Digital Video Communications

Martyn J. Riley Iain E.G. Richardson

Digital Video Communications

Digital Video Communications

Martyn J. Riley
Iain E. G. Richardson

Artech House
Boston • London

Snap - Exhibit 1010
Snap v. Nokia - IPR2025-01439
Page 3 of 78

Library of Congress Cataloging-in-Publication Data
Riley, Martyn J.
Digital video communications / Martyn J. Riley, Iain E. G. Richardson.
p. cm.
Includes bibliographical references and index.
ISBN 0-89006-890-9
1. Digital television. I. Richardson, Iain E. G. II. Title.
TK6678.R55 1996
621.388—dc21 96-37127
CIP

British Library Cataloguing in Publication Data
Riley, Martyn J.
Digital video communications
1. Digital communications 2. Digital video 3. Image transmission 4. Image processing—Digital technique
I. Title II. Richardson, Iain E. G.
621.3'87

ISBN 0-89006-890-9

Cover design by Jennifer Makower

© 1997 ARTECH HOUSE, INC.
685 Canton Street
Norwood, MA 02062

All rights reserved. Printed and bound in the United States of America. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Artech House cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

International Standard Book Number: 0-89006-890-9
Library of Congress Catalog Card Number: 96-37127

10 9 8 7 6 5 4 3 2 1

To Laura and Phyllis
and in memory of Jim Riley

Contents

Chapter 1 Introduction 1
- 1.1 Digital Video Communications 1
- 1.2 The Market 3
- 1.3 Key Issues 4
- 1.4 Structure of This Book 5
- Reference 6

Chapter 2 Applications for Digital Video Communications 7
- 2.1 Introduction 7
- 2.2 Entertainment Broadcasting 7
- 2.3 Video on Demand 8
- 2.4 Conferencing 9
  - 2.4.1 Telemedicine 9
  - 2.4.2 Videoconferencing and Education 13
- 2.5 Summary 14
- References 14

Chapter 3 Video Compression Techniques and Standards 15
- 3.1 Introduction 15
- 3.2 Digital Video 16
  - 3.2.1 Representation of Video Information 16
  - 3.2.2 CCIR 601 Standard 17
  - 3.2.3 The Need for Compression 18
- 3.3 Coding of Still Images 19
  - 3.3.1 Predictive Coding 20
  - 3.3.2 Discrete Cosine Transform Coding 21
  - 3.3.3 Other Image Coding Techniques 25
- 3.4 Coding of Moving Images 26
  - 3.4.1 "Generic" DCT/DPCM CODEC 28
  - 3.4.2 Bidirectional Prediction 28

Snap - Exhibit 1010
Snap v. Nokia - IPR2025-01439
Page 6 of 78

3.5 JPEG: Still Image Coding 30
3.5.1 The Baseline CODEC 31
3.5.2 Other Features and Modes 34
3.6 H.261: Motion Video Coding for Videoconferencing 34
3.7 H.263: Low Bit Rate Video Coding 37
3.8 MPEG: Motion Video Coding for Entertainment and Broadcast 38
3.8.1 MPEG1 for CD Storage 39
3.8.2 MPEG2 for Broadcasting 42
3.8.3 MPEG4 for Integrated Visual Communications 46
3.9 Digital Audio Visual Council (DAVIC) 47
3.10 Future Developments and Research Areas 47
3.10.1 Optimization and Application of Existing Standards 48
3.11 Summary 49
References 50

Chapter 4 The Internet 53
4.1 Introduction 53
4.2 The Internet Architecture 54
4.2.1 Gateways or Routers 54
4.2.2 Internet Addresses 55
4.2.3 The Domain Name System 56
4.3 Internet Protocols 57
4.3.1 Internetworking Protocol (IP) 58
4.3.2 Mapping Internet Addresses to Physical Addresses 59
4.3.3 TCP and UDP 61
4.4 Congestion in the Internet 64
4.5 The Internet and Real-Time Traffic 64
4.5.1 M-BONE (IP Multicast) 66
4.5.2 RSVP 67
4.6 Summary 68
References 68

Chapter 5 B-ISDN and Asynchronous Transfer Mode Networks 71
5.1 Introduction 71
5.2 Principles of the Asynchronous Transfer Mode 72
5.2.1 Virtual Circuit Operation 75
5.2.2 Statistical Multiplexing 77
5.3 ATM Protocol Architecture 78
5.3.1 Physical Layer 79
5.3.2 ATM Layer 79
5.3.3 ATM Adaptation Layer (AAL) 80

Snap - Exhibit 1010
Snap v. Nokia - IPR2025-01439
Page 7 of 78

5.4 ATM Switching Technology 82
5.5 Congestion Control 84
5.6 ATM Signaling 86
5.7 UPC Functions 86
5.8 LAN Emulation 87
5.9 Video Over ATM 87
5.10 Example ATM Networks 89
5.11 Summary 89
References 90

Chapter 6 Quality of Service for Video Communications 91
6.1 Introduction 91
6.2 Video Quality Requirements 93
6.2.1 Image Quality 93
6.2.2 Visible Errors 93
6.2.3 Delay 94
6.3 Quality of Service Requirements for a Video Transport System 94
6.4 Quality of Service for Coded Video 95
6.4.1 Data Transmission Rate 95
6.4.2 Errors and Losses 97
6.4.3 Delay 98
6.5 Multiplexing and Demultiplexing Traffic Streams 100
6.5.1 MPEG2 Systems Layer 100
6.5.2 Transmission of MPEG2 Transport Packets Over ATM 104
6.6 Summary 105
References 105

Chapter 7 The Effect of Bit Errors and Packet Loss on Compressed Video 107
7.1 Introduction 107
7.2 Sources of Errors 108
7.2.1 Bit Errors 108
7.2.2 Packet Loss 108
7.3 Effect of Errors on Coded Video 109
7.3.1 Error Effects and Error Propagation 109
7.3.2 Error Effects: Experimental Results 115
7.4 Coding Techniques to Reduce the Effect of Errors 123
7.4.1 Error Correction 123
7.4.2 Error Concealment 124
7.4.3 Encoding Parameters 125

7.4.4 Case Study: Varying Slice Size to Improve Error Tolerance of MPEG Video 126
7.5 Summary 128
References 128

Chapter 8 Forward Error Correction 131
8.1 Introduction 131
8.2 Some Definitions 132
8.3 Finite (or Galois) Fields 134
8.4 Linear Block Coding 136
8.4.1 Hamming Codes 138
8.4.2 Bose, Chaudhuri Hocquenghem (BCH) Codes 139
8.4.3 Reed-Solomon Codes 141
8.4.4 Fast Implementation Techniques for Fourier Transforms in Galois Fields 148
8.5 Convolutional Coding 148
8.6 Interleaving 151
8.7 Applications of FEC 152
8.7.1 Broadcasting 152
8.7.2 Packet Loss Recovery 153
8.8 Summary 154
References 155

Chapter 9 Layered Video Coding 157
9.1 Introduction 157
9.2 Layered Coding in International Standards 160
9.2.1 SNR Scalability 160
9.2.2 Spatial Scalability 162
9.2.3 Temporal Scalability 164
9.2.4 Frequency Scalability 166
9.3 Research Issues 169
9.3.1 Filtering in the Compressed Domain 169
9.3.2 Case Study: Prioritizing MPEG Picture Classes 171
9.4 Summary 172
References 172

Chapter 10 Rate Control 175
10.1 Introduction 175
10.2 Rate Control Requirements 178
10.2.1 Constant Bit Rate Transmission/Storage 178
10.2.2 Variable Bit Rate Transmission 179
10.3 Effect of Encoding Parameters on Coded Bit Rate 181
10.3.1 Spatial and Temporal Resolution 182
10.3.2 Quantization Step Size 182

10.3.3 Group of Pictures Structure (MPEG) 184
10.3.4 Slice Size (MPEG) 185
10.4 Rate Control Techniques 185
10.4.1 Constant Bit Rate 186
10.4.2 Variable Bit Rate 192
10.5 Summary 195
References 195

Chapter 11 Conclusions 197
11.1 Applications and Requirements 197
11.2 Meeting Video Quality of Service Requirements 198
11.2.1 Encoding 198
11.2.2 Protocol Processing 198
11.2.3 Transmission 199
11.2.4 Decoding 200
11.3 Future Developments 200

Glossary 203

About the Authors 207

Index 209

Introduction 1

1.1 DIGITAL VIDEO COMMUNICATIONS

Digital video has emerged in the last few years as a technology that can provide a new "dimension" to electronic communications. *Digital video* is video information represented in digital form. Digital representation has a number of key advantages over "traditional" analog video and television. All information can be represented in digital form, so the same techniques and systems can be used to store, process, and transmit a wide range of different types of data (multiple media or "multimedia"). The rapid growth in digital processing power means that complex processing and coding operations can be carried out on digital video data in real time. Digital video can be integrated into computer applications and systems. This in turn makes it possible to create interactive applications, where the user is no longer a "passive" observer but has the opportunity to interact with the video information.

