., same VC) which has been established. The VC method also provides other services such as error control and flow control. The error control ensures both that the packets are received in the correct order and that all packets arrive correctly. If a packet fails to reach a node or it

arrives incorrectly, the receiving node can request a retransmission of the packet from the transmitting node. The flow control may be used for communication between a slow receiver and a faster transmitter. At the network level, if the receiving node finds that there is no more space in the buffer for the incoming packets, it can request the transmitting node to stop sending data until further notice. It is important to note that the connection that was established is not a dedicated path between nodes; however, it can be shared by many other virtual connections. The packets are usually buffered at intermediate nodes and queued for output over a line which is part of the virtual circuit.

Datagram

In the *datagram* method, packets are routed independently and with no reference to previously transmitted packets. Packets travel any available route to the destination and arrive out-of-order. The network layer simply accepts messages from the transport layer and sends out each one as an isolated unit. In contrast to the VC method, sequentially transmitted datagrams can follow completely different paths and arrive in a different sequence.

The protocols which are responsible for delivering packets with a VC or datagram are X.25 at the network level and TCP at the transport level. X.25 is a virtual-circuit protocol which operates at the network level and specifies an interface which includes packet format and control procedures between a host system known as DTE (Data Terminal Equipments) and a packet-switched network known as DCE (Data Communication Equipments) [7,13]. It is the means by which a VC is established and removed, and error and flow control is

carried out. X.25 also supports datagram connections. The TCP/IP (Transmission Control Protocol/Internet Protocol) protocols, developed by the Defense Advanced Research Projects Agency (DARPA), are designed to operate over a potentially noisy, low-quality transmission line [14]. The responsibility of TCP is to assemble packets for transmission and to put them back in order upon receipt. The basic operation of TCP is to accept long messages from a user, divides them into small packets, and send each packet as a separate datagram. In addition to providing a reliable and flow-controlled delivery of data, TCP also provides commands for initiating a connection, accepting an incoming connection, sending and receiving data, and closing a connection.

2.2.3 Circuit or Packet Switching

Circuit switching seems to be more attractive for desktop videoconferencing since it allocates more bandwidth for each user and information can be sent with virtually no delay. Although packet switching requires more processing time, its efficient utilization of bandwidth and its ability to accomodate *multipoint* connections encouraged researchers to modify it in order to reduce delays. As a consequence, a new packet-switched network is emerging. *Fast packet switching* which includes B-ISDN and ATM (discussed in Chapter 4) is an advanced version of packet switching based on fiber optic links and simplified protocols [6,7,15]. This new technology meets the demand of high bit rate for image and video services and reduces their transmission delays.

2.3 Desktop Videoconferencing Products

Many products are available in the market for commercial as well as educational use. Several vendors such as Intel, Apple, Silicon Graphics, Sun Microsystems, and many others are actively advertising their products. Most of these companies have worked or are working on their products to support interoperability of videoconferencing over LANs and ISDN lines. Some other products are provided free of charge such as CU-SeeMe from Cornell University. As a result, there are some 50 various products available from all over the world. The cost of many of them ranges from \$300 (over LANs including camera) up to \$6,000 for a full featured product. ITU-T recommendations H.320 and T.120 for videoconferencing over ISDN is already supported by many of them. Recommendation H.324 for videoconferencing over GSTN using V.34 modems will also be supported soon and could make desktop videoconferencing cheaper and more attractive for many users. Below, in Table 2.5, is a list of several products which, from the author's point of view, are the most competitive solutions for this task based on their features. This list was taken from [16].

Table 2.5: List of desktop videoconferencing products from various vendors.

Product Name	Being There	CLI's Desktop	Face 2 Face
Version	2	1000/2000	1
Provider	Intelilgence at Large	Compression Labs, Inc.	Electronic Studio
Description	LAN (Ethernet,Token Ring) WAN (analog dial-up, ISDN, T1/E1)	Audio/Video/Collab. tools over ISDN	Video/Tools: ISDN, analog, Ethernet. Audio: ISDN, analog
Platform	Mac	PC	Mac
Requirments	Mac OS 7.1 or later 8 MB RAM, 3.5 MB disk	Windows 3.1 Windows 95	Apple Com. Toolbox, video capture card
LAN Protocol	Appletalk, TCP/IP	N/A	AppleTalk
Audio Encoding	QuikTime	G.711, G.722, G.728	N/A
Video Encoding	QuikTime Video Codec	H.261	N/A
Interoperability	No	H.320	N/A
Multipoint	4 A/V, 10 collaboration, ISDN point-to-point only	Multipoint bridge	No
Collaboration Features	Whiteboard, file transfer, document sharing, drag and drop interface	Shared whiteboard, file transfer	Text and image exchange

Table 2.5: (Continued).

Product Name	Communique!	InPerson	CU-SeeMe (Enhanced)
Version	N/A	2	2.0 (PC), 2.0 (Mac)
Provider	InSoft, Inc.	Silicon Graphics	White Pine Software
Description	LAN/WAN real-time digital audio, video, chat, text, and graphic tools	Audio/Video/Tools over ISDN, T1, Ethernet and FDDI	Over Internet or any TCP/IP network. Software only solution
Platform	Sun, HP, IBM, DEC, SGI, PC	SGI, Sun, HP, PC	Mac, PC
Requirments	Video card and camera	IRIX 5.3 system software	Video camera. supports all Mac, PC platforms
LAN Protocol	Packet/Cell/Tel. networks	UDP/IP, TCP/IP	TCP/IP
Audio Encoding	G.7xx	G.11, GSM	Non-standard
Video Encoding	Cellb, JPEG, H.261, DVE	H.261, RGB8, HDCC	White Pine 24-bit
Interoperability	H.320	N/A	No
Multipoint	Yes	Yes	Yes, using White Pine reflector
Collaboration Features	Video/Audio/real time collaborative tools	Text, image, 3D model sharin. "Shared self" for file transfer	Object oriented text draw objects. copy and filtering
Notes		Optional G.28 audio compression	As Internet phone with 14.4 modem

Table 2.5: (Continued).

Product Name	nv (Network Video)	PictureTel Live PCS 100	ProShare Video System 200
Version	3.3 (beta)		2
Provider	Xerox/PARC	PictureTel	Intel
Description	Unicast and Multicast over the Internet. Supplemented vat & wb	Video/Audio/Tools over ISDN, Switched 56	Video/Audio/Tools over LAN/WAN/ISDN
Platform	Sun SPARC, DEC 5000, SGI HP9000, IBM RS6000	PC	PC
Requirments	X display and camera compatible with hardware	PC386 or higher, 2 ISA slots, Win. 3.1+	ISDN: NT-1 adapter I. Network interface card
LAN Protocol	UDP/IP, IP Multicast	N/A	NetBIOS, TCP/IP, IPX
Audio Encoding	N/A	G.xx, PT 724	GSM, G.711,G.728
Video Encoding	Native NV, CellB	H.261	Indeo, QCIF H.261
Interoperability	N/A	H.320	H.320, T.120
Multipoint	Yes	Yes (16 maximum)	Yes, using H320- compatible bridges
Collaboration Features	None	Whiteboard, file transfer, screen and application sharing	Whiteboard, application sharring, file transfer
Notes	nv, vat, and wb are	Different versions	Echo canceller is

Table 2.5: (Continued).

Product Name	QuickTime Conferencing	ShowMe	VISIT
Version	1.4	2.0.1	2
Provider	Apple Computer, Inc.	Sun Microsystems	Northern Telecom, Inc.
Description	Video/Audio/Tools over the Internet and ISDN	Video/Audio/Tools over the Internet	Video/Tools over ISDN or switched 56. Audio: separate line
Platform	Mac, PC (future)	Sun SPARCstation	PC, Mac
Requirments	PowerMac AV or later. NuBus-based Mac for ISDN	Solaris 2.3 or later. Sun video board	PC 386 or higher Mac: NuBus-equipped
LAN Protocol	AppleTalk, TCP/IP	UDP/IP, TCP/IP, RTP, IP Multicast	N/A
Audio Encoding	MACE, ISDN kit: G.7xx	G.711	N/A
Video Encoding	H.261, JPEG, Apple video	CellB	H.261
Interoperability	ISDN kit: H.320 and T.120	No	H.320 (future version)
Multipoint	Yes (9 users) ISDN kit: No	Yes, with and without IP Multicast	No
Collaboration Features	Graphics, text, sound, QT movie, Capture screen. Saved shared whiteboard	Whiteboard, application sharing	Whiteboard, file transfer
Notes	Bandwidth monitoring and control via SNMP		

Chapter 3

Desktop Video Systems

A considerable effort has been underway for some time to develop inexpensive transmission techniques that take advantage of recent advances in electronic technology as well as expected future developments [17]. Most of the attention has been focused on digital systems because of their lower sensitivity to noise than analog systems; in addition, signal processing is much easier in a digital format. Progress is being made in the transmission and processing of digital signals to cover video which demands higher bit rates than audio and data. The first improvement, namely transmission media which will be discussed in the next chapter, is achieved through new technology with higher bit rate capacity. The second improvement, which will be the scope of this chapter, is through acquiring and processing data in a way to reduce the amount of information to be transmitted. Figure 3.1 shows a block diagram illustrating the basic elements of a desktop videoconferencing system. Bit rate reduction is accomplished by eliminating, as much as possible, the substantial amount of redundant information that exists in a video signal. This

task is done through video compression where various methods and standardized techniques are available. Some of these standards, such as JPEG, MPEG, and H.261 will be discussed after addressing the issue of acquiring images (still and motion).

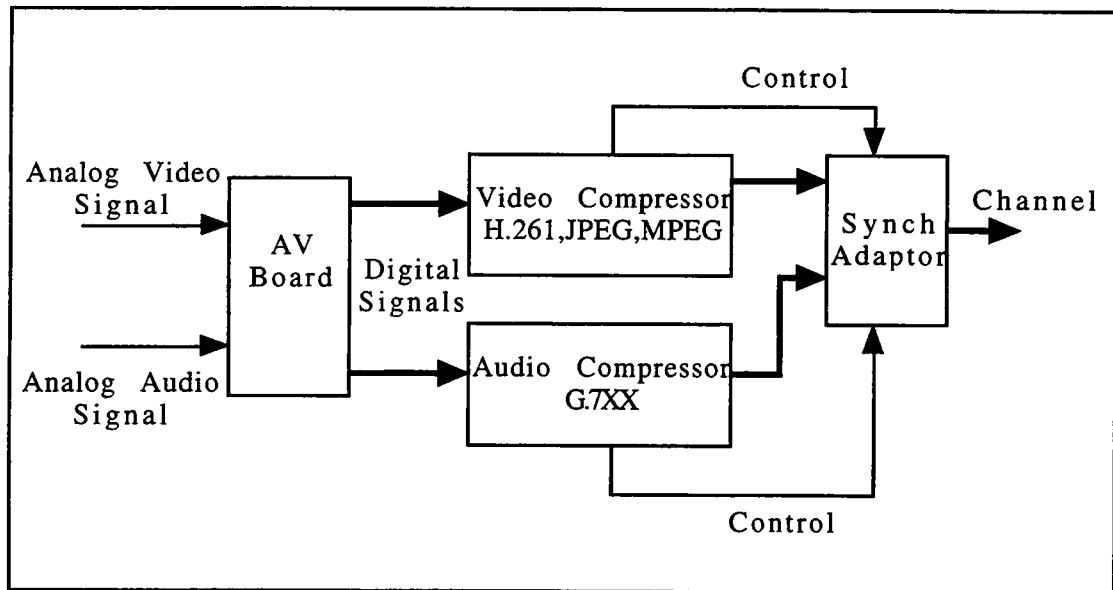


Figure 3.1: Basic components of a desktop videoconferencing system.

3.1 Image Acquisition and Encoding

Imaging and video display have gone through a long evolution in the past half century starting with black and white analog video signals, then color video, and lately to digital color systems. Analog electronic imaging employs a raster scan to acquire an analog signal such as a television picture [18,19]. A frame (or picture) is divided into line segments of amplitude-modulated electrical signals, and these line are stacked sequentially to form a two-dimensional picture. On the other hand, digital imaging technology divides the

image up into a two dimensional array of discrete pixels. Within each pixel (picture element), the signal is represented by digital words for red, green and blue intensities.

The analog video standard in North America and Japan is known as NTSC (National Television System Committee) [20-22] which is to be replaced soon by HDTV (High-Definition TV). A major advance in solid-state imagery came in 1970 with the invention of the CCD (Charge-coupled device [23]. The CCD, a silicon technology, is the most recent advancement in capturing digital images. Desktop videoconferencing tools support one or both, NTSC and CCD for video capture and transmission. The following sections will present a general discussion about these two subjects.

3.1.1 CCD Cameras

A charge-coupled device (CCD) is an array of MOS (metal-oxide-semiconductor) capacitors designed to translate an incident pattern of photons into discrete analog signals ready for digitization [19]. When the MOS capacitors are properly clocked, they move stored charges along from one capacitor to the other in a controlled manner. At the output of the device, the signal charge is converted to a voltage and digitized.

There are several architectures of CCDs. Figure 3.2 shows two types: (a) a linear CCD scanner consists of a single row of about 2,000 to 5,000 photodiodes and CCD shift register for readout. This type of CCD is used mainly in facsimile scanners; (b) a frame-transfer imager where part of the surface is covered with photodetectors and the rest

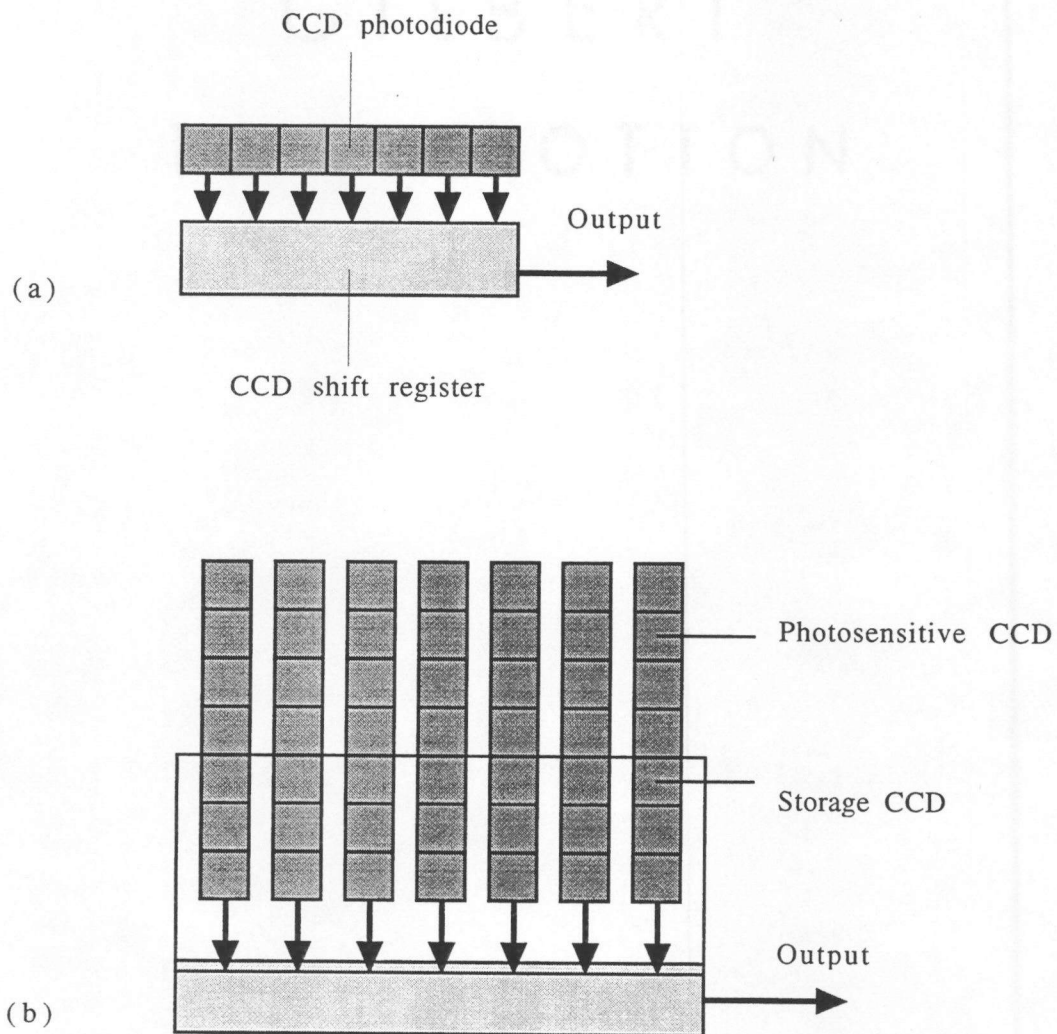


Fig. 3.2: Schematics of CCD imagery. (a): Linear CCD scanner with single row of photodiodes and a CCD shift register. (b): Two-dimensional CCD array with two parts, for sensing and storing.

with storage elements. This type of CCD is used for motion pictures and, as a result, desktop videoconferencing products often include a CCD camera with their packages due to its small size.

In capturing color video, the image sensor can be clocked to produce a frame rate of 30 fps where each frame consists of the original number of pixels of the CCD sensor. An array of red, green, and blue filters is incorporated in the sensor to produce the luminance and chrominance components of an image similar to that of NTSC or others.

3.1.2 NTSC Encoding

The video stream is a composite of many images stacked sequentially and played back at the same rate to the human eye. This is how a stack of images played at about 30 fps (for NTSC) or 25 fps (for PAL) give the illusion of full motion picture. PAL (Phase Alteration Line) is another video standard used mostly in Europe. Human vision is based on three colors: red, green, and blue. But since the human eye can perceive abrupt transitions in brightness much more readily than in hue [21], NTSC first transforms the signal from RGB into a color space whose axes align with the visual ranges, and then apply quadrature modulation to produce a color signal called *chroma*. A typical NTSC encoder is shown in Figure 3.3 [20]. The transformation from RGB space to YIQ space is described by Equation 3.1, where Y represents the brightness of the image and is called *Luminance*. The I and Q axes are simply a scaling and rotation of the color differences $R-Y$ and $B-Y$, respectively, by 33 degrees.

$$Y = 0.299R + 0.587G + 0.114B \quad (3.1a)$$

$$I = 0.596R - 0.275G - 0.321B \quad (3.1b)$$

$$Q = 0.212R - 0.523G + 0.311B \quad (3.1c)$$

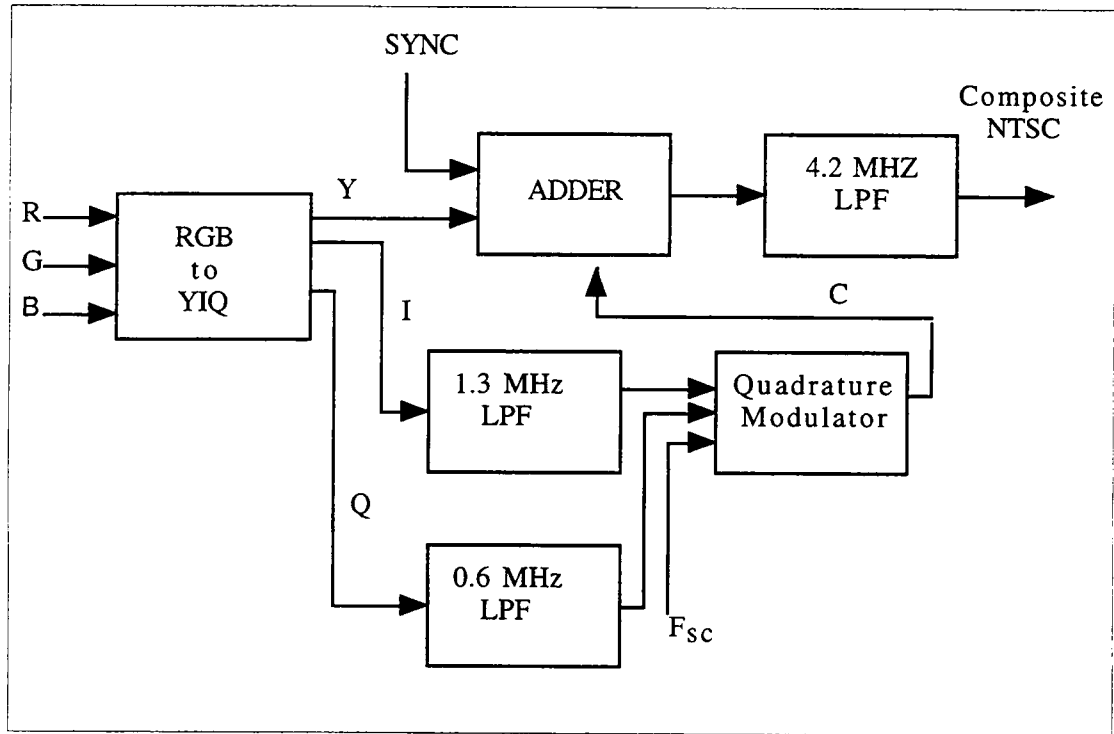


Figure 3.3: Typical NTSC encoder.

Notice that the RGB signal is a normalized version of the original nonlinear one. The signals I and Q are quadrature amplitude-modulated to form one signal called *chroma* (Equation 3.2) with a subcarrier frequency $F_{sc} = 3.579545$ MHz [20, 21].

$$C = I \cos (F_{sc}t) + Q \sin (F_{sc}t) \quad (3.2)$$

The chroma signal, C , has two components: an amplitude, $(I^2 + Q^2)^{1/2}$, called the *saturation* of the color, and phase, $\tan^{-1}(Q/I)$, called the *hue*.

In NTSC format, there are 525 lines per frame which is divided into 2 fields of 262.5 lines each. The 2 fields in each frame are interlaced to produce a field rate of 59.94 fields/sec. Consecutive fields contain odd and even lines of the frame, resulting in 29.97 fps. However, not all 525 lines are visible [17]. Only 483 lines contain video information and the rest are for control-signals.

An improved form of NTSC is the component digital standard called CCIR 601 [22, 24]. A scan line consists of 122 control pixels, 720 visible pixels, and 16 more blank pixels. To represent video in pixels, luminance and chrominance values have to be recalculated and encoded according to Equations 3.3 - 3.5. This is done in order to keep the values of C_R and C_B

$$Y_b = 219Y + 16 \quad (3.3)$$

$$C_B = 112/0.886(B - Y) + 128 \quad (3.4)$$

$$C_R = 112/0.701(R - Y) + 128 \quad (3.5)$$

in the range 0-1 volt and to quantize them into 256 quantization levels. In order to provide some margin for coding, filtering, etc., the full digital video range is not used for the initial digitization of the component signals. The digital luminance, as given by Eq. 3.3, is rounded to the nearest integer which has a range of 220 levels starting from 16. Similarly, the digital chrominance (i.e., $R-Y$ and $B-Y$)

Y), as given by Eqs. 3.4 and 3.5, are also rounded to the nearest integer which has a color-difference range of 225 levels starting from 16. The resulting bit-rate of all three components (Y_d, C_B, C_R) is about 216 Mbps with 486 active lines per frame.

3.2 Compression Standards

Audio and video compression are essential operations in desktop videoconferencing and all other forms of audio and video transmission and storage. Although audio compression can achieve only up to 8:1 compression ratio, a digitized video bit rate of approximately 45 Mbps can be compressed to a rate lower than 64 kbps [2]. To ensure interoperability among various vendors, several international organizations are working on the standardization of audio and video compression. ITU published several recommendations dealing with audio compression standards such as G.711, G.722, G.723, and G.728, as well as several video compression standards jointly with ISO. JPEG, MPEG, and H.261 are some forms of image and video compression standards which will be discussed.

3.2.1 Audio Compression

ITU audio standards (G.711, G.722, G.723, G.728) are based on some compression techniques applied to videoconferencing. These techniques can achieve real-time compression and decompression and can be implemented in software or hardware.

Based on Pulse Code Modulation (PCM), the quantizer step spacing can be transformed from uniform into logarithmic which allows a

larger range of values to be covered with the same number of bits. There are two commonly used transformation: μ -law format which is common in North America and Japan for digital telephony with ISDN, and A -law in Europe. the G.711 standard specified by recommendation H.320 employs these techniques at a bit rate of 64 kbps with 8 kHz sampling rate and 8 bits per sample. μ - and A -law transformations (Eqs 3.6 and 3.7, respectively) can achieve a compression of 1.75:1.

$$y = y_{\max} \frac{\log_e [1 + \mu(|x|/x_{\max})]}{\log_e (1 + \mu)} \operatorname{sgn} x \quad (3.6)$$

$$y = \begin{cases} y_{\max} \frac{A(|x|/x_{\max})}{1 + \log_e A} \operatorname{sgn} x & 0 < \frac{|x|}{x_{\max}} < \frac{1}{A} \\ y_{\max} \frac{1 + \log_e [A(|x|/x_{\max})]}{1 + \log_e A} \operatorname{sgn} x & \frac{1}{A} < \frac{|x|}{x_{\max}} < 1 \end{cases} \quad (3.7)$$

The PCM method encodes each audio sample independently from adjacent samples. However, Adaptive Differential Pulse Code Modulation (ADPCM) can achieve higher compression ratios by encoding the difference between each sample and its predicted value which requires fewer bits per sample to encode the same information. One of the standards specified by H.320 that uses this method is G.722 . The G.722 standard samples at a rate of 16 kHz with 14 bits per sample at three different bit rates: 48, 56, and 64 kbps as described in Chapter 2. This standard achieves some improvement in compression ratio (2:1), but the main improvement

is by increasing the audio bandwidth to 7 kHz which is useful for more than just speech signals with a bandwidth of 3 kHz such as the case with G.711 or G.728.

Since the difference between adjacent time samples for speech is small, sampling techniques have evolved to transmitting sample-to-sample differences rather than the actual values. The number of bits required to move data through the channel with a given fidelity is related to the signal variance. Hence the reduced variance sequence of residues can be moved through the channel with a reduced data rate [28]. A one-tap predictor DPCM can be used for ADPCM by incorporating slow rate, data-dependent adjustment algorithms

Some encoding methods are designed specifically for speech signals which can achieve higher compression ratios. Two of these methods are Linear Predictive Coding (LPC) and Code Excited Linear Prediction (CELP). LPC is the natural extension of N-tap predictive coders where it fits speech signals to a simple vocal tract. The best fit parameters are transmitted and used by the decoder to generate synthetic speech that is similar to the original. In addition to vocal tract modeling as in LPC, CELP computes the error between the input speech data and the model and transmits a representation of the error in addition to the model parameters. Errors are represented as indices in a common code table shared between encoders and decoders. ITU-T recommendation G.728 employs this form of coding for audio compression which requires a bandwidth of 16 kbps.

3.2.2 JPEG Still-Image Coding Standard

The Joint Photographic Expert Group (JPEG) was formed in 1986 by the ISO and CCITT (now known as ITU-T) to develop a general purpose still-image compression standard. From 1988 to 1990, the JPEG committee refined several methods incorporating the discrete cosine transform (DCT) for lossy compression. The committee's work was published in two parts: "Part 1: Requirements and Guidelines" [25], and "Part 2: Compliance Testing" [26]. JPEG standards support four modes of operation:

- *Sequential encoding*: In this mode, 8×8 blocks of an input image are formatted for compression by scanning the image left to right and top to bottom. Each block of samples is transformed to a block of coefficients and then quantized and entropy-encoded.
- *Progressive encoding*: The image is encoded in multiple scans for applications with a low bandwidth medium. The method is similar to sequential encoding, but the quantized coefficients are partially encoded in multiple scans.
- *Lossless encoding*: The decoder reproduces the exact encoded image. The neighboring entropy-decoded samples are used to predict the image.
- *Hierarchical encoding*: The image is encoded at multiple resolutions, so that lower-resolution versions may be accessed without a full resolution decompression.

Videoconferencing employs sequential encoding and therefore discussion will be limited to this mode. The sequential DCT-based mode offers very good compression ratios, while maintaining

excellent image quality [24]. A subset of the sequential DCT capabilities has been identified by JPEG as a *Baseline System* (discussed later). DCT-based JPEG implementations are required to include baseline capability in order to ensure interoperability between codecs from different vendors. Figure 3.4 [4] shows the elements of a typical sequential DCT-based codec.

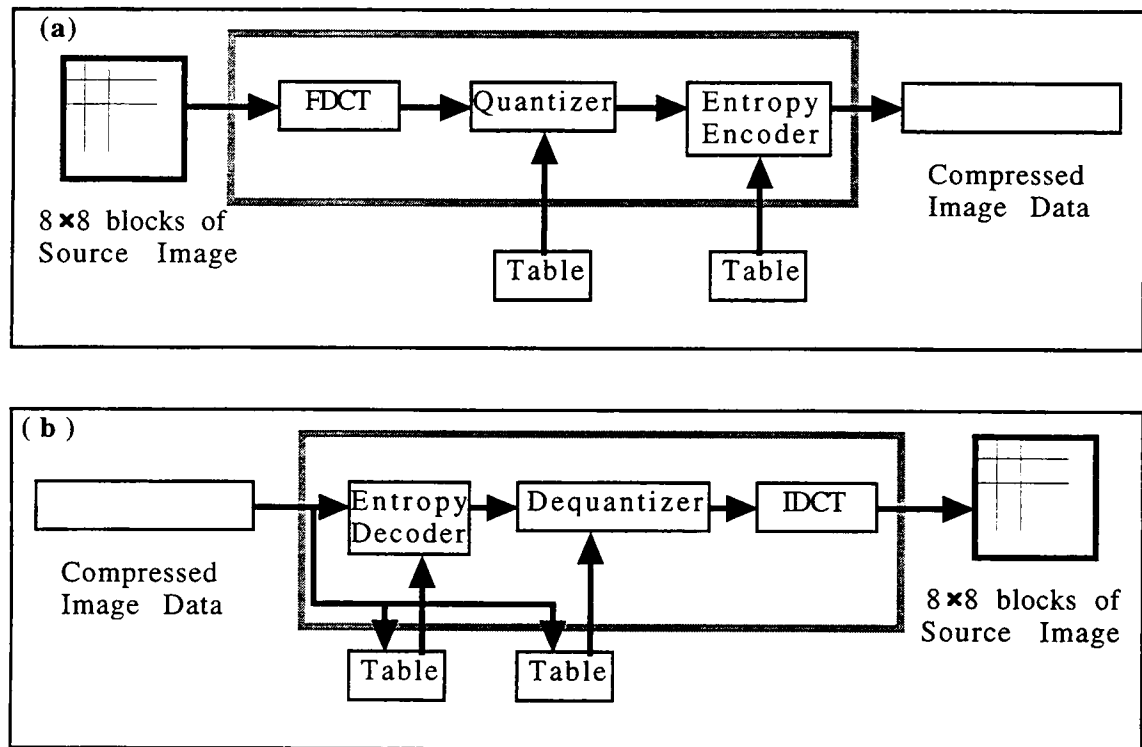


Figure 3.4: JPEG video codec: (a) encoder, (b) decoder.