Advances in computing and processing capabilities have been matched by advances in data networking technology. In a short space of time, the number of computers and systems connected by networks such as the Internet has grown exponentially. As well as becoming more widespread, networks can handle higher volumes of data and higher transmission rates. The current networking structure is loosely defined and "heterogeneous" (consisting of a range of interconnected networks with different technologies and capabilities). Broadband networks based on the asynchronous transfer mode, in which all data is processed and transmitted as small, fixed-length cells, can support efficient transmission of high data rates, and these networks are beginning to replace older, less efficient technologies.

As we shall describe in Chapter 3, digital video has an inherently high bandwidth (i.e., a digitized video signal requires a very high data rate for transmission). In order to store and transmit this information effectively, it has been necessary to develop techniques for compressing the video data (i.e.,

Snap - Exhibit 1010
Snap v. Nokia - IPR2025-01439
Page 11 of 78

encoding it into a smaller number of bits). The emergence of international standards for encoding video data has enabled a wide range of applications that make use of digital video transmission and storage. Image coding provides a means of compressing digitized photographic images by around 10 to 20 times. Current video coding techniques enable video data to be compressed by between 20 and 50 times. By using these techniques, it has become practical to store, transmit, and manipulate digital image and video information using currently available storage systems and data networks. The range of applications for digital video continues to grow and includes the following:

- *Video conferencing and videotelephony.* Two way, real-time video and audio communications. Video and audio information is transmitted via data networks such as integrated services digital network (ISDN) and the Internet (IP-based network). Implementations range from one-to-one videotelephony to studio-based group conferencing.
- *Home entertainment digital video.* The video compact disc (based on the CD-ROM format) provides a digital alternative to analog home video. Compact disc digital video brings many of the advantages of audio compact discs (e.g., durability and random access) to video entertainment, together with further features such as user interaction and interactive video games.
- *Broadcast digital television.* Digital television channels are beginning to replace analog transmissions over cable, satellite, and terrestrial broadcast media. This should provide (arguably) greater choice and picture quality to the viewer together with further advantages and features associated with digital transmission.
- *Video databases and video on demand.* These applications involve the storage of digital video information in large databases. The user can access the stored information and play back the video data. Application areas include home selection and viewing of centrally stored films and programs ("video on demand") as well as archival and access of educational learning video material.
- *Medical applications.* Many of the above techniques can be integrated into the healthcare system. Experimental trials to date have made use of videoconferencing between medical practitioners as well as medical image communication and storage using electronic media.

These applications are all in use at the time of writing. The continual improvements in processing power and network capacities should lead to a wider penetration of digital video into the marketplace. Digital video communication has the capability to change dramatically the way we communicate and do business.

Many more potential applications have been identified that make use of digital video techniques. These include high definition digital television (HDTV), which provides higher spatial and temporal resolutions than existing analog (or digital) television, applications that involve much more interaction from the user (for example, where the video information is manipulated as individual components or objects within a scene) and stereoscopic (three-dimensional) or multiple viewpoint video. Advanced encoding techniques may offer very high compression of video information that could enable high-quality video at low bit rates and should lead to further availability of digital video applications.

1.2 THE MARKET

The rapidly expanding digital video market has been studied in a number of marketing reports. It is widely predicted that in many of the application areas, growth will be extremely rapid over the next few years.

For example, videoconferencing has been identified as one of the major application areas. This can be divided into systems based in specially equipped rooms (referred to as suites) and desktop conferencing. Desktop videoconferencing is carried out from a user's desktop with a specially equipped personal computer (PC). With regard to suites, it has been stated [1] that

> Research commissioned by GPT (and others) estimates that global sales of small and large group [videoconferencing] systems reached 14,000 units and generated a revenue of [US]$621m in 1994. This year [1995] those figures are expected to rise to 20,000 units and [US]$819m and in 1998 it will be 40,000 units and [US]$1.345 Bn.

The situation for desktop systems is considered [1] to be even more promising:

> This steady growth contrasts with the data provided for PC-based desktop systems where 1994 figures suggest sales of 12,500 units and revenue of [US]$31m and 1995 is expected to generate sales of 86,000 units and [US]$120m in revenue. Then, according to the research, the market will soar with 987,000 units likely to be shipped and revenue leaping to [US]$789m.

According to [1], this explosion in revenue will lead to a 1998 market of 5.225m units and revenue of over (U.S.)$2 billion.

This growth is likely to create turmoil in a number of industries since digital video technology involves a combination of telecommunications, televi-

sion, and computer engineering. Companies are being required to develop expertise in new areas and form strategic partnerships to exploit the rapidly developing market. In this book, we aim to provide an introduction to a range of the relevant technologies and then consider some of the key technical issues.

1.3 KEY ISSUES

Many of the applications for digital video that are currently in use have been designed for fixed network topologies and systems. For example, videoconferencing over ISDNs can rely on a fixed transmission bit rate and a well-defined transmission delay and error rate. If these parameters can be accurately predicted, then a video coding system can be designed that provides acceptable performance in this configuration. However, the ever-changing networking infrastructure is made up of an increasingly heterogeneous mix of networks and technologies. The transmission characteristics provided to a particular connection (the *quality of service*) can vary from connection to connection and even during a connection, depending on the path taken through the network and on the other traffic currently in transit through the network.

A communication network offers a particular quality of service (QoS) to traffic transmitted through the network. Important QoS parameters include data rate (mean data rate and rate variation), delay characteristics (transmission delay, delay variation), and error or loss probability (bit error rate, cell or packet loss probability, error patterns).

All of these parameters can affect the quality and reliability of digital video communications. The available data rate has a significant effect on the quality and resolution of coded video information that can be transmitted through the network. Different coding techniques are appropriate for different data rates. Transmission delay and delay variation are very important since real-time video applications are sensitive to changes in transmission delay. Decoded video frames must be presented to the viewer at a constant rate, and if a particular frame is delayed too long then it cannot be displayed and is "lost." Errors and losses have a particularly severe effect on the quality of decoded video. Video coding schemes achieve compression by removing redundant information from the video data, which means that all of the coded data is significant and hence a transmission error is likely to have a significant impact on video quality.

Some of these problems are addressed by the video coding standards themselves. However, many of the implementation details are left to be addressed by the developer of a particular application. The video coding standards have intentionally been kept as broad as is practical. There is therefore considerable scope for optimization of video coding techniques in order to address these problems and provide reliable video communications.

Ideally, a video communication system should provide the user with a service that offers video at an acceptable quality. The networking environment is not likely to "stabilize" in the foreseeable future and so a practical video communication system should be adaptable, reliable, and robust in order to cope with varying QoS from the network. A further useful feature is scalability: a flexible video communication system should be able to scale the quality of the video depending on the capabilities of the network and of the display.

1.4 STRUCTURE OF THIS BOOK

This book aims to introduce the concepts and systems required to support digital video communications and to discuss some of the issues that need to be addressed to provide reliable, flexible video communications in the changing networking environment. The intended audience for the book includes:

- Practicing engineers and technical staff who have an involvement in the video, television, computing, and communications industries;
- Academics and researchers in electronic engineering, computer science, and related fields who have a research or teaching interest in digital video communications;
- Final year undergraduate and postgraduate students in electronic engineering, computer science, and related fields who require an introduction to the concepts and issues involved in digital video communications.

In the early chapters, we describe the applications for digital video and introduce some of the key technologies that are required to support digital video communications.

Chapter 2 describes application areas in which digital video is used, including those areas mentioned in Section 1.1.

Chapter 3 introduces the techniques for encoding (compressing) and decoding digital video data. In this chapter, we concentrate in particular on the international standards for video encoding.

Chapters 4 and *5* provide an introduction to key network technologies that are being used for digital video transmission. In Chapter 4 we describe the structure and protocols of the Internet and in Chapter 5 we introduce asynchronous transfer mode (ATM) networks and broadband ISDN.

Chapter 6 is a pivotal chapter. Here, we discuss the QoS issues related to the transmission of coded digital video through networks. These QoS requirements must be met in order to provide flexible, reliable, and robust video transmission.

The later chapters concentrate on a series of issues related to supporting the QoS required for digital video communications and are based on research work by the authors and others.

Chapter 7 analyzes the effect of transmission errors on coded video data. Any practical transmission system or network will introduce errors and losses into transmitted data. We discuss the problems that this can cause for video quality and look at ways of improving quality in the presence of errors.

Chapter 8 deals with forward error correction coding and its application to error control for digital video communications. Forward error correction is particularly suited to real-time information such as digital video and we introduce the fundamental techniques and some of the applications for video communications.

In *Chapter 9* we discuss scalable or layered video coding. We describe some of the main layered coding techniques and how they might be applied to improve the QoS available to coded video data.

Chapter 10 introduces the topic of rate control for coded video. Practical networks place restrictions on the rate of data that can be transmitted and for this reason rate control is a key part of any video coding system.