The coding process begins by grouping the input image into non-overlapping 8×8 blocks of component samples, shifted to signed integers and input to a forward DCT (FDCT). This discussion assumes the encoding of a single component image (grayscale). For multiple component color images, the components are interleaved. At the

output of the decoder, the inverse DCT (IDCT) outputs 8×8 sample blocks to form the reconstructed image. Equations 3.8 and 3.9 are the mathematical definitions of the 8×8 FDCT and IDCT, respectively. The output of the FDCT, consists of 64 DCT coefficients and is uniformly quantized according to a prespecified 64-element quantization table. The purpose of quantization is to achieve further compression by representing DCT coefficients with no greater precision than is necessary and to discard information which is not visually significant [4, 24]. Many JPEG implementations control the compression ratio by using a q-factor, which is usually a scale factor applied to the quantization tables.

$$\text{FDCT} : C_{uv} = \frac{1}{4} K_u K_v \sum_{u=0}^7 \sum_{v=0}^7 b_{xy} \cos \frac{\pi u(2x+1)}{16} \cos \frac{\pi v(2y+1)}{16} \quad (3.8)$$

$$\text{IDCT} : b_{xy} = \frac{1}{4} \sum_{u=0}^7 \sum_{v=0}^7 K_u K_v C_{uv} \cos \frac{\pi u(2x+1)}{16} \cos \frac{\pi v(2y+1)}{16} \quad (3.9)$$

where $K_w = 0.707$ for $w = 0$; $K_w = 1$ otherwise

After quantization, a bit stream is formed from the block. The so-called DC coefficient, the first coefficient of each block which contains the average value of the other 63 so-called AC coefficients, is encoded as the difference between the current DC coefficient and the DC coefficient of the previous block. The AC coefficients usually contain many zeros and are reordered in a zig-zag scan as shown in Figure 3.5 so as to have long runs of zeros. The zero run (number of zero elements) is entropy-encoded using Huffman codes (or

arithmetic codes) where each number of consecutive zeros is assigned a variable length code (VLC) based on statistical characteristics.

Baseline sequential coding is a special case of sequential coding for images with 8-bit samples and uses Huffman coding only. The description of baseline sequential entropy coding is given in two steps: conversion of the quantized DCT coefficients into an intermediate sequence of symbols and assignment of variable length codes to the symbols. An intermediate symbol (runlength/nonzero coefficient combination) contains three types of information: runlength, size, and amplitude. Runlength is the number of consecutive zero-valued AC coefficients in the zig-zag sequence preceding the nonzero AC coefficient being represented. Size is the number of bits used to encode the amplitude which is a signed integer in the range $[-1023, 1023]$. If the end of the block is reached, then a special symbol EOB is inserted. For both DC and AC coefficients, the combination runlength/size is encoded with a VLC, and the amplitude is encoded with a variable-length integer (VLI). These codes are stored as tables in the codec.

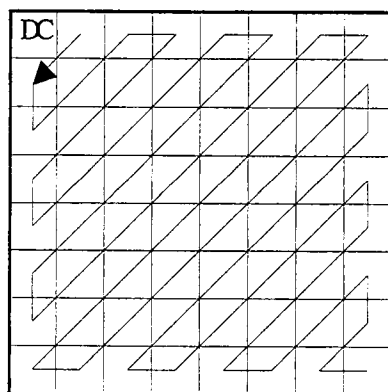


Figure 3.5: Zig-zag scanning.

3.2.2 MPEG Video Compression Standard

The Moving Picture Expert Group (MPEG) was chartered by the ISO to standardize a coded representation of video and associated audio suitable for digital storage and transmission media [2, 24]. The group's goal has been to develop a *generic* standard which can be used by various applications such as digital storage, teleconferencing, videophone, and many others. The MPEG first phase (MPEG-1) video compression standard aimed for video and audio coding at a bit rates up to 1.5 Mbps. The second phase (MPEG-2), however, is aimed at video bit rates above 2 Mbps. The official names of Phases 1 and 2 are ISO/IEC 11172 and 13818, respectively [24].

MPEG Phase 1

The MPEG-1 standard for video compression employs two major techniques to achieve low bit rates: block-based DCT for the reduction of spatial (intraframe) redundancy and block-based motion compensation for the reduction of temporal (interframe) redundancy. *Intraframe Coding*: As in JPEG, intraframe coding exploits the spatial redundancy found in each picture of video. The picture is individually and independently compressed through the same DCT-based techniques applied to JPEG image. Pictures coded using only *intraframe* coding are called *I-Pictures*.

Interframe Coding: A high degree of correlation between adjacent pictures results in temporal redundancy which can be exploited by an interframe difference signal called *prediction error*. Block-based motion compensation is computed on a basis of a 16×16 macroblock (MB). Prediction error can be computed in two ways: (1)

Unidirectional Motion Estimation based on Forward Prediction. An MB in the picture to be encoded is matched with another MB from the previous picture. The prediction error is computed as the difference between these two blocks. Pictures coded using *Forward Prediction* are called *P-Pictures*. (2) *Bidirectional Temporal Prediction* which is also called *Motion Compensated Interpolation*, is a key feature of MPEG video [24]. This type of coding uses two reference pictures, one in the past and one in the future. It uses *Forward Prediction*, *Backward Prediction*, or an average of the two which is known as *Interpolation*. Bidirectional prediction, shown in Figure 3.6, can produce higher compression ratios but results in more delay since more complex computations have to be done. Pictures coded using this technique are called *B-Pictures*. MPEG-1 does not specify the encoder, but any encoder design is expected to follow a specified *syntax* of the compressed bit stream. The syntax is flexible and is constructed in a hierarchy of several *layers*, each performing a different logical function. In addition, the bit stream must satisfy a particular constraint so that it can be decodable with a buffer, known as Video Buffer Verifier (VBV), of an appropriate size. Decoding should occur without buffer underflow or overflow [3,17]. The syntax of a bit stream contains six layers, Table 3.1 lists them with their function. The encoding process follows the following procedures: Reordering frames so that I- and P-Pictures are coded and transmitted first so that to be used as a reference for B-Pictures. For each MB, the inter/intra classifier then compares the input pixels with the prediction error output to determine if inter- or intra-operation should take place. The future and past pictures to be used

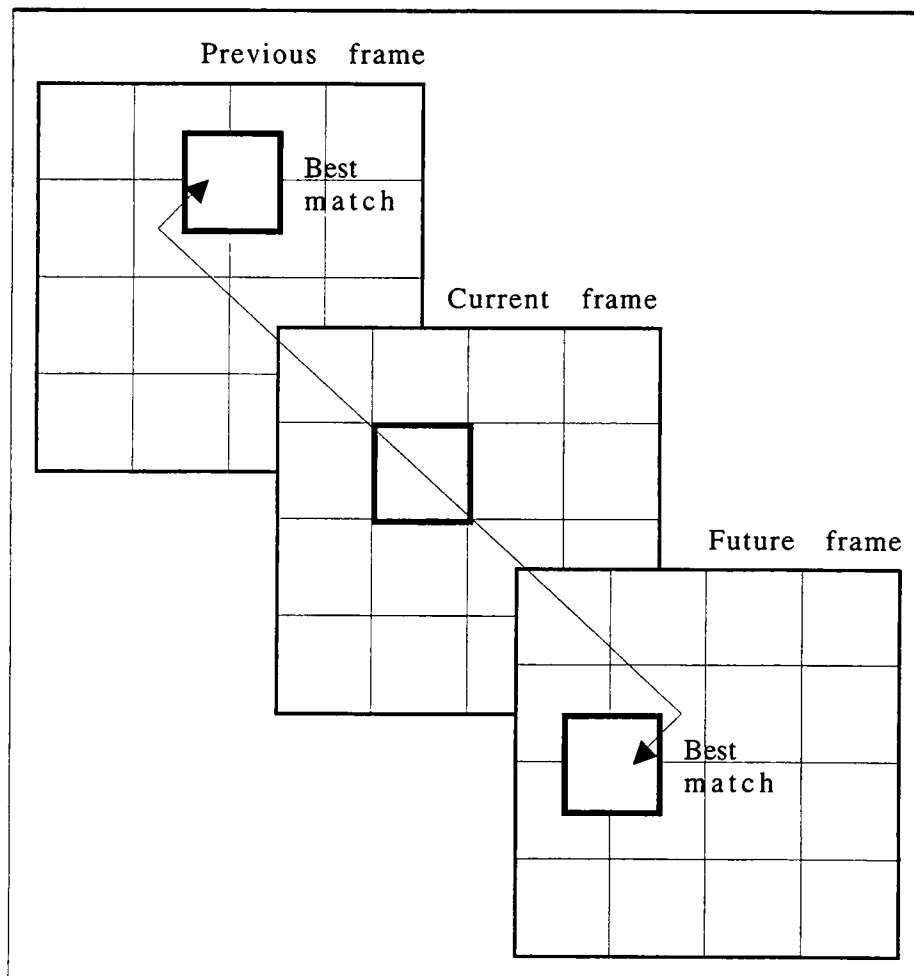


Figure 3.6: Motion-Compensated Interpolation using Bidirectional prediction.

Table 3.1: Layers and their function in MPEG-1.

Syntax Layer	Function
Sequence	Context unit
Group of pictures	Random access unit
Picture	Primary coding unit
Slice	Resynchronization unit
Macroblock	Motion compensation unit
Block	DCT unit

by B-Pictures are stored and remain unchanged during B-Picture coding. I- and P-Pictures are transferred from Future Store to Previous Store while new pictures are stored in Future Store.

On the decoder side, the expected decoding process can be viewed as shown in Figure 3.7 [3]. The bit stream is demultiplexed into overhead information such as motion information, quantizer step size, and quantized DCT coefficients. The quantized DCT coefficients are dequantized and are input to the IDCT. The reconstructed waveform from the IDCT is added to the result of the prediction. Two reference pictures will be used for Bidirectional prediction.

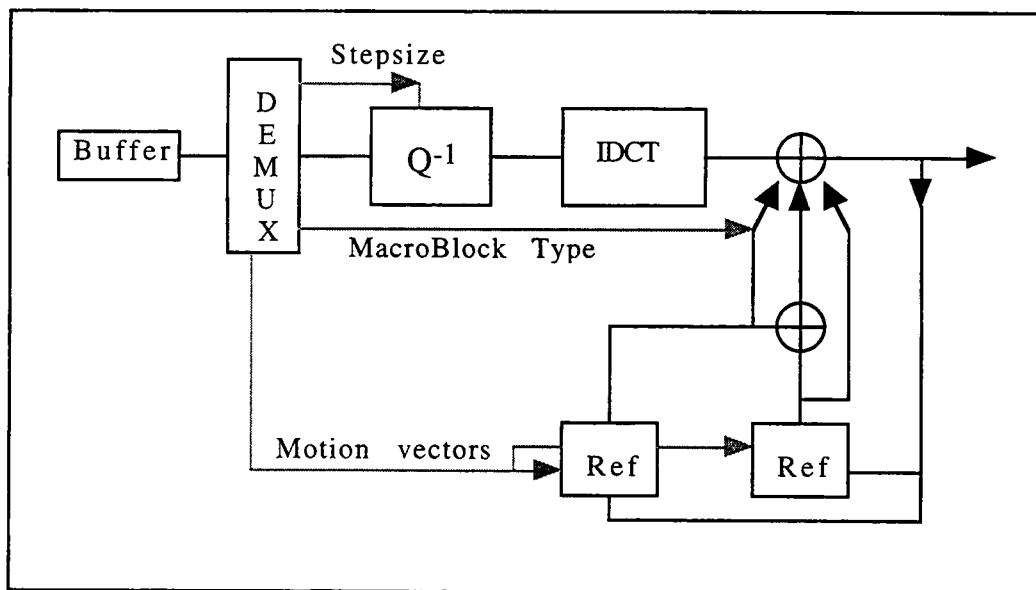


Figure 3.7: Schematic block diagram of an MPEG-1 decoder.

MPEG Phase 2

The second phase of MPEG was initiated in 1991 and aimed to code interlaced CCIR 601 video at bit rates of 2 Mbps or higher. Since MPEG-2 is not of practical use for videoconferencing because of its

relatively high bit rate, the discussion here will be limited so as to give a general idea. The reader can refer to [27] for detailed information.

The main feature of MPEG-2 is the use of scalable video coding [3, 24] for a range of applications where video needs to be decoded and displayed at a variety of resolution scales. Among the major applications of MPEG-2 are HDTV and video communications over ATM networks. Since it deals with interlaced video, MPEG-2 specifies a choice of two pictures. *Field Pictures* consist of the fields that are coded independently, and *Frame Pictures* where field pairs are combined into frames before coding. The main syntax structure of MPEG-2 consists of additions to MPEG-1 such as 10-bit quantization instead of 8-bits, flags for picture structure, and information for the reconstruction of a composite NTSC signal.

Since the MPEG-2 standard should achieve the highest possible quality of decoded video at a given bit rate and different requirements for different applications, additional constraints are placed on its syntax. For this task, MPEG-2 designates two sets of parameters: (1) *Profile*, deals with coding features supported, and (2) *Level*, deals with picture resolutions, bit rate, etc. The most used Profile-Level combination is known as *Main Profile at Main Level* (MP@ML) and its parameter constraints are shown in Table 3.2.

MPEG-2 defines an *Alternate Scan* instead of zig-zag scan in order to increase correlation between interlaced frames. This type of scanning, shown in Figure 3.8, is used in conjunction with the DCT for interframe coding. In intraframe coding, MPEG-2 defines, in addition to MPEG-1 techniques, other methods to deal with interlace and prediction between frames and fields.

Table 3.2: Parameters Bounds for MPEG-2
MP@ML [24].

Header	Content
Picture	20-bit picture start code 5-bit input frame number CIF/QCIF selection
GOB	16-bit GOB start code 4-bit GOB address Quantizer information
MB	MB address Type of information Quantizer information Motion vector data Coded block pattern

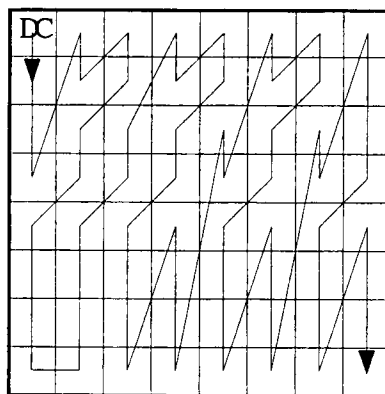


Figure 3.8: Alternate scan for
MPEG-2 coding.

3.2.3 ITU-T Recommendation H.261

As mentioned earlier, H.261 is a part of a series of recommendations called H.320 aimed at video compression at a bit rate of $P \times 64$ kbps, where P is between 1 and 30. The H.261 standard was designed primarily for videoconferencing over ISDN lines which support up to 1.544 Mbps. Basically, videoconferencing is carried through at what is known as the *Basic Rate*, which consists of two B-channels (64 kbps each), and one D-channel for signaling (16 kbps).

To achieve a high compression rate, H.261 utilizes both intraframe spatial and interframe temporal coding. In intraframe coding, as in JPEG and MPEG, DCT-based spatial compression is used, followed by VLC, and runlength/entropy encoding. In interframe coding, motion compensation is performed to compute the difference between the current and the previous frames. Since H.261 is designed for real-time applications, it uses the closest previous frame to reduce delay.

The H.261 standard defines only the decoder and expects the encoder to be well compatible with the bit stream. The compressed bit stream expected consists of several layers and supports two picture formats: namely, common intermediate format (CIF) and quarter-CIF (QCIF). The CIF picture consists of 352 pixels per line (ppl) by 288 lines per frame (lpf) for the Y component and 176 ppl by 144 lpf for the two color components, C_B and C_R . All standard codecs must be able to operate with QCIF: CIF is optional. Figure 3.9 shows the bit stream layers and Table 3.3 lists the content of each layer header. The differential motion vectors in the case of interframe coding are coded using VLC and then ordered in the zig-zag scanning pattern as shown in Figure 3.5.

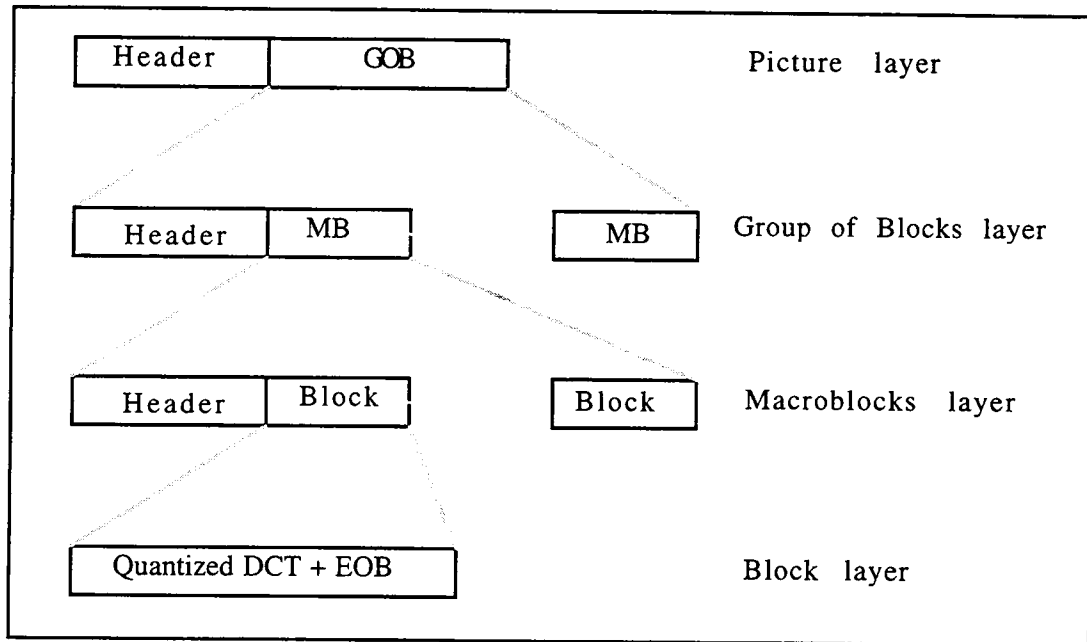


Figure 3.9: Layered structure of H.261 bit stream.

Table 3.3: Header content of H.261 bit stream layer.

Header	Content
Picture	20-bit start code(PSC) 5-bit temporal reference 6-bit type information PET + PSPARE (variable)
GOB	16-bit GOB start code (GSC) 4-bit GOB address 5-bit quantizer information GEI + GSPARE (variable)
MB	MB address (VLC) MTYPE (variable) 5-bit Quantizer information Motion vector data (VLC) Coded block pattern (VLC)

Video Stream

Each picture contains its header and data for Group of Blocks (GOBs). A header is made up of fixed length parts in addition to 1-bit extra insertion information (PEI) and variable length spare information (PSPARE) for future additions. A PEI set to "1" means the presence of an 8-bit PSPARE. The combination PEI + PSPARE could be repeated various times. The 5-bit temporal reference code is formed by incrementing its value in the previous picture whether it is transmitted or not.

A GOB comprises $1/12$ of the CIF or $1/3$ of the QCIF picture areas which consists of 176 pels by 48 lines of y and 88 pels by 24 lines of each of C_B and C_R . The GEI and GSPARE are the same as in picture layer, and The 4-bit GOB address indicates the position of each of the 12 GOBs in a picture. The last 3 positions are reserved for future use. Each GOB is divided into 33 MBs where each MB consists of 16 pels by 16 lines of Y and 8 pels by 8 lines of each of C_B and C_R . The MB address is a VLC indicating the position of each MB within a GOB. The MTYPE is also a VLC which determines the presence of each MB data elements. These elements are 5-bit quantizer, VLC motion vector data, and VLC coded block pattern. The motion vector data consists of 2 VLCs, one for horizontal component followed by one for vertical component. The coded block pattern gives a pattern number which indicates the transmitted transformed coefficients of each block in an MB. The block structure is the same as in JPEG and MPEG.

Multipoint and Bit Transmission

Some facilities are provided by H.261 to support multipoint. These facilities are: freeze picture request which causes the decoder to

freeze its displayed picture to allow transmission of a picture from different user and fast update request which forces the encoder to switch to intraframe coding in order to avoid buffer overflow. The buffer size in CIF is 256 kbps and in QCIF is 64 kbps which is the number bits created by coding single picture in each case. H.261 uses BCH (511,493) forward error correction code in order to correct any error caused by transmission.

This standard was designed primarily as a part of H.320 for videoconferencing over ISDN. But the emerging H.323 standard for LAN conferencing will likely be the basis for interoperability in conferencing on LAN which could include H.261 in addition to H.263, a video compression standard for LAN. The following chapter will discuss the issue of transmission media and the emerging technologies in this field.

Chapter 4

Transport Protocols

Chapter 3 addressed the issues of image acquisition and bit rate reduction using various compression techniques. This chapter will discuss some of the popular transport protocols which are used to carry traffic across the network. Some of these protocols are LAN-based such as Ethernet, others are WAN-based such as ISDN and ADSL, while some others can be used for both topologies such as ATM. All these protocols will be discussed, as well as their implementation for desktop videoconferencing.

4.1 Ethernet

LANs are widely deployed everywhere today, and most existing LAN installations have enough bandwidth to support desktop videoconferencing connections. With appropriate bandwidth on wide area connections, videoconferencing can occur in a LAN/WAN environment that extends around the world. But because each caller must share the transport medium in a LAN environment, managing bandwidth consumption can become an issue as more conferencing

sessions are added. However, communications infrastructure improvements such as 100 Mbps Fast Ethernet and higher-bandwidth WAN connections are available to help resolve bandwidth issues. Most networks today consist of 10 Mbps Ethernet and many are moving towards 100 Mbps Fast Ethernet. The physical limitation of Ethernet is being resolved or expanded by means of repeaters, hubs, bridges, and routers. These devices are used to connect different network types. On the WAN side, usually fractional or multiple T-1 (1.54 Mbps) links are used to connect various LANs.

There are several LAN technologies in use today, but Ethernet is by far the most popular. The 10 Mbps Ethernet media varieties include the original thick coaxial system, as well as thin coaxial, twisted-pair, and fiber optic systems. The most recent Ethernet standard defines the new 100 Mbps Fast Ethernet system which operates over twisted-pair and fiber optic media.

4.1.1 Ethernet Standard

After the success of Ethernet technology, the Institute of Electrical and Electronics Engineers adopted it as a standard for LANs through the LAN standard committee (IEEE 802) and published it under the title of "IEEE 802.3 Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications"[28]. The physical medium of any LAN consists of a combination of repeaters, bridges, hubs, and routers where each has a designated task:

Repeater: A repeater receives and then immediately retransmits each bit. It has no memory and does not depend on any particular

protocol. It duplicates everything, including the collisions.

Bridge: A bridge receives the entire message into memory. If the message was damaged by a collision (discussed later) or noise, then it is discarded. Otherwise, the message is queued and will be retransmitted on another Ethernet segment. The bridge has no address and it is transparent to the client and server workstations. It also connects LAN segments of different types.

Router: A router acts as an agent to receive and forward messages. The router has an address and is known to the client or server machines. Typically, machines send messages directly to each other when they are on the same segment, while they send the router messages addressed to another zone, department, or sub-network. TCP/IP is an example of routing protocol.

Hub: A hub can be used for connecting different segments such as a repeater, and for bridging multiple logical networks (local and remote). It can also perform routing.

The logical topology of an Ethernet provides a single channel (or bus) that carries Ethernet signals to all stations. Multiple Ethernet segments can be linked together in any physical topology to enlarge the physical limitation of a network. However, Ethernet segments must not be connected in loops.

All stations attached to Ethernet are connected to a shared signaling system, also called the medium. Ethernet signals are transmitted serially, one bit at a time, over the shared signal channel to every attached station. To send data a station first listens to the channel, and when the channel is idle the station transmits its data in the form of an Ethernet frame, or packet. After each frame transmission, all stations on the network must wait for the next

frame transmission opportunity. This ensures that access to the network channel is fair, and that no single station can lock out the other stations. Access to the shared channel is determined by the medium access control (MAC) mechanism embedded in the Ethernet interface located in each station [14].

The MAC mechanism is based on a system called Carrier Sense Multiple Access with Collision Detection (CSMA/CD). If more than one station happens to transmit on the Ethernet channel at the same time, then the signals are said to collide. The stations are notified of this event, and instantly reschedule their transmission using a given algorithm. As part of this algorithm the stations involved each choose a random time interval to schedule the retransmission of the frame. As more computers are added to a given Ethernet, and as the traffic level increases, more collisions will occur as part of the normal operation of an Ethernet. Repeated collisions for a given packet transmission attempt indicate a busy network. CSMA/CD does not ensure data delivery, it is up to a higher-level protocol that is sending data over the network to make sure that the data are correctly received at the destination station.

High-level network protocols such as TCP/IP, AppleTalk, or Novell IPX can do this by establishing a reliable data transport service using sequence numbers and acknowledgment mechanisms in the packets that they send over the LAN which are carried between nodes in the data field of Ethernet frames. The system of high-level protocols carrying application data and the Ethernet system are independent entities that cooperate to deliver data between computers. To discover the Ethernet addresses of other IP-based stations on the network, yet another high-level protocol called the Address

Resolution Protocol (ARP) is used [14]. ARP will send a request in a broadcast mode to every Ethernet interface to send the Ethernet address corresponding to a specific IP address. The station with this specific IP address will send its Ethernet address in a packet form to the station that requested it.

4.1.2 Frame Structure

The structure of an Ethernet packet, shown in Figure 4.1, consists of several fields, each with a specific length. These fields are:

Preamble	Destination address	Start frame delimiter	Source address	Type	Data field	CRC
(7)	(6)	(1)	(6)	(2)	(46-1500 bytes)	(4)

Fig. 4.1: Ethernet frame structure.

Preamble: A 56-bit synchronization field of alternating 1s and 0s.

Start frame delimiter: An 8-bit field indicating the start of frame.

Destination address: A 48-bit field which specifies to which interface the packet is addressed. If the first bit is 0, the packet is sent to one recipient; if it is 1, then the packet is for multicast use. An address of all 1's indicates broadcast mode.

Source address: A unique 48-bit address is commonly preassigned to each Ethernet interface when it is manufactured, which vastly simplifies the setup and operation of the network.

Length: This 2-byte field indicates the number of bytes used to

send data

Data field: The data field carries a variable 46 to 1500 bytes of data stream. if the length is less than 46, then an integer number of bytes are padded to the end of data field.

Frame check sequence: A 32-bit cyclic-redundancy-check code to check parity in each packet.

4.1.3 Fast Ethernet

There are two LAN standards that can carry Ethernet frames at 100 Mbps which are known as Fast Ethernet. When the IEEE standardization committee met to begin work on a faster Ethernet system, two approaches were presented. One approach was to speed up the original Ethernet system to 100 Mbps by applying changes to the physical medium and keeping the original CSMA/CD MAC mechanism. This approach is called 100 Mbps Ethernet. Another approach, known as 100VG-AnyLAN, presented to the committee was to create an entirely new MAC mechanism which is based on hubs that controlled access to the medium using a *demand priority* mechanism [29]. This new access control can transport standard Ethernet (IEEE 802.3) and Token Ring (IEEE 802.5) frames.

The 100 Mbps Ethernet standard proposes three types of cable wiring. They are: (1) 100Base-TX by using the standard Ethernet access over two pairs of type 5 Unshielded Twisted Pair (UTP) cable. (2) 100Base-T4 by using four pairs of type 3, 4, or 5 UTP cable. (3) 100Base-FX by using two strands of fiber optic cable [29] .

100VG-AnyLAN is defined as the IEEE 802.12 standard, that provides a data rate of 100 Mbps on four pairs of type 3 (voice

grade), 4, or 5 UTP cable. The Demand Priority Protocol (DPP), used by 100VG-AnyLAN, is a simple, deterministic request method that maximizes network efficiency by eliminating network collisions. It uses two levels of priority (normal priority and high priority) for each user request to guarantee support for emerging time-critical multimedia applications such as real-time video and audio for videoconferencing. The basic configuration of 100VG-AnyLAN network consists of a hub, known as level 1 or root, and a link connecting each node creating a star topology. A LAN segment of different type (802.3, 803.5, ATM, etc.) can be connected via a bridge. The hub is an intelligent central controller that manages the network access by continually performing a rapid "round robin" scan of its network port requests and checking for service requests from the attached nodes. The hub receives the incoming data packet and directs it only to the port with a matching destination address or addresses. this increases data security.

4.2 ATM

Recently, ATM has been advocated as an important technology for the wide area interconnection of heterogeneous network types, including traditional packet and synchronous switched networks. The advantages of ATM for the wide area network also extend into the local area network. The interconnection of these LANs/WANs to provide an end-to-end ATM service is an important component in providing a flexible service to applications. Computer manufacturers are beginning to promote the use of ATM techniques in privately owned and operated digital networks. Many networking companies

are coming forward offering private ATM solutions with a promise to migrate to upcoming standards (e.g., ITU) and emerging de facto standards (e.g., ATM Forum).

The interest in ATM has grown from the need for a worldwide standard to allow interoperability of information, regardless of the end-system or type of information [30]. With ATM, the goal is one international standard and a method of communication which can be used as the basis for both LAN and WAN technologies. Over time, as ATM continues to be deployed, the line between local and wide area networks will blur to form a seamless network based on one standard, ATM. Today, in most instances, separate networks are used to carry voice, data and video because these traffic types have different characteristics, but with ATM, separate networks will not be required. ATM is the only standards-based technology which has been designed from the beginning to accommodate the simultaneous transmission of data, voice and video. As described above, ATM is the emerging standard for communications. This is possible because ATM is available at various speeds from megabit to gigabit speeds.

4.2.1 ATM Technology

ATM, as a switching technology, is designed to meet the needs of high-speed networking. It implements fast packet switching in which data are carried in fixed length units called *cells*. Each cell consists of 53 bytes in length with the first 5 bytes used as a header. When a connection is requested, ATM employs a virtual circuit between users through the use of fast-switch hardware. The result is to provide users with the advantages of circuit switching and a

guarantee in transmission capacity and level of service needed [30]. In fact, the end users can negotiate a Quality of Service (QoS) contract before establishing a connection. This means that each user can request the network bandwidth that is required for a particular application, and ATM network provides it. ATM networks offer a specific set of service classes, each with a particular mix of traffic and QoS parameters. The set of QoS classes includes Continuous Bit Rate (CBR), Variable Bit Rate (VBR) for real time, and Variable Bit Rate (VBR) for non-real time. The transmission medium can also be shared by different users with each utilizing only the bandwidth requested.