Digital video communications offer considerable scope to change the way we communicate. Providing practical and robust digital video communications systems requires an understanding of coding and transmission techniques as well as an appreciation of how video QoS requirements may be supported. We hope that this book goes some way towards meeting these needs.

Reference

[1] "Bringing Video Conferencing into Line," *Audio-Visual Communications for Business Magazine*, May 1995.

Applications for Digital Video Communications

2

2.1 INTRODUCTION

Analog video communications technology has significantly affected the lives of millions of people worldwide. The vast majority of households in the developed world own at least one television set, which provides a major source of entertainment to the home.

Digital video offers many advantages over analog video for the vast majority of applications. Therefore, as digital video technology becomes more mature we would expect to see a gradual shift from analog to digital technology. The ease with which digital video can be integrated with computers, together with the introduction of global communication networks that support the integration of video with audio and conventional computer data make a whole range of new applications possible.

Many people are now familiar with personal computers (PCs) both at home and in their working environment. These PCs are now being produced with support for multimedia, including audio and video. The costs of this multimedia support are dropping rapidly. The increasing power of the microprocessors in these PCs makes more and more processing possible in software, thereby reducing the need for specialist hardware. Multimedia PCs represent a powerful platform and motivating factor in the development of digital video technology.

In this chapter, we provide a discussion of some of the applications of digital video communications and concentrate on two case studies: one relating to the use of digital video to support telemedicine and the other to support education and distance learning.

2.2 ENTERTAINMENT BROADCASTING

It is expected that analog television technology will gradually be replaced with digital technology. Entertainment television may be delivered to the home using

Snap - Exhibit 1010
Snap v. Nokia - IPR2025-01439
Page 17 of 78

terrestrial broadcasting, fixed cable networks, and/or satellite broadcasting. Standards [1–3] have been defined for digital television broadcasting. Digital television offers a number of advantages over analog television. Many fixed cable networks are already digital and provide a whole range of digital communication services such as Internet access, home banking, and interactive shopping in addition to entertainment television. Digital compression techniques combined with advanced modulation allow very efficient utilization of bandwidth for broadcasting. This is extremely important when the radio spectrum is very congested. Four to five digital channels of comparable quality can be broadcast with the same bandwidth as one analog channel. Digital television facilitates the integration of television with the home computer system, thereby providing the user with much greater flexibility. Digital video can be stored on computer disks at a wide range of quality levels and can be used as an integral part of a computer software application.

The replacement of analog with digital technology is likely to be an evolutionary one where in the first instance analog and digital systems coexist. The first users of digital television technology will buy a set-top box, which will demodulate and decode the digital video signal. This will be converted to the equivalent analog signal and fed into a standard analog television. Eventually this technology will be integrated into a single consumer product, possibly also including the home computer.

2.3 VIDEO ON DEMAND

Video on demand allows a user to request the video that they want to watch and have the requested program supplied to them. In some senses, the video rental shop provides a very slow video on demand service. Fixed cable networks provide the communications infrastructure to allow a user to select a video of interest electronically and have that video transmitted to them. Video on demand can be implemented in many different ways. The Internet does not yet reliably support real-time video traffic. In Chapter 4, we discuss how the Internet is evolving to become a suitable communications infrastructure for providing video on demand. Many cable television companies have communication infrastructures capable of two-way communication and therefore are suitable for supporting video on demand. A video server consisting of a high-speed disk subsystem would be used for storing the video material. In Chapter 5, we describe an example of a video on demand system that is being developed to deliver video-based computer-assisted learning material across a metropolitan area ATM network (ABMAN).

2.4 CONFERENCING

Conferencing applications are concerned with real-time two-way communication and are designed to replace face to face communication. In order to provide effective interactive communication it is necessary to limit the end-to-end delay. For delays greater than about 300 ms, interaction between people becomes difficult. Traditionally, videoconferencing has been performed using videoconferencing suites. These are rooms specially equipped with videoconferencing equipment. Traditionally, this videoconferencing equipment has been expensive and only available to large multinational organizations.

Low-cost desktop videoconferencing systems are now emerging. These systems are usually integrated with a PC and consist of a small video camera and hardware- or software-based video compression/decompression adapters (CODECs). Conference participants are viewed in one or more small video windows on the PC screen. In addition, these systems usually support audio communication and data interchange. Increases in PC processing power are enabling high-performance video CODECs to be implemented in software, causing the cost of the systems to fall rapidly.

Videoconferencing can be used effectively in a number of different application areas, usually with the aim of avoiding the inconvenience and/or cost of traveling. In the following sections we consider how videoconferencing is useful in the fields of telemedicine and distance learning.

2.4.1 Telemedicine

Telemedicine can be defined as medicine practiced when the patient is at a site that is at some distance from the doctor or medical expert. The practice of telemedicine makes use of telecommunications and information technology, including digital video communications, to provide this healthcare at a distance.

In many areas of the world, major hospitals are a considerable distance apart. Healthcare is provided in the local community by a local health provider, which may be a smaller community-based health center or general practice hospital. Although these local centers can provide effective healthcare for many medical problems, there is frequently a need for patients with more specialized problems to be referred to a major hospital. The costs required for this patient transfer can be considerable and may involve transporting the patient by ambulance, helicopter, or airplane and at least an overnight stay at the major hospital. In many cases, this patient transfer could be avoided if sufficient details of the patient's conditions could be communicated to the consultants at the major hospital. Following a remote consultation with the general practitioners at the local center, it may be possible to decide on a course of treatment that can be administered locally. To make this procedure effective, it is necessary to

communicate sufficient patient details. Following this communication of patient details, it is also necessary to provide high-quality communication between the consultant at the major hospital and the local health expert.

Patient information may include medical records stored on a computer database, digital images from patient scans or x-ray images, vital signs such as electrocardiograms, and either recorded or live video images of the patient. Live video images of the patient may provide the consultant with the ability to carry out a remote examination of the patient with the aid of the local health expert. This technique is referred to as telepresence.

In order to make such a system work well, it is necessary to effectively integrate many separate pieces of equipment together and communicate the data from this equipment across the wide area. This requires that multimedia data is communicated over a communication network that supports the transmission of integrated services. In many situations the most challenging part of implementing this communications is the support of real-time two-way videoconferencing. The video requires large bandwidths even when compressed, which may still be significantly higher than many of the other data types. Two-way conferencing also requires that end-to-end delays are minimized to provide effective interaction across the communication link. It is only in recent years that the communication systems capable of supporting this integration of services and the necessary digital video compression techniques have become available at practical costs.

An example of a project concerned with implementing a telemedicine application is the Study of Video Image Transfer: Orthopedics Up to Rehabilitation (SAVIOUR) project [4] that commenced in 1993. The project is funded as part of the European Union (EU) Advanced Informatics in Medicine (AIM) program. The SAVIOUR project is directed by the research unit of RGIT Ltd. (based in Aberdeen, Scotland) and the other project partners are The Robert Gordon University, Aberdeen Royal Hospitals Trust, and ABB-NERA Ltd.

One of the purposes of this project is to establish the practice of telemedicine between a local general practice hospital in Peterhead, which is situated 40 miles north of Aberdeen, and the Accident and Emergency Department of Aberdeen Royal Hospitals Trust. The project aims to largely use existing technology to perform a clinical trial and in particular to investigate the interface between the telemedicine equipment and the clinicians rather than develop new technology. Two alternative communication systems are available for use during the project. These are a basic rate Integrated Services Digital Network (ISDN) connection providing 128 Kbps and a satellite connection providing 64 Kbps. In addition, the clinicians can communicate over the public switched telephone network (PSTN). ISDN connections are widely available in the North of Scotland, but there are many other locations where this is not the case and the purpose of the satellite connection is to allow evaluation of the system for

Snap - Exhibit 1010
Snap v. Nokia - IPR2025-01439
Page 20 of 78

subsequent possible use in these situations. For example, the Health Division of RGIT Ltd. specializes in providing healthcare to remote and inhospitable locations, including a medical service to the British Antarctic Survey and to the British sector of the North Sea oil fields. In such situations, remote healthcare may be extremely beneficial since patient evacuation is very costly or even impossible.