ATM is a layered architecture allowing multiple services like voice, data and video, as well as, different types of connections to be mixed over the network. Three lower-level layers have been defined to implement the features of ATM [31]: (1) The Adaptation layer that makes sure of the appropriate service characteristics and divides all types of data into the 48-byte payload that will make up the ATM cell. (2) The ATM layer which takes the data to be sent and adds the 5-byte header information that assures the cell is sent on the right connection. (3) The Physical layer which defines the electrical characteristics and network interfaces. ATM is not tied to a specific type of physical transport.

The 5-byte header of an ATM cell is divided into several fields as shown in Figure 4.2. They are a 4-bit Generic Flow Control (GFC), a 24-bit Virtual Path Identifier/Virtual Channel Identifier (VPI/VCI) as a virtual connection number, a 3-bit Payload Type Indicator (PTI) to distinguish between services, 1-bit Cell Loss Priority (CLP) as a flag to assign higher reliability to certain packets, and 8-bit Header

Error Control (HEC) to detect any error in header. Note that ATM does not perform data error detection or correction, instead it delivers corrupt packets to higher-level protocols which are responsible for detecting and correcting any error. However, it does check errors in the header to avoid wrong delivery.

G F C	VPI	VCI	P T U	C L P	HEC	Payload
(4)	(8)	(16)	(3)	1	(8)	(48 bytes)

(*) : No. of bits

Fig 4.2: ATM cell format.

An ATM network consists of a set of ATM switches interconnected by point-to-point ATM links or interfaces. ATM switches support two kinds of interfaces: user-network interfaces (UNI) and network-node interfaces (NNI). UNI connect ATM end-systems (hosts, routers, and so on) to an ATM switch, while an NNI may be defined as an interface connecting two ATM switches together where slightly different cell formats are defined across UNI and NNI. In NNI cells, unlike UNI cells, there is no GFC field, and the first four bits of the cell are used by an expanded (12 bit) VPI field. ATM circuits are of two types: virtual paths identified by VPI and virtual channels, identified by the combination of a VPI and a VCI. A virtual path contains several virtual channels, all of which are switched transparently across the ATM network on the basis of the common VPI.

ATM signaling is initiated by an ATM end-system that desires to

set up a connection through an ATM network and signaling packets are sent on exclusive virtual channels. The signaling is routed through the network to set up the connection. The other end-user can either accept and confirm the connection request, or can reject it, clearing the connection. Note that the data flow along the same path of the connection request. The routing of the connection request and data flow is governed by the ATM routing protocol called Private-NNI (P-NNI) [30]. The protocol routes the connection request based upon both the destination address, and the traffic and QoS parameters requested by the source end-system. Unlike current packet switches which need to process every packet that is relayed, ATM switching systems only need to process a connection setup. Following connection setup, cells can be relayed without route processing. A new layer defined by ATM called ATM Adaptation Layer (AAL) was designed to support various kinds of services with different characteristics [31].

4.2.2 Adaptation Layer and Support

The basic operation for this layer is to adapt different types of applications to the ATM layer. Four types of AAL were recommended by ITU-T with the following characteristics:

AAL1: Supports connection-oriented services that require constant bit rates and have specific timing and delay requirements.

AAL2: Supports connection-oriented services that require variable bit rate applications such as video.

AAL3/4: This AAL is intended for both connectionless and

connection oriented variable bit rate services. Originally two distinct adaptation layers AAL3 and 4 were defined, they have been merged into a single AAL.

AAL5: Supports connection-oriented variable bit rate data services. Some organizations have proposed AAL5 for use with both connection-oriented and connectionless services.

All AAL types are composed of a convergence sublayer (CS) and a segmentation and reassembly (SAR) sublayer. CS is further composed of a common part (CPCS) and a service specific part (SSCS). SAR segments higher layer packets into 48-byte packets that are fed into the ATM layer to generate 53 byte cells. CPCS provides services such as padding and CRC checking. It also takes SSCS packets, adds padding if needed, and then adds an 8-byte trailer such that the total length of the resultant packet is a multiple of 48. The trailer consists of 2 bytes reserved, 2 bytes of packet length, and 4 bytes of CRC. SSCS is service dependent and may provide services such as assured data transmission based on retransmission.

ATM is a promising switching technology due to its wide range of application support and speed. In addition to AAL, ATM support LAN Emulation (LANE) protocol which defines mechanisms for emulating either an IEEE 802.3 Ethernet or 802.5 Token Ring. This will ensure a gradual shift towards adopting ATM for LAN and WAN as well. Although switching to ATM is not moving fast, future provisions support a bit rate of up to 600 Mbps using B-ISDN technology. This rate is sufficient not only for videoconferencing, but for all audio, video, and data applications.

4.3 ISDN

ISDN is a set of international digital transmission standards developed by ITU-T which aims to provide digital services including data, audio, and video over the public switched network. It offers a dramatic increase in the speed of data and document transfer as compared to modems: Bearer channels (or B channels) transmit user information at relatively high speeds, while separate Data channels (or D channels) carry call setup, signaling and other information. It is a network "that provides end-to-end digital connectivity to support a wide range of services by a limited set of standard multipurpose user-network interfaces" [32]. In this section, the ISDN basic structure and transport protocols will be discussed.

End-users connect to ISDN service through simple standardized interfaces. In addition, an ISDN switch could connect various networks to users through an access line. These networks could be circuit-switched, packet-switched, public, or private network. The access line, which could be PSTN that uses local copper loops or a fiber optic link, is terminated at the user premises with an ISDN interface [6,33]. Figure 4.3 shows an ISDN interface with all components and reference points.

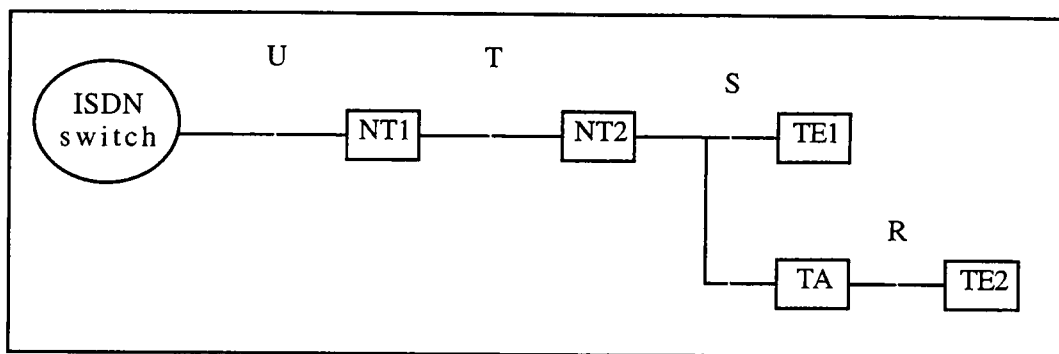


Fig. 4.3: ISDN interface between switch and terminal equipments.

4.3.1 ISDN Architecture and Standards

The transmission line coming to a user premises, a twisted-pair wire with reference U, is connected to a Network Termination 1 (NT-1) unit. NT-1 performs OSI layer 1 functions associated with the physical and electrical termination of access lines, and serves as an isolation point for the user from the transmission loop. It also converts the 2-wire U reference into 4-wire T reference. The T reference, as well as NT-2, are associated with the OSI data link layer (layer 2), where one pair is used for receiving data and the other for transmission. NT-2 handles OSI layers 2 and 3 (network layer) functions such as switching and concentration. NT-2 may be a digital private branch exchange (PBX), a LAN, or a terminal multiplexer. Terminal Equipment 1 (TE1) is an ISDN device (i.e., a device that is ISDN compatible such as a digital telephone). TE2 is a non-ISDN terminal that requires a Terminal Adaptor (TA) to connect to ISDN reference point S. Any ordinary telephone is a TE2 device.

The physical access connecting users to an ISDN node is a digital transmission facility that can support a number of communications channels bundled together. ISDN supports two rates: Basic Rate

Interface (BRI) with a bit rate of 144 kbps (2B + 16 kbps D channel), and Primary Rate Interface (PRI) with a bit rate of 1.544 Mbps in North America and Japan (23B + 64 kbps D channel), or 2.048 Mbps in Europe (30B + 64 kbps D channel). Note that the bit rate of each B channel is 64 kbps.

On the network side of the ISDN interface, the switch must separate the channels and route them appropriately. For instance, the B channel can be routed to a circuit-switched or packet-switched network. The same is the for the D channel, but the signaling information goes to the signaling network and the data part is routed to a packet-switched network.

A typical point-to-point videoconferencing session can be carried over BRI; however, the video quality is reduced with a frame rate about 5-8 fps. PRI, on the other hand, offers enough bandwidth to support high quality for several users at one time. In fact, as mentioned earlier, ITU-T recommendation H.320 incorporating the video compression standard H.261 can achieve more than 20 fps over PRI. The following section will discuss the frame structure and routing protocols of ISDN.

4.3.2 ISDN Protocol

To be consistent with other network services, ISDN exchange data according to layered protocols. The ISDN protocol reference model is based on the OSI model in information transfer and signaling. For signaling between the user and the network, the D channel uses what is known as Link Access Protocol on the D channel (LAP-D) data link layer, as specified in the ITU-T recommendation Q.921 [34]. Layer 3 of the ISDN access signaling protocol, specified in ITU-T

recommendation Q.931 [35], provides user-network interface specifications and procedures for call establishment, maintenance, and termination.

The following is a brief look at the specific user access protocols defined by ISDN and their frame structure. Namely, the physical layer, the data link layer, and the network layer.

The Physical Layer

The physical layer is the means by which the user terminal gains access to the channels. The standardized specifications for layer 1 are specified in ITU-T recommendation I.430 [36]. The D and B channels as well as control signals are combined by time-division multiplexing and transmitted in each direction on a wire pair. The transmitted signals are encoded by Alternate Mark Inversion (AMI). In this coding scheme, the "1s" are represented by the absence of pulses and the "0s" by equal-amplitude alternating pulses.

An ISDN frame consists of 48 bits $\{16(2)B + 4D + 12\}$ with a duration of 250 μ sec. this corresponds to a total bit rate of 192 kbps ($2B + D + 48$ kbps overhead). Each 8 bits from B channel and 1 bit from D channel are separated by control bits for synchronization and cyclic redundancy check (CRC). In PRI, the frame length is 193 bits at 1.544 Mbps and 256 bits at 2.048 Mbps with a duration of 125 μ sec each. The frame is divided into 24 or 31 8-bit time slots preceded by 1- or 8-bit, respectively, for control purposes.

Data Link Layer

The purpose of LAP-D is to convey user information between layer 3 entities across ISDN using the D-channel. It offers two types

of operation:

Unacknowledged operation: In this type, layer 3 information is transferred in unnumbered frames. Error detection is used to discard corrupted frames but with no error control. This is an important point for videoconferencing since delay must be kept at a minimum.

Acknowledged operation: This type of operation is the opposite of the first one.

The LAP-D frame consists of several fields: An 8-bit flag is inserted at both ends of each frame and has a pattern of 01111110. LAP-D uses bit-stuffing to provide transparency by inserting a 0 bit after all sequences of five continuous 1s. The second field is a 16-bit address which consist of a 7-bit Terminal Endpoint identifier (TEI), 6-bit Service Access Point Identifier (SAPI), 1-bit Command/Response (C/R) identifier, and one extension bit for each of TEI and SAPI to indicate the end. The third field is a frame level control field which indicates the frame type. Its length is 16 bits for information transfer or supervisory, or 8 bits for unnumbered frames which provide supplemental link control. The fourth field is for information where it could be of any length of integer number octets. Before the end flag, the fifth field is a 16-bit Frame Check Sequence using a CRC code.

Network Layer

This layer is the user-network interface [35]. It specifies procedures for establishing connections on the B channels that share the same interface to ISDN as the D channel. It also provides user-to-user control signaling over the D channel such as call setup. The

length of the message frame is variable and depends on the type of service being offered. The message frame contains the following fields: (1) 8-bit Protocol Discriminator to identify layer 3 protocol (i.e., Q.931, X.25, etc.); (2) 8 or 16-bit Call Reference Value (CRV) to identify the B channel call to which this message refers, preceded by an 8-bit length field to identify CRV field length; (3) 8-bit Message Type followed by a variable-length Information Element field.

4.3.3 Broadband-ISDN

B-ISDN is a set of services being defined by the ITU-T for the transport of all types of data at speeds starting at about 150 Mbps and employing ATM switching technology [37]. The transport backbone for B-ISDN is an optical time-division-multiplexed network based on synchronous-optical-network (SONET) technology in the United States and on synchronous-digital-hierarchy (SDH) in Europe. Table 4.1 lists the digital hierarchy of SONET and SDH channels.

Table 4.1: Bit rate hierarchy of fiber optic channels.

SONET	SDH	Bit rate (Mbps)
STS-1		51.84
STS-3	STM-1	155.52
STS-9	STM-3	466.56
STS-12	STM-4	622.08

The STS-3 signal is formatted into frames of 2430 bytes with a duration of 125 μ sec. Signaling in B-ISDN is similar to ISDN format, carried on a separate Virtual Channel (VC).

4.4 ADSL

The Asymmetric Digital Subscriber Line (ADSL) standard, developed by ANSI group T1E1.4, aims to provide over a single twisted-pair line a number of variable bit rate channels (up to 6 Mbps) from the central office to a subscriber (downstream) and a number of duplex channels (upstream) (up to 640 kbps) while leaving Plain Old Telephone Service (POTS) unaffected[38]. This technology is basically designed to match the requirements for video on demand and information services over data networks using modems. The asymmetric nature is due to physical limitations which are imposed by line attenuation increases with line length and frequency, and decreases as wire diameter increases. The following section will discuss ADSL in its general form

The first step in using Digital Subscriber Line (DSL) was in High-bit-rate DSL (HDSL) which delivers a bidirectional T1 (1.544 Mbps) connection over two pairs of twisted copper wire up to a range of 12,000 feet assuming 24 AWG cable. By limiting high-rate transmission to one direction only, crosstalk from the same service is reduced, allowing further increase in the data rate. Given that many service applications such as entertainment video and information services (Internet) do not require bidirectional transmission, the concept of ADSL was well suited.

An ADSL circuit connects an ADSL modem on each end of a twisted-pair telephone line, creating three information channels; a high speed downstream channel, a medium speed duplex channel, and a POTS channel. The high speed channel ranges from 1.5 to 6.1 Mbps, while duplex rates range from 16 to 640 kbps. These channels

are Frequency Division Multiplexed (FDM) which will ensure uninterrupted POTS services, while downstream and upstream channels are divided by Time Division Multiplexing (TDM) into several low or high speed channels.

In order to reduce the effect of noise, the downstream channel employs Discrete Multitone (DMT) modulation [38]. This technique divides the available bandwidth into 256 subchannels of width 4 kHz each and allocates transmission power to the individual subchannels so that their throughput is maximized. As with all technologies designed for real-time applications, ADSL uses FEC for error correction which further increases its throughput.

In addition to ADSL, three major switching and transmission technologies were discussed in this chapter. Some of them are widely deployed such as Ethernet and ISDN; others are still under further development such as ADSL and B-ISDN. After presenting in Chapter 3 some of the compression methods are used for bit rate reduction and in Chapter 4 several transmission protocols to carry such traffic, Chapter 5 will evaluate some of the available desktop videoconferencing systems that fit campus environments.

Chapter 5

Cross Platform Issues

In early development, desktop videoconferencing systems relied on proprietary tools and so different platforms were not interoperable. In addition, the high cost of systems was a reason for the low deployment of videoconferencing. However, with the emergence of standards and the reduction in cost, systems are becoming more popular.

Some of the ITU-T finished standards such as H.320, H.324, and T.120 were discussed in Chapter 2, while others such as H.323 for videoconferencing over LANs are still under development. These standards will form the core of cross-platform solutions for videoconferencing products from different vendors. As a result, other video conferencing methods such as the MBone will probably lose some of their interest.

This chapter discusses the basic concepts behind three popular videoconferencing solutions including the MBone and its videoconferencing tools. It also discusses two commercial products, Apple QTC and Intel ProShare, and their features. The information presented here about these two products was compiled mostly from documents published on the web sites of their prespective companies [4, 39].

5.1 Multicast Backbone (MBone)

The MBone is a virtual network layered on top of portions of the physical Internet to support routing of IP multicast packets. Using the MBone, it is possible to transmit audio, video, and other data in real-time to multiple destinations throughout the Internet. Originated in 1992 by the Internet Engineering Task Force (IETF), the MBone was first used in an experiment to transmit live audio and video from their meetings to the Internet community.

The MBone network is composed of islands that can directly support IP multicast, linked by virtual point-to-point links called *tunnels*. The tunnel endpoints are typically UNIX workstations having operating system support for IP multicast and serve as multicast routers (mrouter)[1]. Mrouters are necessary for routing packets through non-IP multicast gateways. Unlike unicast datagrams which could pass through any gateway in its path, an IP multicast packet is encapsulated using the address of another mrouting host as the destination. The mrouter at the other end of the tunnel receives the packet, removes the encapsulating IP header, and forwards the packet as appropriate. IP multicast as defined by Request For Comment 1112 (RFC 1112), “is the transmission of an IP datagram to a *host group*, a set of zero or more hosts identified by a single IP destination address. A multicast datagram is delivered to all members of its destination host group with the same *best-efforts* reliability as regular unicast IP datagrams.” [40]. The host communicates its group membership by using the Internet Group Management Protocol (IGMP) which reports if there are any active listeners on a certain subnet.

Routing

The IP host group address uses class D IP addresses (the first 4 high order bits are 1110- in dotted decimal notation; the range is 224.0.0.0 to 239.255.255.255) followed by the particular multicast group in the remaining 28 bits. The mapping of the IP host group address to the Ethernet multicast address is done by placing the low order 23 IP host group address bits into the low order 23 bits of the Ethernet multicast address. Routing between mrouter is carried through the Real-Time Transport Protocol (RTP) which provides sequenced datagram delivery, but makes no guarantees about delivery and does not provide flow control. In order not to flood the network with unnecessary traffic, Mbone includes procedures for limiting the travel of multicast packets. Each packet has a time-to-live (TTL) value, a counter that is decremented every time the packet passes through an mrouter until it reaches zero.

MBone Tools

MBone supports tools based on IP multicast of real-time video, audio, and shared whiteboard. Some of the better known Mbone tools are visual audio tool (*vat*), whiteboard (*wb*), session directory (*sd*), and network video (*nv*). The first three tools were developed by Lawrence Berkeley Laboratory (LBL) and the fourth one by Xerox PARC. *Vat*, used for audioconferencing, provides a μ -law encoded audio at 64 kbps. *Wb* provides a shared whiteboard for collaboration on text files, graphs, drawings, etc. *Sd* is a tool that lists all available Mbone sessions. From *sd*, a user can start *vat*, *nv*, or *wb* sessions, or join any prescheduled session. When starting a new session, *sd* provides the user a randomly selected, unique IP multicast address,

a port number and application discriminator [41].

MBone is seen by many as a temporary tool until multicasting is a standard feature in commercial routers; however, MBone tools can still be used. Although all MBone tools can be acquired free, finding a UNIX-based workstation for each user is somewhat expensive and inappropriate for wide spread campus use. In addition to this, the bandwidth capability of the MBone (about 500 kbps) is limited to about four *nv* or eight *vat* sessions, making it an unattractive solution for videoconferencing.

5.2 QuikTime Conferencing (QTC)

QTC is a Macintosh system extension which enables videoconferencing and collaboration on the Mac platform by adding real-time media communications features. It includes Apple Media Conference as a conferencing tool that supports various network protocols, compression and communications standards. Running on any Power Macintosh supported with an AV board, QTC is a software-only solution capable of point-to-point, multipoint, and broadcast connections. It supports applications running on existing LANs, WANs, and through international standards, ISDN connections.

Features

QTC comes in two basic packages: a low-end package which runs on existing LANs but does not support any cross platform operation, and a high-end package for use over ISDN which is interoperable with any system supporting the H.320 standard. It is expected that support for new standards will be adopted when those standards

become available. This includes H.324 for videoconferencing over POTS, and H.323 for videoconferencing over LANs.

Several video and audio compression techniques are supported by QTC; some of them are standards-based, others are proprietary. Table 5.1 lists all the audio and video codecs which are supported by the QTC LAN kit and ISDN kit. An extension of the previous one, supports ISDN connections in addition to LAN connections. Since this technology is still under development, future releases would include any new ratified standards.

Table 5.1: Video and audio codecs supported by QTC LAN and ISDN kits.

	Video Compressor	Audio Compressor
LAN support	H.261	Mu-law
	JPEG	MACE 3:1
	Apple Video	MACE 6:1
ISDN support	H.261	G.711, G.722, G.728

Features of QTC include the support of various transport protocols such as AppleTalk and TCP/IP the ability to establish multipoint connections where several locations can communicate and collaborate at one time. Although QTC can support up to eight simultaneous sessions, video frame rates and audio quality drop drastically with more than two connections at one time. In fact, this feature was tested personally by connecting three different locations and it was founded that, with the available bandwidth, the maximum frame rate which could be achieved is about 1-3 fps with high audio jitter. Bandwidth monitoring and control can be done via SNMP. Other

features that are supported by QTC are the ability to broadcast or to watch a broadcast, and to record any session as a QuickTime movie and play it back. It also allows for taking snapshots and saving them as PICT files.

On the collaboration side, users can share data in real-time such as graphics, sound files, QuickTime movies, and text documents by dropping them into the file sharing window. In addition, the shared whiteboard window offers annotation capabilities including basic drawing and writing tools. The QTC version which includes the ISDN kit supports data conferencing standards (i.e., T.120).

User Interface

With the QTC LAN kit, a user can interact with three basic windows in addition to a separate viewing window for each end user. These three windows are: a self view window that shows the user his own video input, a file sharing window where virtually any file can be dropped in or dragged out, and a whiteboard window for collaboration and annotations. Each video display window contains audio level control, pause, record, and snapshot buttons. Users can also establish more than one session at the same time, start a broadcast, or watch a broadcast using AppleTalk or TCP/IP.

QTC offers the user a wide range of settings and control options such as answering calls, displaying messages or warnings, remembering most recent calls, changing the ring sound, and whether to compress shared files or not. A user can also choose to send audio, video, or both with several options. For video settings, QTC offers various compression options and direct control for quality level, number of fps, color resolution, and input format (i.e., NTSC,

PAL, or SECAM). The sound setting includes audio compression options, sampling rate (8 kHz and 22.05 kHz), and number of bits per sample (8 or 16 bits).

Requirements

To run QTC, a user needs any Macintosh AV model or Power Macintosh 7500/8500. In addition, 16 MB of RAM and system 7.5 or later are required. An Ethernet using TCP/IP or AppleTalk protocols for LAN/WAN connection, or an ISDN BRI connection terminated with NT-1 for conferencing over ISDN is needed.

5.3 Intel ProShare

The Intel ProShare desktop videoconferencing system is a PC-based solution for video, audio, and data conferencing. As with QTC, it also provides notebook and file sharing, and supports international communications standards such as H.320 and T.120 for data conferencing. Intel offers various videoconferencing solutions, each intended for a specific use. These products include ProShare 150 as a LAN-based solution, ProShare 200 for use over ISDN and LAN, and ProShare Video Phone for use over POTS.

Based on the information posted at Intel's web site, the ProShare 200 supports H.320 standards and can communicate with any cross platform H.320-compliant system. The T.120 standard is supported by both systems, ProShare 150 and 200. The Video Phone with ProShare technology conforms to the H.324 standard which ensures interoperability with other H.324-compliant systems. However, the discussion here will be limited to the first two systems since it is intended as a comparison with Apple QTC.

Features

The ProShare 150 supports most of the major features of video and data conferencing systems such as point-to-point video and multipoint data connections, whiteboard sharing and file transfer, the ability to store frequently dialed addresses or ISDN numbers in the case of ProShare 200, and recording sessions. But it does not support broadcasts as in QTC. It does include application sharing based on T.120 standard. Application sharing permits users to have access to software which is installed on remote units. Intel's LANDesk Conferencing Manager lets users control bandwidth usage and manage conferencing traffic by choosing the quality level.

Various network protocols are supported, such as IPX, TCP/IP, and NetBIOS. It is also independent of the LAN physical layer, i.e., Ethernet, TokenRing, FDDI, or T-1. The compression techniques supported by ProShare 150 are Indeo, a proprietary video compressor, with video resolution of 160×120 and 320×240, and GSM for audio compression with echo cancellation option and full-duplex video and audio conferencing capabilities. An ISA video input card supporting Composite and S-Video inputs, and audio subsystem with connectors for headset and microphone is included with the package.

The ProShare 200, in addition to the features of ProShare 150, supports multipoint videoconferencing over ISDN in compliance with the H.320 standard. Using Multipoint Control Units (MCU), it supports up to 8 participants with audio/video/data capabilities and 24 participants for audio/video only. On LANs, it supports 5 data connections, while audio/video is point-to-point only. The video resolution is CIF for receive and display, and QCIF for transmit to any

interoperable system. The video and audio compression techniques supported, as with any H.320-compatible system, are H.261 and G.711 and G.728, respectively. The package includes an ISDN card and NT-1 interface.

Requirements

To run the ProShare system, the minimum requirements are an Intel DX2 with at least a 66 MHz microprocessor, 16 MB of RAM, and 32 MB of disk space. A network connection supported by TCP/IP, IPX, or NetBIOS protocols for LAN/WAN conferencing, or an ISDN BRI connection terminated with NT-1 for ISDN is needed.

Future Implementations

Videoconferencing systems are still in their evolution phase and many requirements are still needed. In addition to Intel ProShare and Apple QTC systems, many systems (some of them are listed in Chapter 2) are available for users but they still do not satisfy their needs which may be summarized by "better quality and interoperability for an affordable price". The development of international standards will certainly give a boost to these systems which will be of no practical use if two different systems cannot communicate. With these standards, as the need for better performance increases, more complex and extensive computations are required and can not be delivered without the enhancement in desktop computers' processing speed.

In order to experiment with such systems and their capabilities to deliver what they promise, some testing was done using the Apple QTC solution. Chapter 6 explain all of the testing procedures and their results.

Chapter 6

Field Trials and Results

In order to evaluate various video and audio compression systems, some quantitative measurements have to be made. Measurements of quality for each compressor were based on how efficient it was in bandwidth utilization. In other words, a compressor with a lower bit rate at a given quality and frame rate, is deemed a better one. To do so, three network topologies running Apple QTC were considered. The network topologies are as follows:

- 1- Sending and receiving computers are on the same LAN segment.
- 2- Sending and receiving computers are on different LAN segments.
- 3- Sending computer on UT campus and receiving computer at a remote location connected via an ISDN WAN link.

For each topology, a two-way audio/video conference was established using TCP/IP for transport. Testing was done on three video compressors and three audio compressors under various video quality settings and audio sampling rates. However, all systems were tested for a constant frame rate, specifically 10 fps. Testing was done using a 6100/60AV Power Macintosh computer running the MacSNMP utility which reports bandwidth requirements for each

setting.

The following three sections describe the three different topologies used for testing and list all data statistics results. Since measurements can vary from time to time, depending on network traffic conditions, data were taken at five different times and averaged. As a result, the numbers presented in tables and their corresponding graphs are empirical values.

6.1 First Topology

The sending and receiving computers used for this topology were connected to the same 10 Mbps Ethernet LAN segment as shown in Figure 6.1. As a result, it was expected that bandwidth utilization would be higher than the other two topologies since data do not have to pass through a router which means less traffic and packet loss.

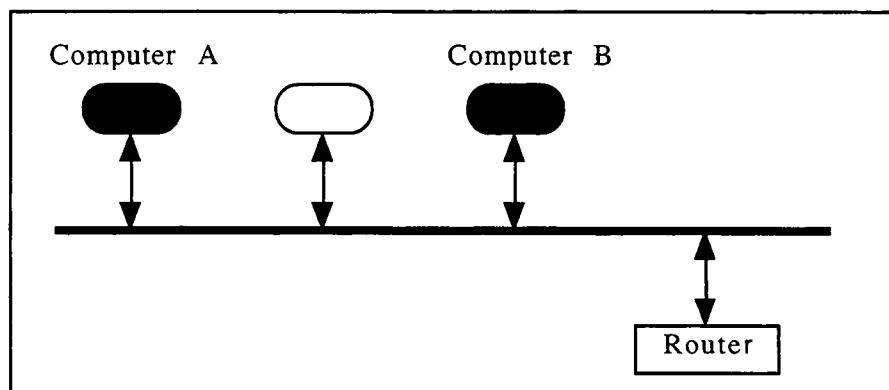


Fig. 6.1: Schematic representation of the first topology.

The first task under this topology was to measure the bandwidth requirements for three video compressors under the condition of transmitting video and audio. The audio compressor selected for this task was μ -law with a compression ratio of about 2:1, 8 kHz sampling rate, and 16 bits per sample. The choice of 16 bits per sample was mandatory since QTC does not support 8 bits per sample for this compressor. The three video compressors which were chosen included H.261, JPEG, and the proprietary compressor Apple Video. Also, measurements were made for various video qualities. The results for these measurements are listed in Table 6.1 and plots of the average values are shown in Figure 6.2. These results show that, at high video quality for example, H.261 requires about 260 kbps, JPEG requires 445 kbps, and Apple Video requires about 950 kbps.

Table 6.1: Bandwidth measurements for video compressors with μ -law audio compressor.

Video compressor	Quality	High BW (kbps)	Low BW (kbps)	Average BW (kbps)
H.261	Low	250.54	200.25	234.28
	Medium	257.04	238.82	245.44
	High	266.22	243.38	260
JPEG	Low	267.47	212.12	253.88
	Medium	384.22	324.82	367.9
	High	469.66	359.98	444.5
Apple Video	Low	621.27	524.98	584.88
	Medium	783.89	718.03	764.32
	High	982.23	872.55	948.72

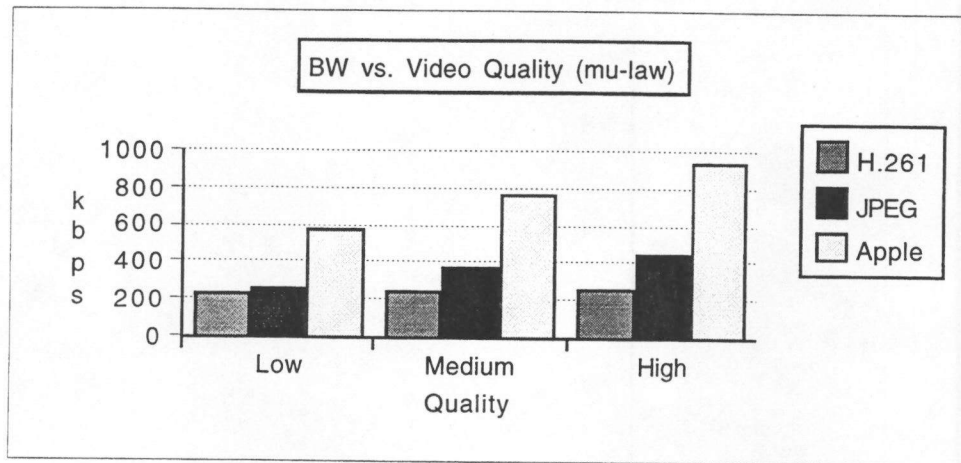


Fig 6.2: Plots of the average bandwidth of Table 6.1.