A block diagram of the SAVIOUR system is shown in Figure 2.1. British Telecom VC-7000 videotelephone units are located at both of the sites to provide a videoconferencing capability. This provides the ability for "face to face" consulting. The videophones implement the ITU-TS H.261 videoconferencing standard, which compresses the video to a data rate of either 64 Kbps or 128 Kbps. We discuss this and other video compression techniques in Chapter 3. The quality of the video is acceptable, but significantly lower than (for example) standard VHS entertainment quality video. Another British Telecom product referred to as the CamNet headset is used in Peterhead to provide the consultants in Aberdeen with a telepresence in Peterhead. This provides the consultants in Aberdeen with the ability to see "over the shoulder" of a paramedic or doctor while the patient is examined. The CamNet headset includes

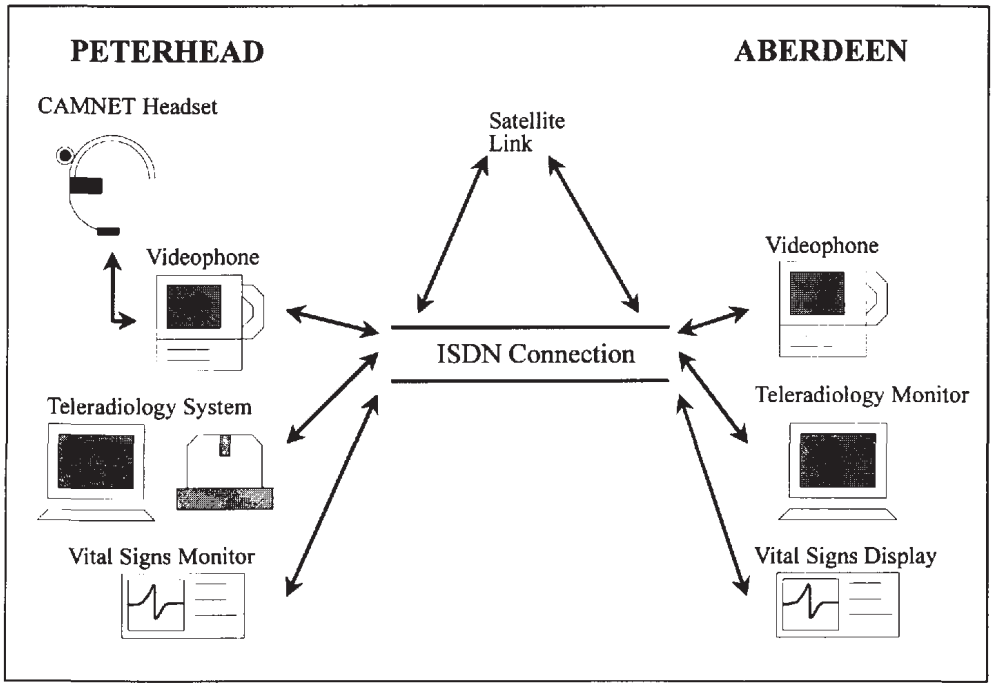

Figure 2.1 SAVIOUR telemedicine system block diagram.

a small video camera, a microphone and ear piece, and a miniature view screen. The consultant can talk to the paramedic or general practitioner and request them to look in the appropriate directions and thereby point the camera at the features of interest to the consultant. In addition, the facilities are available for the expert to access computer-based information relevant to the problem and send this over the communications link to be displayed on the miniature Cam-Net monitor. In many accident and emergency cases, x-ray information is very important in making a preliminary diagnosis. A high-resolution scanner is used to digitize a radiograph for subsequent transmission to Aberdeen. The radiograph can then be displayed on a high-resolution monitor at both ends of the link and the two practitioners can discuss the image. The facilities exist for both practitioners to manipulate a shared pointer onto the image. In addition, many types of vital signs monitoring equipment such as electrocardiographs have data ports on the back to allow data files to be transferred to/from a computer. This data can then be transferred across the link for display at the central site.

Most of the technology used in the SAVIOUR project is not new technology, but has been in existence for some time. However, in order to make a practical telemedicine system useful this equipment must be integrated together to provide a consistent and user friendly interface. This is extremely difficult as it is desirable to include a wide range of equipment in a telemedicine system. Therefore, in the future we would postulate that telemedicine systems should be based on open standards rather than completely proprietary solutions. Such standards are already under development within European AIM projects. Within a hospital environment, space is often at a premium and it is highly desirable to integrate the various pieces of equipment into a single cabinet, thereby also alleviating interconnection problems. Recent developments in desktop videoconferencing have lead to the situation where much of the necessary equipment could in the future be integrated into a single desktop PC.

Developments in communication networking both within the hospitals and in the wider area will provide a much greater capability for high-quality video communications. However, in the near future there is likely to be a proliferation of different types of networks, resulting in the need for a range of protocols to ensure interworking and flexible video CODECs that are capable of adapting their operation to make the best use of the networks that are available.

The clinical trial carried out within the SAVIOUR project has shown that many patient transfers can be avoided following a teleconsultation, as transfers were found to be unnecessary. In a smaller number of cases, the teleconsultations resulted in the immediate transfer of a patient where a potentially serious condition was identified.

The authors believe that telemedicine is set to become an important technique for providing high-quality and cost-efficient healthcare, especially in

remote rural locations such as the North of Scotland. As well as the technology issues that must be addressed to make this happen, there are also a number of clinical issues that must be resolved. For example, even after further development on equipment integration and user friendly interfaces, a significant training requirement concerning the use of a telemedicine system is still likely to remain. For efficient operation of the telemedicine system, integration with medical records systems must be achieved. This has a number of implications for confidentiality and privacy where information is transmitted over public networks. A further issue remains concerning clinical responsibility during a teleconsultation. The situation of a medical practitioner at a remote site seeking the opinion of a specialist is standard practice, and used in this context the telemedicine system simply provides a better method of obtaining advice. However, it is possible that the telemedicine system could be used in other contexts such as, for example, when the patient interacts directly with the remote consultant. In such cases, the question of who is clinically responsible for the patient is not clear.

2.4.2 Videoconferencing and Education

A large-scale project is under way to allow desktop videoconferencing between school children all over the world, referred to as the Global Schoolhouse (GSH) project. The original idea for the GSH project began in 1985 and a large grant was obtained in 1992, allowing the project to expand considerably. The GSH is funded by the U.S. National Science Foundation, Cornell University, the University of Illinois, and corporate sponsors including AT&T, Cisco, Farallon, Sprint, SuperMac, and Zenith Electronics. The project makes use of a free (but copyrighted by Cornell University) videoconferencing software program called Cu-SeeMe to enable videoconferencing over the Internet.

Cu-SeeMe is designed to run on PCs and to support multiparty videoconferencing over the Internet. Development of Cu-SeeMe began in 1992 with the aim of producing an Internet-based videoconferencing system that would use low-cost hardware. The program now supports conferencing using video, audio, and data exchange. Cu-SeeMe uses a simple efficient video compression technique largely based on frame differencing. Multiparty conferencing is handled by using a special computer called a reflector. The reflector performs the multicasting function.

The GSH project currently allows school students in 12 U.S. states and six other countries to communicate and share information. The aim of the project is to demonstrate how the Internet can be used as a tool for research and as a medium for interactive and collaborative learning. In the initial phase of the project, from 1992 to 1994, elementary and high-school students focused on the four topics of alternative energy resources, solid-waste management,

space exploration, and weather and natural disasters. These topics were studied by conducting live videoconferences and exchanging e-mail messages over the Internet. Many leaders in the various fields have participated in these conferences, including scientists, senators, and members of the White House. One of the main advantages of these conferences is the ability to bring expert outside resources into the classroom.

2.5 SUMMARY

The applications for digital video communications technology are very wide. It is expected that as costs fall, digital technology will gradually replace analog technology in the majority of applications. The integration of digital video technology into PCs is making possible a whole range of new applications. Desktop videoconferencing is a particularly powerful application that combines video, computing, and communications technology. In the future, desktop video communications may have an impact on a similar scale to the telephone. Desktop videoconferencing is useful in a whole range of application areas, including distance learning and telemedicine. In this chapter, we hope to have provided the reader with an awareness of the likely impact of digital video communications and an idea of some of the many applications for this technology.

References

[1] ETSI, "Digital broadcasting systems for television, sound and data service; Satellite Master Antenna Television (SMATV) distribution systems," ETS 300 473, 1994.
[2] ETSI, "Digital broadcasting systems for television, sound and data service; Framing structure, channel coding and modulation for 11/12 GHz satellite services," ETS 300 421, 1994.
[3] ETSI, "Digital broadcasting systems for television, sound and data service; Framing structure, channel coding and modulation cable system," ETS 300 429, 1994.
[4] AIM Project A2050, "Biosignal representation, integration and telecommunication services in rehabilitation," 1993.

Video Compression Techniques and Standards

3

3.1 INTRODUCTION

All of the applications described in Chapter 2 have been made possible by recent advances in digital video techniques. *Digital video* is video information that is stored and transmitted in digital form. Digital video techniques have been used for a number of years, for example in the television broadcasting industry. However, until recently a number of factors have prevented the widespread use of digital video. Analog video occupies a bandwidth of a few megahertz. When an analog video signal is converted into digital form at an equivalent quality, the bit rate of the digital signal is typically well over 100 Mbps. This bit rate is too high for most networks and processors to handle. The solution is to *compress* (or encode) the digital video information and to store and transmit it in a compressed form. Compression techniques for digital video have been continually improving over the last two decades. International standards now provide standard techniques for digital video coding. At the same time, processor technology has improved dramatically in recent years. The availability of cheap, high-performance processors together with the development of international standards for video compression have enabled a wide range of video communications applications such as those described in Chapter 2.