The second task was to measure bandwidth utilization of the three supported audio compressors: μ -law, MACE 3:1, and MACE 6:1. The settings were changed so that one video compressor, H.261 with low quality, is now used along with these three audio compressors tested at 8 kHz and 22.05 kHz sampling rate. The results are shown in Table 6.2 and plots of the average values in Figure 6.3. The best bit rate was found to be achieved by MACE 6:1 at an 8 kHz sampling rate. However, sampling at 22.05 kHz produces better audio quality.

The third task was to measure bandwidth utilization of the three audio compressors without transmitting video signals. Table 6.3 shows the resulting measurements which are plotted in Figure 6.4. In this task, results reflect audio requirements for each compressor where a clear comparison can be made, for instance, between MACE 3:1 and MACE 6:1 since they both were measured at a sampling size of 8 bits. By comparing their results, MACE 6:1 requires half the bandwidth of MACE 3:1 to transmit the same information. The results of μ -law were higher because of its lower compression ratio and larger sample size of 16 bits.

Table 6.2: Bandwidth measurements for audio compressors with H.261 video compressor.

Audio compressor	Sampling rate (kHz)	High BW (kbps)	Low BW (kbps)	Average BW (kbps)
mu-law	8	292.73	236.51	271.39
	22.05	385.85	343.52	370.08
MACE 3:1	8	257.28	183.66	222.46
	22.05	296.66	248.09	269.62
MACE 6:1	8	241.7	184.8	208.21
	22.05	262.28	200.85	224.07

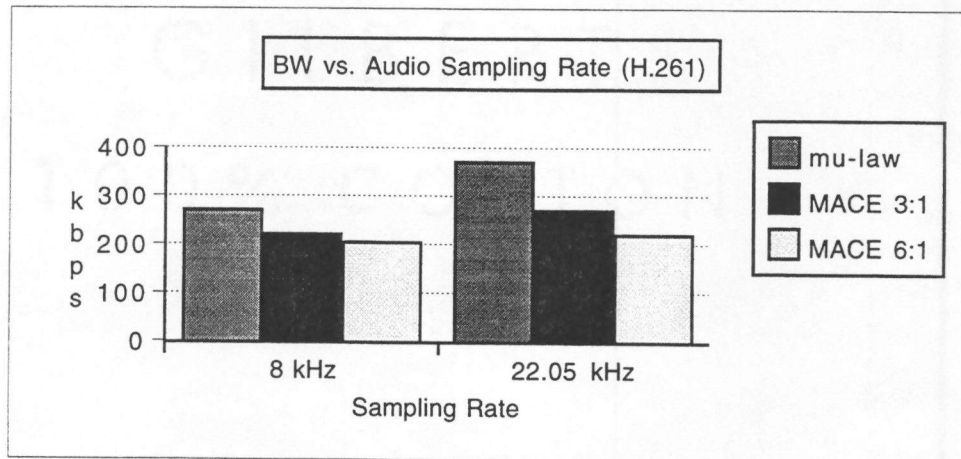


Fig. 6.3: Plots of the average bandwidth of Table 6.2.

Table 6.3: Bandwidth measurements for audio compressors with no video transmission.

Audio compressor	Sampling rate (kHz)	High BW (kbps)	Low BW (kbps)	Average BW (kbps)
mu-law	8	79.9	78.85	79.4
	22.05	220.25	219.73	220
MACE 3:1	8	26.8	26.45	26.75
	22.05	73.34	72.83	73.17
MACE 6:1	8	13.22	13.22	13.22
	22.05	37.15	36.5	37.72

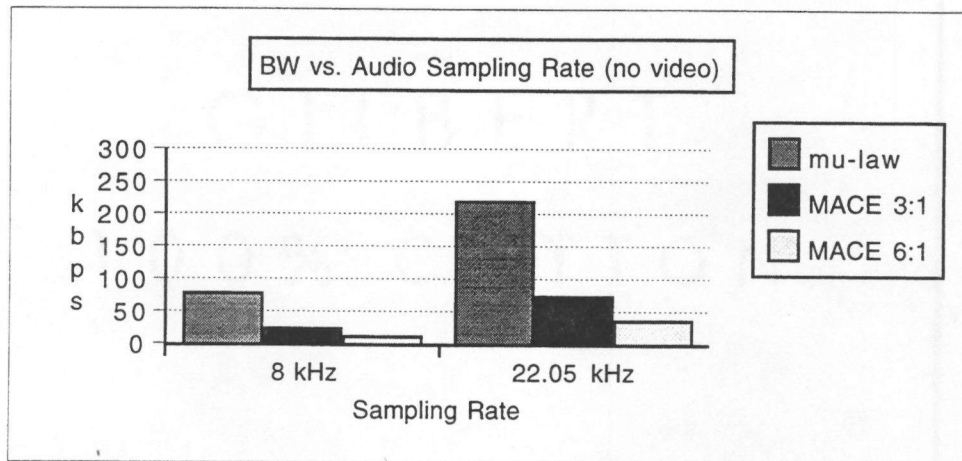


Fig 6.4: Plots of the average bandwidth of Table 6.3.

6.2 Second Topology

The same measurements were made for a different network topology such that the sending and receiving computers were on different LAN segments as shown in Figure 6.5. For this task, a connection was established between the Science and Engineering Building and Ferris Hall. As in the previous section, measurements were taken for three scenarios and under different video quality settings and sampling rates. Table 6.4 list the results for measuring video quality of the three video compressors and Figure 6.6 shows a graphical representation of the average bandwidth obtained. Similarly, Table 6.5 and Figure 6.7 show the results of the three audio compressors for different sampling rates with H.261. Table 6.6 and Figure 6.8 show the results for the case of audio only.

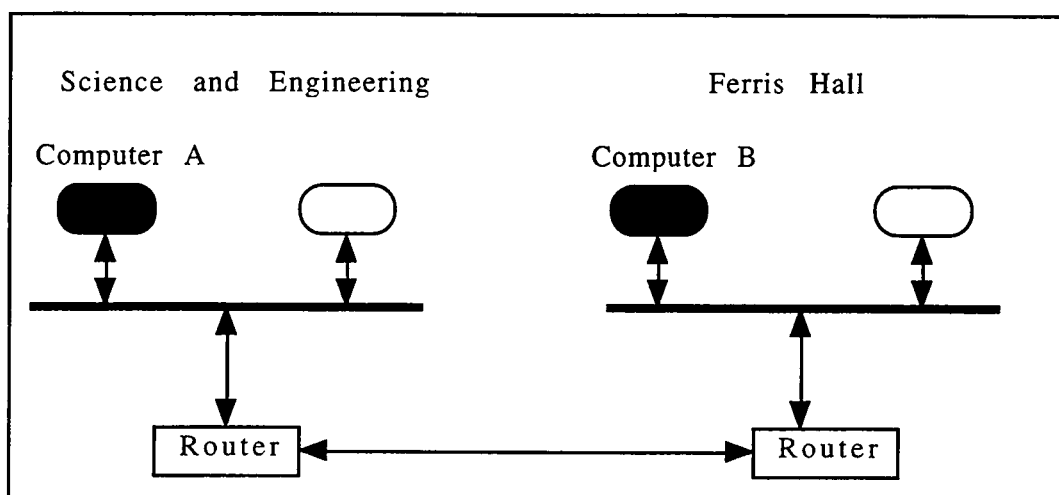


Fig. 6.5: Schematic representation of the second topology.

Table 6.4: Bandwidth measurements for video compressors with μ -law audio compressor.

Video compressor	Quality	High BW (kbps)	Low BW (kbps)	Average BW (kbps)
H.261	Low	169.83	148.85	163.99
	Medium	180.44	171.62	176.7
	High	182.81	177.02	180.77
JPEG	Low	172.2	162.24	168.89
	Medium	213.43	199.54	205.37
	High	267.4	228.82	252.75
Apple Video	Low	275.68	259.41	268.7
	Medium	328.05	310.86	325.56
	High	391.18	340.68	360.68

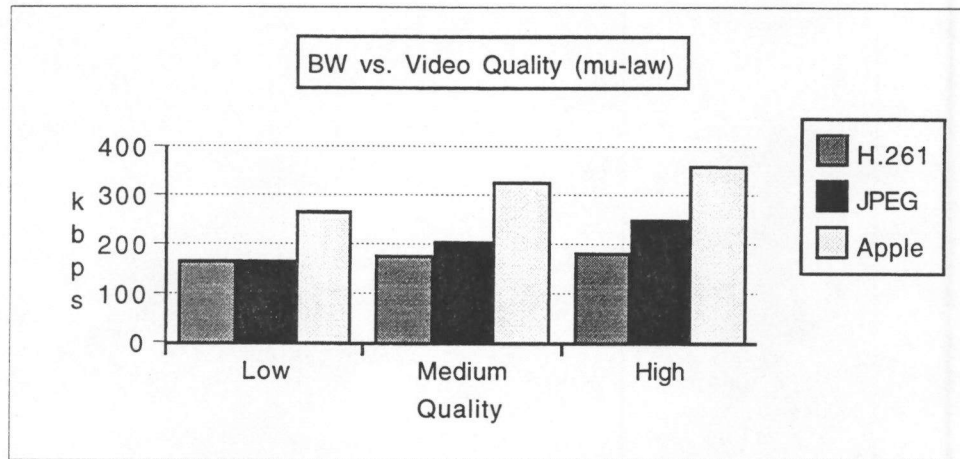


Fig 6.6: Plots of the average bandwidth of Table 6.4.

Table 6.5: Bandwidth measurements for audio compressors with H.261 video compressor.

Audio compressor	Sampling rate (kHz)	High BW (kbps)	Low BW (kbps)	Average BW (kbps)
mu-law	8	167.31	161.33	164.32
	22.05	331.22	327.02	328.62
MACE 3:1	8	104.83	93.58	101.52
	22.05	157.28	141.36	151.97
MACE 6:1	8	89.88	71.82	83.92
	22.05	110.75	108.7	111.23

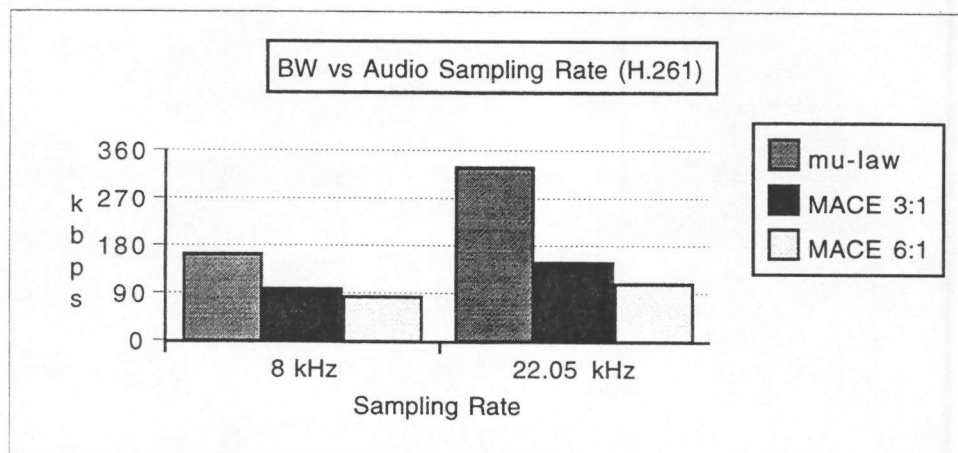


Fig. 6.7: Plots of the average bandwidth of Table 6.5.

Table 6.6: Bandwidth measurements for audio compressors with no video transmission.

Audio compressor	Sampling rate (kHz)	High BW (kbps)	Low BW (kbps)	Average BW (kbps)
mu-law	8	96.69	95.98	96.4
	22.05	261.65	258.84	260
MACE 3:1	8	32.46	32.46	32.1
	22.05	87.25	87.25	86.35
MACE 6:1	8	15.94	15.94	15.8
	22.05	43.57	43.57	43.1

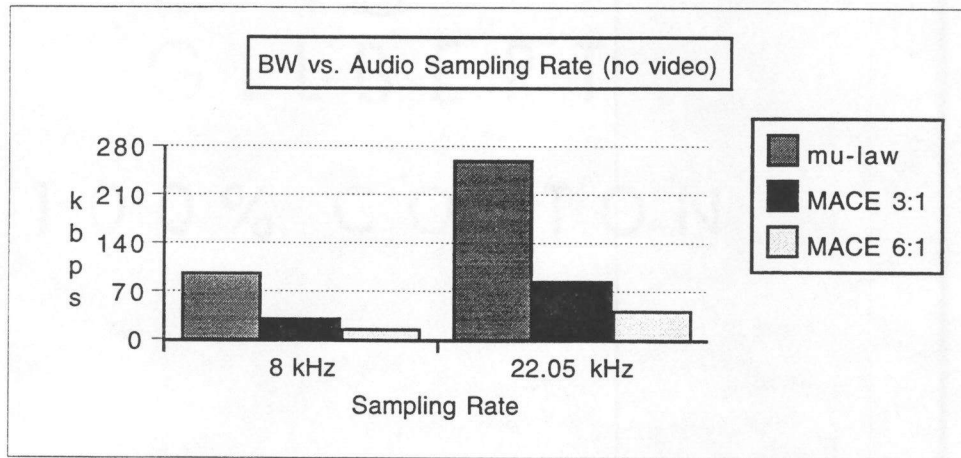


Fig. 6.8: Plots of the average bandwidth of Table 6.6.

By comparing the results obtained from this topology with those from the previous one, it is necessary to make two remarks. First, bandwidth utilization among various compressors and under the same conditions is still consistent with the previous results. For example, H.261 still has the lowest bit rate among other video compressors, and MACE 6:1 among audio compressors. Second, the bit rate is lower in this case due to more traffic and packet loss. As a

result, quality was degraded and the delay time between sending and receiving nodes increased.

6.3 Third Topology

This topology represents a WAN connection between the UTK campus and a remote location. Figure 6.9 shows a schematic representation of such a connection. The measurements were taken for this section were the same as before. Bandwidth utilization results for this section are listed in Tables 6.7 - 6.9 and plots of the average data in Figures 6.10 - 6.12, respectively.

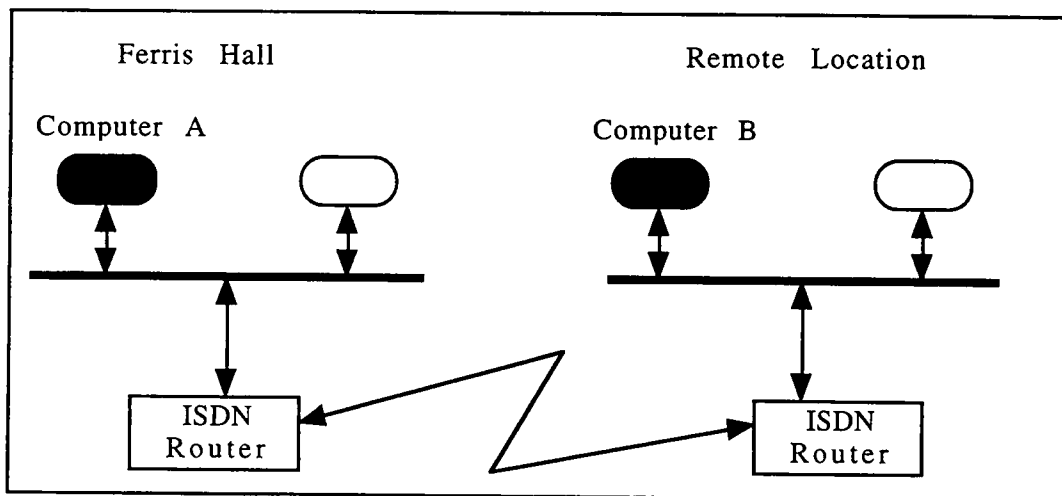


Fig. 6.9: Schematic representation of the third topology.

Table 6.7: Bandwidth measurements for video compressors with μ -law audio compressor.

Video compressor	Quality	High BW (kbps)	Low BW (kbps)	Average BW (kbps)
H.261	Low	155.56	143.04	147.88
	Medium	160.82	151.12	156.32
	High	154.82	146.21	152.51
JPEG	Low	150.28	145.3	147.54
	Medium	311.94	285.08	293.66
	High	390.94	372.1	386.51
Apple Video	Low	358.58	344.39	351.62
	Medium	442.89	433.34	438.88
	High	585.7	566.18	577

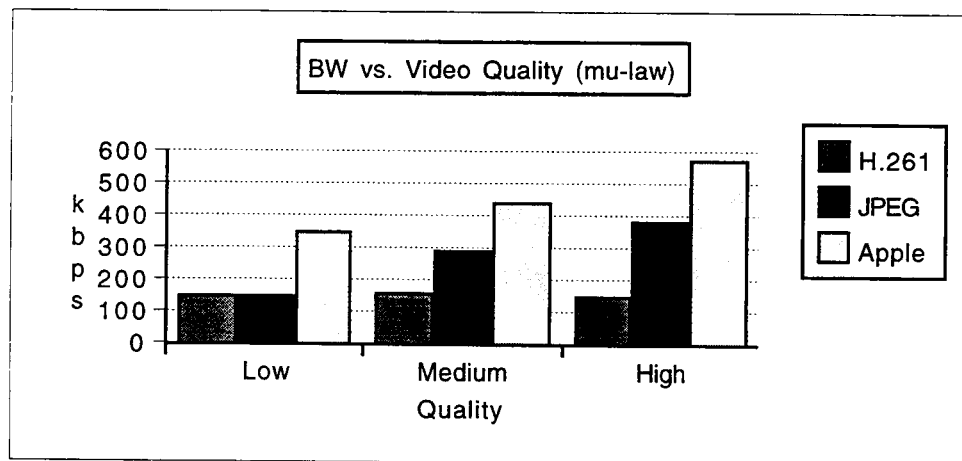


Fig 6.10: Plots of the average bandwidth of Table 6.7.

Table 6.8: Bandwidth measurements for audio compressors with H.261 video compressor.

Audio compressor	Sampling rate (kHz)	High BW (kbps)	Low BW (kbps)	Average BW (kbps)
mu-law	8	231.18	157.36	209.12
	22.05	368.68	349.34	358.3
MACE 3:1	8	170.08	155.83	161.75
	22.05	214	199.44	208.72
MACE 6:1	8	139.08	102.83	120.13
	22.05	178.66	159.81	168.6

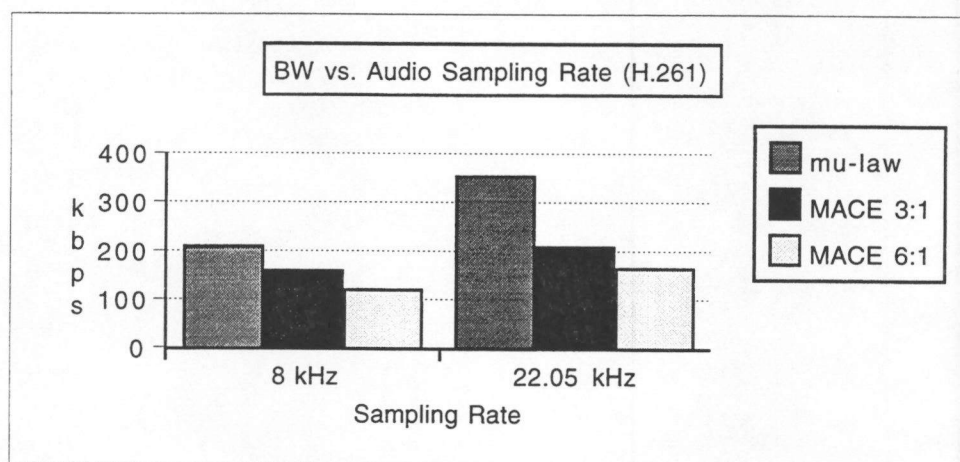


Fig 6.11: Plots of the average bandwidth of Table 6.8.

Table 6.9: Bandwidth measurements for audio compressors with no video transmission.

Audio compressor	Sampling rate (kHz)	High BW (kbps)	Low BW (kbps)	Average BW (kbps)
mu-law	8	80.57	79.1	79.4
	22.05	220.25	219.22	220.54
MACE 3:1	8	26.9	26.36	26.7
	22.05	73.34	73.3	73.32
MACE 6:1	8	13.39	13.21	13.3
	22.05	36.67	36.41	36.61

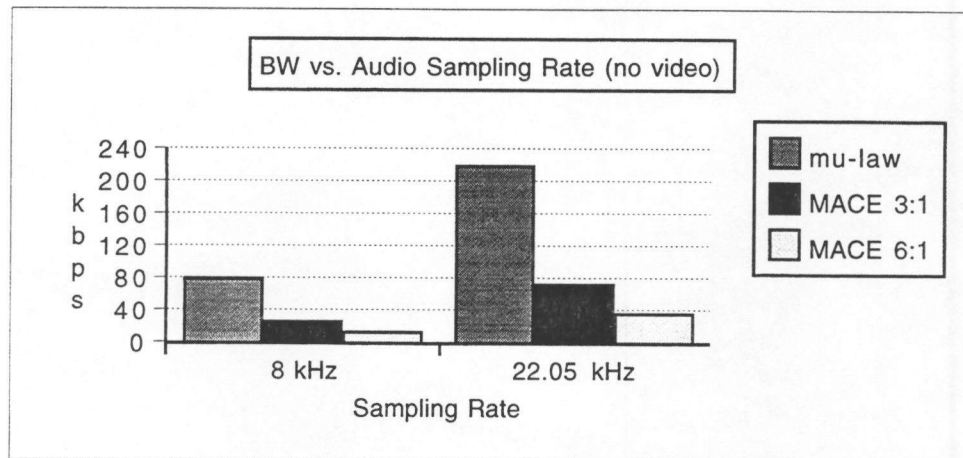


Fig 6.12: Plots of the average bandwidth of Table 6.9.

The following chapter will present a conclusive remarks on the results obtained and offer solutions for enhancing the performance of videoconferencing in the future.

Chapter 7

Summary and Recommendations

Desktop videoconferencing is a new technology that will play an important role in the future of communications bringing real-time video in addition to speech and the ability to share and collaborate on documents. This will certainly have positive results whether for a corporate meeting around the world, or two students in different labs on the same collage campus. The question to be asked is will this technology mature enough to deliver what it promises? This question was partially answered in this thesis project by discussing the important technical aspects of each of the desktop videoconferencing elements and testing some of them.

This work addressed many of the issues of desktop videoconferencing including image acquisition and encoding, audio and video compression, and transport media suitable for this application. It also includes an evaluation test for one of the videoconferencing commercial products. As was seen, raw audio and video signals demand considerable bandwidth and are computational complex to compress. Several compression standards were discussed for this task as well as their importance in improving

videoconferencing. Also different communication channels and transport protocols were discussed. This included circuit-switched channels such as ISDN and ADSL which offer dedicated bandwidth and predictable timing for data delivery, and packet-switched channels which are the basis for LAN operation on a shared network. B-ISDN and ATM show a promising future for solving many of the problems encountered today such as speed and reliable delivery, and developing standards for interoperability.

Although most of the infrastructure for desktop videoconferencing is available on a campus network, upgrading it to meet the increased needs is still costly. ATM switches are still expensive and standards for deploying them are still under development. As a result, videoconferencing over existing networks will continue to be delivered with average quality until further development.

Some of the tested compression standards such as H.261 and MACE achieved better bit rate results, while others such as Apple video achieved better image qualities, but with higher bit rates. This means that network-limited resources could either be used to transport more frames at lower quality or the opposite. The problem which was encountered during testing was the difficulty of maintaining good and synchronized audio and video quality with a relatively acceptable frame rate.

Future work for enhancing desktop videoconferencing should further address the issues of increasing channel capacities and improved methods for bit rate reduction. By addressing these issues, in addition to implementing international standards, desktop videoconferencing in the future will be a part of the human daily activities much as television and telephone are now.

REFERENCES

References

- [1] M. Macedonia and D. Burtzman, "MBone provides audio and video across the Internet," *Computer*, Vol. 27 No. 4, Apr. 1994, pp. 30-46.
- [2] R. Aravind et al., "Image and Video Coding Standards," *AT&T Technical Journal*, Vol 72 No. 1, Jan/Feb 1993, pp.67-89.
- [3] D. LeGall, "MPEG: A Video Compression Standard for Multimedia Applications," *Communications of the ACM*, Vol. 34 No. 4, April 1991, pp. 60-63.
- [4] Apple Computers, Inc., "QTC Product Information," WWW page, March 1996,
<<http://qtc.quicktime.apple.com/qtc/qtc.products.html>>.
- [5] M. Steenstrup, "Routing in Communications Network," Prentice-Hall, Inc., 1995.
- [6] W. Stallins, "ISDN An Introduction," Macmillan Publishing Co., 1989.
- [7] M. Smouts, "Packet Switching Evolution From Narrowband to ISDN," Artech House, 1992.
- [8] A. Girard, "Routing and Dimensioning in Circuit-switched Networks," Addison-Wesley Publication Co., 1990.
- [9] R. Gibbens et al., "Dynamic Routing in Multiparented Networks," *IEEE/ACM Transaction Networking*, Vol. 1 No. 2, 1993, pp. 261-270.
- [10] G. Fantauzzi, "Digital Switching Control Architectures," Artech House, Inc., 1990.

- [11] G. Ash et al., "Design and optimization of Networks with Dynamic Routing," *BellSouth Technical Journal*, Vol. 60, 1981, pp. 1787-1820.
- [12] P. Green, Jr., "Computer Communications: Milstones and Prophecies," *IEEE Commun. Mag.*, Vol. 22, Apr. 1984, pp. 24-31.
- [13] S. Schlar, "Inside X.25: A Maneger's Guide," McGraw-Hill, Inc., 1990.
- [14] S. Keiser, "Local Area Networks," McGraw-Hill, Inc., 1989.
- [15] J. Turner, "New Directions in Communications," *IEEE Commun. Mag.*, Vol 24, Oct. 1986, pp. 8-15.
- [16] K. Hewitt, "Desktop videoconferencing products," Sep. 1996, <<http://www3.ncsu.edu/dox/video/products.html>>
- [17] H. Hang, "Handbook of Visual Communications," Academic Press, Inc., 1995.
- [18] J. Allebach, "Processing Digital Color Images: From Capture to Display," *Physics Today*, Vol. 45, Dec. 1992. pp. 32-39.
- [19] R. Khosla, "From Photons to Bits," *Physics Today*, Vol. 45, Dec. 1992. pp. 42-49.
- [20] Y. Faroudja, "Improving NTSC to Achieve Near-RGB Performance," *SMPTE Journal*, Aug. 1987. pp. 750-761.
- [21] J. Blinn, "NTSC: Nice Technology, Super Color," *IEEE COmputer Graphics and Applications*, March 1993, pp. 17-23.
- [22] J. Blinn, "The World of Digital Video," *IEEE COmputer Graphics and Applications*, Sep. 1992, pp. 106-111.
- [23] J. Beynon, "Charge-Coupled Devices and Their Applications," McGraw-Hill, Inc., 1980.

- [24] A. Netravali, "Digital Pictures: Representation, Compression, and Standards," Plenum Press, 1995.
- [25] "Digital Compression and Coding of Continuous Tone Still Images, Part 1: Requirements and Guidelines," ISO/IEC IS 10918-1, 1991.
- [26] "Digital Compression and Coding of Continuous Tone Still Images, Part 2: Compliance Testing," ISO/IEC IS 10918-1, 1991.
- [27] ITU-T Recommendation H.320, "Narrowband visual telephone systems and terminal equipments," ITU, 1991.
- [28] ANSI/IEEE Std. 803.2, "Carrier Sense Multiple Access with Collision Detection (CSMA/CD)," 1985.
- [29] J. Enck, "LAN to WAN Interconnection," McGraw-Hill, 1995.
- [30] W. Goralski, "Introduction to ATM Networking," McGraw-Hill, 1995.
- [31] A. Alles, "ATM Internetworking," May 1995,
<<http://cio.cisco.com/warp/public/614/12.html>>
- [32] ITU-T (formerly CCITT) Study Group XVIII I Series Recommendations, Oct. 1984.
- [33] C. Hoppit, "ISDN Evolution: From Copper to Fiber in Easy Stages," *IEEE Commun. Mag.*, Vol. 24, pp.17-22, Nov. 1986.
- [34] ITU-T, Recommendation Q.921, "ISDN User-network Interface Data Link Layer Specification," Vol. VI.10, ITU 1989.
- [35] ITU-T, Recommendation Q.931, "ISDN User-network Interface Layer 3 Specification for Basic Call Control," Vol. VI.11, ITU 1989.
- [36] ITU-T, Recommendation I.430, "Basic User-network Interface - Layer 1 specification," Vol. III.8, ITU 1989.

- [37] ITU-T, Recommendation I.150, "B-ISDN Asynchronous Transfer Mode Functional Characteristics," ITU 1995.
- [38] K. Maxwell, "Asymmetric Digital Subscriber Line," *IEEE Commun. Mag.*, Vol. 34 No. 10, Oct. 1996, pp. 102-109.
- [39] Intel, "Intel ProShare Videoconferencing," WWW page, <<http://www.intel.com/comm-net/proshare>>.
- [40] S. Deering, "Host Extensions for IP Multicasting," Internet Request For Comment 1112, August 1989, <<http://ds.internic.net/rfc/rfc1112.txt>>.
- [41] S. Deering, "IP Multicast Extensions for 4.3 BSD UNIX and Related Systems," MULTICAST 1.2 Release, 1989, <<file://gregorio.stanford.edu/umtpp/ipmuticast>.README.

VITA

Raymond M. El-Bitar was born in Baalbek, Lebanon on January 24, 1969. He attended Notre Dame elementary school in Beirut and graduated from Pine Hill College after earning the Lebanese Baccalaureate in 1986. He joined the Lebanese-American University (formerly known as Beirut University College) in the spring of 1989 for a dual degree program. In the summer of 1993, he entered Georgia Institute of Technology and graduated in fall of 1994 with a Bachelor of Electrical Engineering. In June 1995, he entered the University of Tennessee, Knoxville to pursue a Master of Science with a major in Electrical Engineering which he is expecting to receive in May 1996.