The standards for video coding all make use of the *redundancy* inherent within digital video information in order to achieve a substantial reduction in bit rate. A still image, or a single frame within a video sequence, contains a significant amount of *spatial* redundancy. It is possible to represent or encode the information in a more compact form that eliminates some of this redundancy. For example, the image data may be transformed from the spatial domain into another domain in which the information is represented more compactly. Certain components of the transformed information can be eliminated without seriously reducing the visual quality of the decoded image. The information can then be efficiently encoded using entropy encoding (for example, variable-length encoding such as Huffman coding).

Snap - Exhibit 1010
Snap v. Nokia - IPR2025-01439
Page 25 of 78

A moving video sequence contains *temporal* redundancy (i.e., successive frames of video are usually very similar). It is possible to achieve further compression by sending only the parts of the image which have changed from the previous frame. In most cases, changes between frames are due to movement in the scene which can be approximated as simple linear motion. By predicting the motion of regions from the previous transmitted frame (*motion prediction*) and then sending only the prediction error, the transmitted data can be reduced further.

In this chapter, we describe some of the more popular techniques for image and video compression. We describe the main international standards for image and video coding and show how these compression techniques are used within the standards.

The Joint Photographic Experts Group (JPEG) standard describes techniques for compressing still images or individual frames of video. Using an encoder/decoder (CODEC) based on the JPEG standard, it is possible to compress images by between 10 and 20 times without seriously reducing the visual quality of the decoded image. Motion video coding for videoconferencing and videotelephony applications is supported by the H.261 standard. This standard is optimized for two-way video communications at the bit rates provided by ISDNs. A considerable amount of interest is currently being shown in video coding and compression for very low bit rate applications (less than 64 Kbps). This has led to the recent development of the H.263 draft standard for low bit rate videoconferencing. Video coding for entertainment and broadcast applications is dealt with in the Motion Picture Experts Group (MPEG) standards. MPEG1 is optimized for coding of video and associated audio for digital storage media such as CD-ROM. MPEG2 enhances the techniques of MPEG1 to support video coding for a range of video communication applications, including broadcast digital television (at an equivalent resolution and quality to analog television and also at higher resolutions). The MPEG4 initiative is addressing generic, integrated video communications.

We discuss the recent work of the Digital Audio Visual Council (DAVIC). One of the key aims of DAVIC is to provide "standard" specifications that address many of the wider issues associated with digital video communication. Finally, we introduce some of the areas that are the subject of current research.

3.2 DIGITAL VIDEO

3.2.1 Representation of Video Information

A color image is represented in electronic form in terms of *component* signals. A color can be synthesized by combining the three color primaries red, blue, and green (RGB). The RGB color component system is one means of representing

color images. Each of the three primary components contains information about brightness (*luminance*) and color (*chrominance*).

Alternatively, the luminance and chrominance information can be represented separately. A luminance signal (Y) is used to represent the "brightness" of the color. This signal can be produced from a weighted sum of the three R, G, and B components. *Color difference* signals are then required to represent the *chromaticity* of the color. These can be formed by subtracting the luminance from each primary component:

$$C_r = w_r (R - Y)$$
$$C_b = w_b (B - Y)$$
$$C_g = w_g (G - Y)$$

where w_r, w_b, and w_g are weighting factors for the red, blue, and green difference signals, respectively.

In practice, in a given color representation system the sum of three color difference signals C_r, C_b, and C_g is always the same. This means that the original color can be fully represented using only the luminance Y and any two of the color difference signals, since the remaining color difference signal can be derived from the other two. This system of color representation has certain advantages over the RGB system. The human visual system (HVS) is less sensitive to chrominance or color information than it is to brightness information. The chrominance signals can therefore be represented with a lower resolution than the luminance without significantly affecting the visual quality.

The main standards for analog color television in current use are PAL, SECAM, and NTSC. The Phase Alternation Line (PAL) system is in use throughout most of Western Europe. Sequential Colour avec Memoire (SECAM) is used in France and in parts of Eastern Europe, and National Television System Committee (NTSC) is in use in the United States and Japan. Each system uses three components: luminance Y, blue color difference U (equivalent to C_b above) and red color difference V (equivalent to C_r above). In the PAL and SECAM systems, each complete frame of video information consists of 625 active lines. The field rate is 50 Hz, which means that 25 complete frames are sent per second. NTSC has fewer active lines (525 lines per frame) and a slightly higher field rate of 60 Hz.

3.2.2 CCIR 601 Standard

In order to process and transmit video information using digital techniques, it is necessary to convert analog video into the digital domain. Analog television and video signals are sampled and converted to digital form. The choice of

sampling rate and resolution has an important effect on the quality of the digital video signal. The CCIR 601 standard [1] (proposed by the former CCIR, now the International Telecommunications Union-Radio or ITU-R) provides a standard method of encoding television information in digital form.

The luminance and color difference components described in Section 3.2.1 are sampled with a precision of 8 bits. The sampling rate is chosen to give acceptable quality compared with the original analog television signal. The luminance component of an NTSC frame is sampled to produce an image of 525 lines, each containing 858 samples. The active area of the digitized frame is 720 × 486 picture elements (pixels). The luminance component of a PAL or SECAM frame is sampled to give a 625 × 864 sample image and the active area in this case is 720 × 576 pixels. The color difference signals are sampled at a lower rate in CCIR 601: the vertical resolution is the same (i.e., 486 or 576 active lines) but the horizontal resolution is halved. This means that only the odd-numbered luminance pixels in each line have associated color difference pixels. This sampling structure is described as a 4:2:2 component system.

The bit rate of a CCIR 601 digital television signal is 216 Mbps and can be calculated as follows:

Bit rate, NTSC original:

30 frames per second, 858 × 525 luminance samples, 429 × 525 × 2 chrominance samples, 8 bits per sample.

Bit rate = 30 × 8 × ((858 × 525) + (429 × 525 × 2)) = 216.216 Mbps

Bit rate, PAL/SECAM original:

25 frames per second, 864 × 625 luminance samples, 432 × 625 × 2 chrominance samples, 8 bits per sample.

Bit rate = 25 × 8 × ((864 × 625) + (432 × 625 × 2)) = 216.0 Mbps

This high bit rate can be handled within the television studio, but is too large for transmission over practical networks.

3.2.3 The Need for Compression

A single digital television signal in CCIR 601 format requires a transmission rate of 216 Mbps, which is unacceptably high for most practical purposes. An analog television signal of comparable visual quality occupies a bandwidth of

around 6 to 7 MHz. Digital television in this format therefore compares unfavorably with analog television in terms of transmission requirements. This bit rate is too high for most existing communication networks. Most local area networks (LANs), for example, offer data transmission at rates of up to a few tens of megabits per second, and most wide area networks (WANs) support much lower data rates than this. The emerging ATM networks are capable of transmitting higher bit rates. However, distributing a CCIR 601 bitstream over these networks would still be prohibitively expensive.

There is a need to reduce this data rate before digital television and video can be integrated into existing and emerging communication systems. This means that the digital video information must be *compressed* (encoded) prior to transmission and *decompressed* (decoded) before displaying it at the receiver. A number of video coding techniques and standards have been developed within the last few years that exploit some of the inherent redundancy in still images and moving video sequences in order to provide significant data compression.

Fortunately, digital video data contains a considerable amount of redundancy [2]. Some compression can be achieved by exploiting the *statistical redundancy* within the data. For example, video data is often highly correlated both spatially and temporally. This redundancy can be removed by coding the data in a more efficient way. Compression of this nature does not destroy any of the original data and is hence a reversible process (*lossless compression*). In order to achieve higher levels of compression, it is necessary to remove *subjectively redundant* information. This is information that is not visually obvious to the viewer and can be removed without severely reducing the quality of the signal. This type of compression destroys some of the original image information, which cannot later be recovered (*lossy compression*).

Coding techniques can be grouped into two main classes: still image coding and motion video coding. *Still image coding* exploits the spatial redundancy within images. *Motion video coding* takes into account temporal as well as spatial redundancy.

3.3 CODING OF STILL IMAGES

A typical photographic-quality still image contains a large amount of spatial redundancy. The pixel values are often highly correlated. The redundancy (both statistical and subjective) can be removed to achieve compression of the image data. By removing this redundant information and by representing the data in a more efficient form, the capacity required to store or transmit the image can be greatly reduced.

The general procedure for compression image information is shown in Figure 3.1. The *encoder model* models the image in some way to exploit its

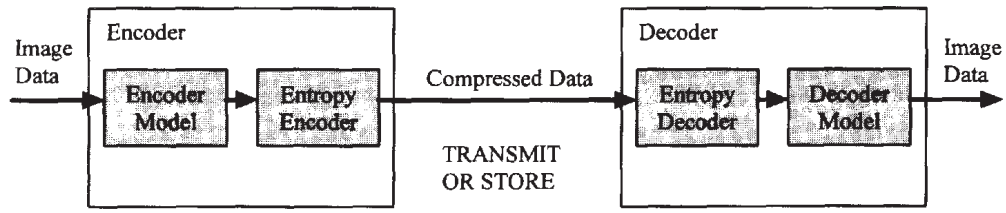

Figure 3.1 Image compression.

statistical properties and to remove redundancy. The encoder produces *symbols* that represent the information in the original image. These symbols are then *entropy encoded* (for example using Huffman encoding) to code them as efficiently as possible. The decoder carries out the reverse procedure to recreate a copy of the original image.

3.3.1 Predictive Coding

One of the simplest image compression techniques is differential pulse code modulation (DPCM). The image to be transmitted is scanned in a raster fashion (i.e., the pixels within the image are ordered in lines from left to right and from top to bottom).

Each pixel is represented as a number with a limited precision (e.g., eight bits). In a DPCM system, the pixel value itself is not transmitted. Instead, a prediction of the probable pixel value is made by the encoder based on previously transmitted pixel values. The *prediction error* between the predicted pixel value and the actual value is quantized (i.e., its precision is reduced) and transmitted. This system takes advantage of the fact that most images contain significant amounts of spatial redundancy. Because of this, neighboring pixels tend to be highly correlated and the prediction error tends to be small. Further compression can be achieved by encoding the quantized error values using variable length codes (entropy encoding). Since small errors are likely to occur more frequently than large errors, the more common values can be encoded using a shorter code and the less common values using a longer code (for example, using Huffman coding).

The decoder must use the same prediction method as the encoder. The prediction, based on previous reconstructed pixels, is added to the error value to reconstruct the current pixel.

The coding efficiency of this system depends on the accuracy of the prediction. The simplest predictor is the previous transmitted pixel value. In the example shown in Figure 3.2, the current pixel would be predicted from the value of pixel a. A slightly better prediction can be made by averaging

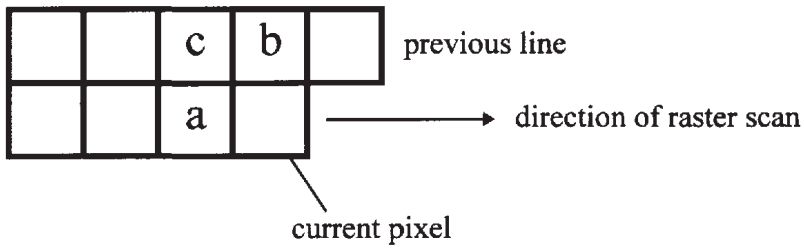

Figure 3.2 DPCM prediction.

the values of other neighboring pixels (for example, pixels a, b, and c in Figure 3.2).

Better predictions still can be made by taking more pixel values into account. The choice of the predictor and the weights of the neighboring pixel values have a significant effect on the efficiency of the algorithm. Further improvements can be gained by modifying the predictor based on the statistics of the current image (adaptive prediction).

3.3.2 Discrete Cosine Transform Coding

Transform coding is a very widely used method of compressing image information. In a transform coding system, pixels are grouped into *blocks*. A block of pixels is transformed into another domain to produce a set of transform coefficients, which are then coded and transmitted.

An effective transform will *compact* the energy in the block of pixels (i.e., the "useful information" in the block will be concentrated into a few of the coefficients). Compression is achieved by quantizing the coefficients so that the "useful" coefficients are transmitted and the remaining coefficients are discarded. The most effective compaction or decorrelation of continuous-tone image data is achieved using the Karhunen-Loeve Transform (KLT) [3]. However, the KLT has the disadvantage of being very computationally intensive and hence impractical in an image coding system.

The discrete cosine transform (DCT) is a popular alternative to the KLT for image coding [4]. For most continuous-tone photographic images, the DCT provides energy compaction that is close to the optimum. A number of fast algorithms exist for calculating the DCT of a block of pixels [5]. These factors have led to the widespread use of the DCT for image and video compression systems.

An image coding system based on the DCT will typically carry out the following steps:

Separation of the Image Into Blocks

An image is made up of a rectangular array of pixel values, which are grouped into blocks. Most existing systems use blocks of a regular size, such as 8 × 8 or 16 × 16 pixels. A larger block size leads to more efficient coding, but requires more computational power.

This block-based segmentation of the source image is a fundamental limitation of the DCT-based system. Real-world images do not generally contain features that can be neatly partitioned into regular blocks. The use of uniformly sized blocks simplifies the coding system, but does not take into account the irregular shapes within real images. It is possible to achieve better compression efficiency by using a combination of blocks of different shapes to cover the image, at the expense of increased complexity [6].

Discrete Cosine Transform

The DCT converts a block of pixels into a block of transform coefficients (of the same dimensions). These coefficients represent the *spatial frequency* components that make up the original pixel block. Each coefficient can be thought of as a weight that is applied to an appropriate basis function. The basis functions for an 8 × 8 DCT are shown in Figure 3.3. The top-left function is the

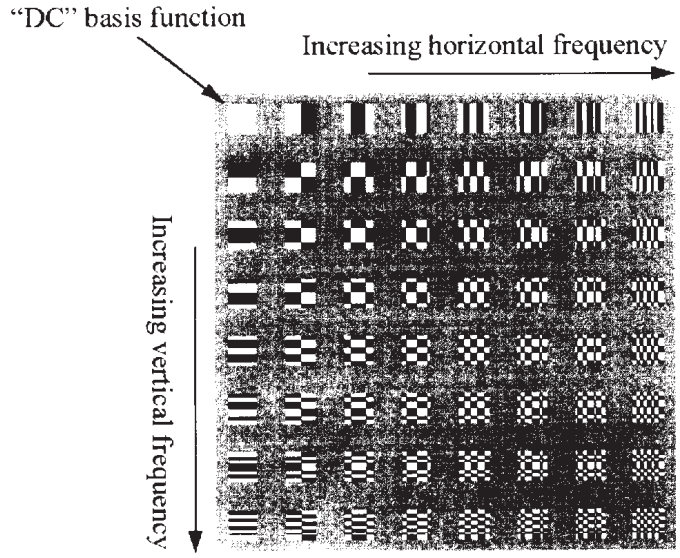

Figure 3.3 DCT basis functions.

basis function of the "dc" coefficient and represents zero spatial frequency. Along the top row, the basis functions have increasing horizontal spatial frequency content. Down the left column, the functions have increasing vertical spatial frequency content, and along the diagonal the functions have combinations of horizontal and vertical spatial frequencies. All of the coefficients other than the dc coefficient are known as "ac" coefficients.[1]

Any gray-scale 8 × 8 pixel block can be fully represented by a weighted sum of these 64 basis functions. The appropriate weights that are required to produce a particular block are the DCT coefficients for that block. The DCT operation therefore transforms a block of pixels into the set of DCT coefficients that represent the block in the spatial frequency domain. This in itself does not give compression since the information is merely being represented in a different form.

The two-dimensional DCT of an $N \times N$ block of pixel values is described by the following formula, where the pixel values are represented as $f(i, j)$ and the transform coefficients as $F(u, v)$.

$$F(u, v) = \frac{2}{N}C(u)C(v)\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} f(i, j)\cos\left(\frac{(2i+1)u\pi}{2N}\right)\cos\left(\frac{(2j+1)v\pi}{2N}\right) \quad (3.1)$$

where

$$C(x) = \begin{cases} \frac{1}{\sqrt{2}} & , x = 0 \\ 1 & \text{otherwise} \end{cases}$$

The inverse DCT of an $N \times N$ block of coefficients is given by

$$f(i, j) = \frac{2}{N}\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} C(u)C(v)F(u, v)\cos\left(\frac{(2i+1)u\pi}{2N}\right)\cos\left(\frac{(2j+1)v\pi}{2N}\right) \quad (3.2)$$

with $C(x)$ as above.

It can be seen that both forward and inverse transforms require a large number of calculations: Each coefficient in the forward transform (3.1) is calculated by summing 64 separate calculations, and so a total of 64 × 64 = 4,096 calculations are needed to transform an 8 × 8 block of pixels.

The forward and inverse DCTs are separable. Instead of performing the two-dimensional transform described by the above equations, the same result

[1]The basic functions in this figure are shown as two-valued (i.e., only black/white). In practice, the functions are cosine functions, so the intensity varies more gradually between black and white.

can be achieved by applying a one-dimensional transform along all the rows of the block and then down the columns of the block. This reduces the number of calculations required: each coefficient in the one-dimensional transform requires 8 calculations and so a row or column requires $8 \times 8 = 64$ calculations. The total number of calculations is 64×8 (for the 8 rows) + 64×8 (for the 8 columns) = 1,024. The computational complexity can be further reduced by replacing the cosine form of the transforms with a fast algorithm such as that described in [5], which reduces the operation to a short series of multiplications and additions.

Quantization

For a "typical" block within a photographic image, most of the DCT coefficients will be near zero. On average, the larger DCT coefficients will be clustered around the dc value (the top left of Figure 3.3); that is, they will be low spatial frequency coefficients. Most blocks within an image tend to contain little high-frequency content, so in general only a few of the DCT coefficients for a block will have significant values.

The DCT coefficients are quantized so that the near-zero coefficients are set to zero and the remaining coefficients are represented with a reduced precision. This can be achieved by dividing each coefficient by a positive integer (the quantizer scale or quantizer step size) and rounding the result to the nearest integer: the larger the quantizer scale, the "coarser" the quantization. This results in loss of information, but also in compression since most of the values in each block become zero. Increasing the quantizer scale leads to coarser quantization, which gives high compression and poor decoded image quality. Reducing the quantizer scale has the opposite effect.

Encoding

The statistical redundancies within the quantized DCT coefficient data can be exploited to give further compression. Nonzero coefficients are encoded using an entropy coding scheme (such as Huffman encoding). In Huffman coding, the more frequently occurring values are represented with short codes and the less frequent values with longer codes. Zero coefficients can be efficiently encoded using run-length encoding (i.e., instead of transmitting the zero values themselves, simply transmit a number representing the length of the current "run" of zeros).

The result is a compressed representation of the original image. To decode the image, the reverse procedure is carried out. First, the variable-length codes (VLCs) are decoded to recreate the quantized coefficients. These are then

multiplied by the appropriate quantizer scale to obtain an approximation to the original DCT coefficients. (This rescaling operation is sometimes referred to as "inverse quantization," but this is not strictly correct as the rescaling does not restore the original precision of the coefficients). Each block of coefficients is processed using an inverse discrete cosine transform to get back to a representation of the block in the spatial domain. This will not be identical to the original block of pixels since some of the information is lost during the initial quantization process. A lossy DCT CODEC produces characteristic distortions or "artifacts" due to the quantization process. These include "blocking" effects, where the block structure used by the CODEC becomes apparent in the decoded image, and "mosquito noise," where lines and edges in the image are surrounded by fine lines.

DCT-based image coding systems can provide compression of between 10 and 20 times while maintaining reasonably good image quality. The compression efficiency depends to some extent on the content of the image: an image with lots of detail will contain many nonzero high-frequency components and will produce more coded data than a similarly sized image with less detail. Compression can be increased by increasing the quantization scale factor. In general, higher compression can be achieved at the expense of poorer decoded quality.

3.3.3 Other Image Coding Techniques

A number of other coding techniques have been developed for still images. We will briefly discuss some of these here. A more detailed treatment of these and other coding techniques can be found in [7].

Subband coding involves filtering the source image using a series of bandpass filters to produce a series of *subbands*, each containing a band of spatial frequencies from the original image. The subbands are coded separately, for example by quantizing and then entropy encoding the subband coefficients. The image "energy" tends to be concentrated in the lower subbands. The HVS is less sensitive to losses and degradation in the higher subbands and so these can be quantized more coarsely than the lower bands. A *wavelet transform* can be applied to the image to produce a similar decomposition into spatial frequency bands. These techniques have the advantage that they do not require the image to be subdivided into blocks, so the characteristic blocking artifacts of DCT-based systems are avoided.

It has been claimed that *fractal coding* can provide very high compression with little loss in image quality. The image is described in terms of a *fractal code*, a series of transformations which, when applied to an arbitrary source image, can be used to recreate the original image. The decoder can reconstruct the image using the information contained within the fractal code. This tech-

nique has been reported as giving very high compression ratios (though there is also some skepticism about these claims; see, for example, pp. 191–193 in [7]). The process of finding the optimum fractal code to describe an image is extremely computationally intensive. This process can be simplified by splitting the image into small blocks and finding the optimum code to represent each block. Decoding is relatively straightforward, but encoding using this technique is much more complex and takes typically several hundred times longer than decoding.

3.4 CODING OF MOVING IMAGES

The concept of differential prediction discussed in Section 3.3.1 can be extended to enable efficient encoding of moving video sequences. Most video sequences contain high temporal redundancy (i.e., successive frames within the sequence are very similar). The differences between successive frames are usually due to movement in the scene and this can often be closely approximated as linear movement, for example due to a camera pan across the scene.

Figure 3.4 shows an example of a frame from a video sequence (frame 1). The following frame (frame 2) from the same sequence is shown in Figure 3.5. It can be seen that there is very little difference between the two. In this case, the camera viewpoint is fixed and the only difference is due to the movements of the two figures. Figure 3.6 was obtained by subtracting frame 1 from frame 2, pixel by pixel. In Figure 3.6, gray represents a difference of zero and white and black represent positive and negative differences, respectively. Most of the difference frame is gray (i.e., the difference between pixel values is zero). The nonzero areas are around the moving figures in the scene.

Figure 3.4 Frame 1.

Figure 3.5 Frame 2.

Figure 3.6 Difference between frames 1 and 2.

The amount of information that needs to be transmitted can be reduced considerably by encoding and transmitting the difference between the current frame and the previous transmitted frame, rather than by encoding the frame data itself: this is differential prediction (DPCM) in the temporal rather than the spatial domain. In Figure 3.6, for example, most of the blocks within the difference frame contain no information and do not therefore need to be transmitted.

Further compression can be achieved by using *motion prediction*. Each block of pixels in the current frame is matched with a similar block in the previous frame, within its immediate vicinity. The offset between the two blocks is known as a *motion vector*. The error between the current block and the

similar block in the previous frame is encoded and transmitted along with the motion vector for the block. At the receiver, the original block can be recreated by decoding the error block and adding this to the offset block in the previous reconstructed frame.

3.4.1 "Generic" DCT/DPCM CODEC

Figure 3.7 shows a block diagram of a generic CODEC which uses motion-compensated DPCM (prediction coding) and DCT-based encoding of the difference frames. This generic CODEC will serve as a useful reference for the discussion of specific video coding standards.

Encoder

The *frame store* contains a reconstructed copy of the previous encoded frame: this is used as a reference for temporal prediction. The *motion estimator* calculates motion vectors for each block of the current frame. A motion-compensated version of the previous frame is *subtracted* from the current frame to create a difference or error frame. Each block of this difference frame is then transformed using the *DCT* and the coefficients are *quantized* and, together with the motion vectors, are *entropy coded*. At the same time, the quantized coefficients are rescaled (the *Iquant* block) and inverse transformed (IDCT) to create a local copy of the encoded and decoded frame. This is used as the prediction reference for the next frame. (This ensures that the encoder and decoder use identical reference frames for motion compensation.)

Decoder

The coded data is *entropy decoded* and the coefficients are rescaled ("inverse quantization") and inverse transformed to recreate the difference frame. A motion-compensated reference frame is created using the previous decoded frame and the motion vectors for the current frame. The current frame is reconstructed by adding the difference frame to this reference frame. This frame is displayed and is also stored in the decoder frame store.

3.4.2 Bidirectional Prediction

Motion estimation based on the previous transmitted frame does not work in every situation. In a scene containing movement, it is often the case that moving objects reveal areas of the background that were previously hidden. These

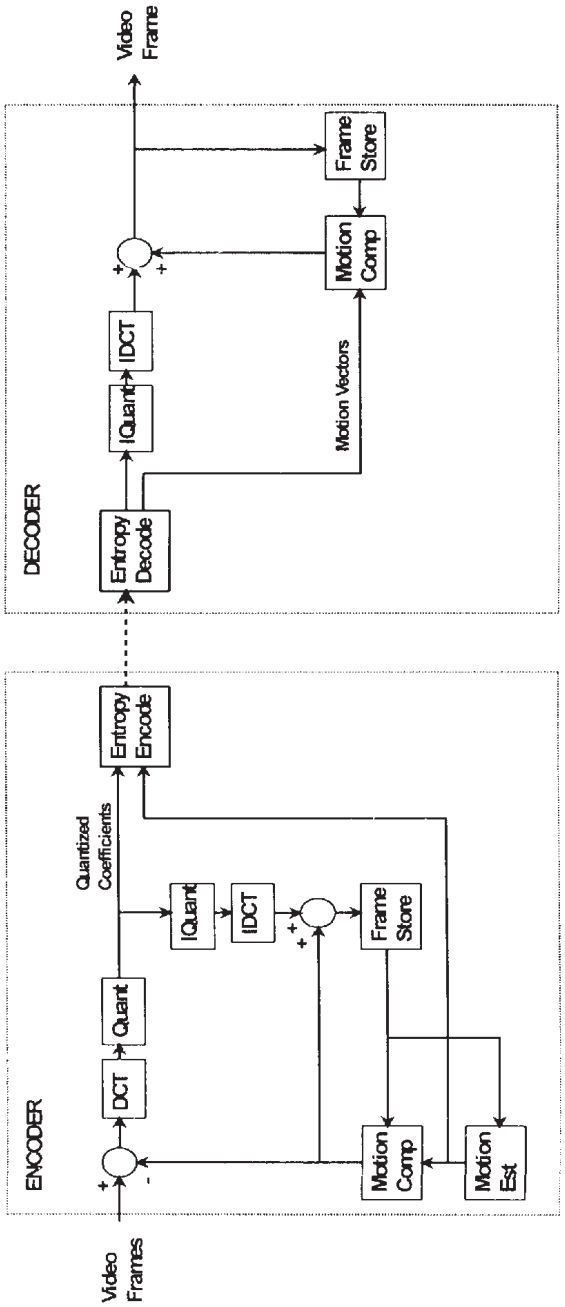

Figure 3.7 "Generic" DCT/DPCM CODEC.

hidden areas cannot be successfully predicted from the previous frame. Better prediction can be achieved by using a combination of motion prediction from a previous frame and motion prediction from a *future* frame. An attempt is made to find a matching block in the previous and in the future frame. The best match is chosen and the appropriate motion vector is encoded. This is known as interpolated or bidirectional motion prediction. This technique can give a significant improvement in compression efficiency over forward motion prediction, since the number of successfully matched blocks is increased. However, it introduces several potential problems. Firstly, the complexity of both the encoder and the decoder is increased. Secondly, a delay is introduced into the system, since the interpolated frame cannot be encoded until a future frame is read into the encoder. A delay of even a few frames can cause problems, particularly in a two-way videoconferencing application.

3.5 JPEG: STILL IMAGE CODING

ISO International Standard 10918 provides a standard format for coding and compressing still images [8]. This standard is commonly known by the acronym of the working group that developed the standard, the Joint Photographic Experts Group (JPEG). An overview of the standard and its scope is given in [9]. The standard consists of two parts: Part 1, "Requirements and Guidelines," describes the features of the standard and gives details of the syntax of the coded bitstream; and Part 2, "Compliance Testing," specifies test procedures that can be used to verify whether a particular implementation conforms to the standard.

Part 1 of the standard defines four *coding modes*. A particular application will use one of these modes depending on its requirements for image coding. The color components within the image (see Section 3.2.1) are processed separately.

- *Sequential encoding:* Each component is encoded in a single pass or *scan*. The scanning order is left to right, top to bottom. The encoding process is based on the DCT.
- *Progressive encoding:* Each component is encoded in multiple scans. Each scan contains a partially encoded version of the image. The scans can be decoded sequentially so that a "rough" image is quickly decoded and this is then built up using further scans until the full detail is achieved. (This mode will be described in more detail in Chapter 9.)
- *Hierarchical encoding:* Each component is encoded at multiple resolutions. The image can be decoded at a low resolution without decompressing the full-resolution image. (This mode will be described in more detail in Chapter 9.)

- *Lossless encoding:* Unlike modes 1 to 3, the lossless encoding mode is based on a DPCM system. As the name suggests, this mode provides compression without any loss of quality, at the expense of a considerable reduction in compression efficiency.

The standard defines a *baseline CODEC* that implements a minimum set of features (a subset of the sequential encoding mode). The baseline CODEC provides sufficient features for many general-purpose compression applications and has been widely adopted as the most common implementation of JPEG.

3.5.1 The Baseline CODEC

The baseline CODEC uses the sequential encoding mode described above. It is assumed that the samples (pixel values) within each image component are represented with 8-bit precision. The procedure for encoding an image component is as follows.

DCT

Each block is transformed using the DCT to produce an 8×8 block of transform coefficients.

Quantization

Each DCT coefficient is quantized to reduce its precision. The quantizer scale can be (and generally is) different for different coefficients in the block. Figure 3.8 shows an example of a quantization scale table that can be used in a JPEG CODEC. Prior to quantization, the DCT coefficients lie in the range −1,023 to +1,023. Each coefficient is divided by the appropriate scale factor (the top left is the scale for the DC coefficient) and is then rounded to the nearest integer. Notice that the scale tends to increase as the spatial frequency increases (to the right and down): this takes advantage of the fact that the HVS is less sensitive to quality loss in the high spatial frequencies.

Zigzag Ordering

The lower spatial frequency coefficients are more likely to be nonzero than the higher frequency coefficients. The quantized coefficients are reordered in a zigzag scanning order, starting with the dc coefficient and ending with the highest frequency ac coefficient. Figure 3.9 shows the zigzag ordering of the

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Figure 3.8 JPEG quantization table.

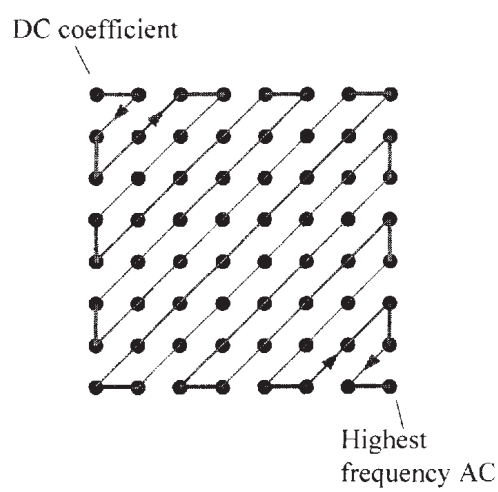

Figure 3.9 Zigzag scanning pattern in JPEG.

64 coefficients. Reordering in this way tends to create long "runs" of zero-valued coefficients.

Encoding

The dc coefficients are coded separately from the ac coefficients. There is usually a strong similarity between dc coefficients in adjacent blocks and so each dc coefficient is encoded differentially from the dc coefficient in the previous transmitted block. In a given block, most of the ac coefficients are likely to be zero, so the 63 coefficients are converted into a set of run-length/value symbols. Each symbol includes a figure for the number of zero coefficients prior to the current nonzero coefficient, followed by the coefficient itself. The set of symbols is then encoded using Huffman coding. The end result is a series of variable-length codes (VLCs) that describe the quantized coefficients in a compressed form. The baseline decoder carries out a reverse procedure to get back to a representation of the original image. The quality of the decoded image depends on the quantization scale chosen in the encoder. The compression efficiency depends very much on the image content. Two gray-scale decoded images are shown in Figure 3.10 and Figure 3.11. In Figure 3.10, the source image (shown in Figure 3.4) has been encoded into a compressed form that uses about 0.85 bits per pixel (approximately 9.5 times compression). The decoded image is of a reasonably high quality, though some quality loss is visible around the edges of the figures. This distortion appears as a "fringing" effect next to edges in the image. Distortions of this nature due to the coding process are known as "coding artifacts." The image is coded to about 0.55 bits

Figure 3.10 JPEG decoded image, 0.85 bits per pixel.

Figure 3.11 JPEG decoded image, 0.55 bits per pixel.

per pixel (approximately 14 times compression) in Figure 3.11. Here, the coding artifacts are more obvious, but the image is still clearly recognizable.

3.5.2 Other Features and Modes

The baseline CODEC is suitable for many image coding applications. However, the JPEG standard supports further options and modes that may be useful in certain situations. For example, the sequential mode can be implemented with options other than those provided by the baseline CODEC, the number of bits used to represent each image sample can be extended to 12 to support higher quality images, and arithmetic coding (another form of entropy coding) can be used as an alternative to Huffman coding. Arithmetic coding of the quantized coefficients can give higher compression, at the expense of what may be an increase in CODEC complexity.

3.6 H.261: MOTION VIDEO CODING FOR VIDEOCONFERENCING

Video coding for videoconferencing applications is supported by ITU-T Recommendation H.261, part of the H.320 group of standards. Recommendation H.320 is an "umbrella" standard that describes the various components of a videoconferencing system [10]. The components of an H.320 system are themselves the subject of further ITU-T recommendations. They include

- H.261 (video CODEC);
- G.711, G.722, or G.728 (audio CODEC);
- H.230 (audiovisual control and synchronization);
- H.221 (frame structure for ISDN channel).

A simplified diagram of the relationship between these components in a videoconferencing system is shown in Figure 3.12.

The video coding component of an H.320 videoconferencing system is described in Recommendation H.261 [11]. The H.261 algorithm was developed with the aim of supporting videoconferencing and videotelephony over ISDNs that provide data rates at multiples of 64 Kbps. An overview of H.261 is given in [12].

H.261 is designed for two-way video communications over ISDN. Only two image resolutions are supported: common intermediate format (CIF) and quarter CIF (QCIF). Video information is represented in Y, C_r, and C_b components. The resolution of the luminance (Y) component is 352×288 pixels for CIF and 176×144 pixels for QCIF. In both cases, the C_r and C_b chrominance components have exactly half the horizontal and vertical resolution of the luminance component; that is, one C_r and one C_b sample is stored for every four Y samples (described as 4:2:0 sampling). The maximum picture rate is 29.97 frames per second, but this can be reduced if required by dropping up to three frames between each pair of encoded and transmitted frames.

Many H.261 applications involve video communications over a 64 or 128 Kbps ISDN connection. At these bit rates, substantial data compression is required. QCIF resolution is often chosen in this case and the source frame rate is restricted to around 10 frames per second.

The H.261 coding algorithm is based on DCT image compression and motion-compensated interframe prediction: the CODEC design is likely to be similar to the "generic" CODEC shown in Figure 3.7. The image data is pro-

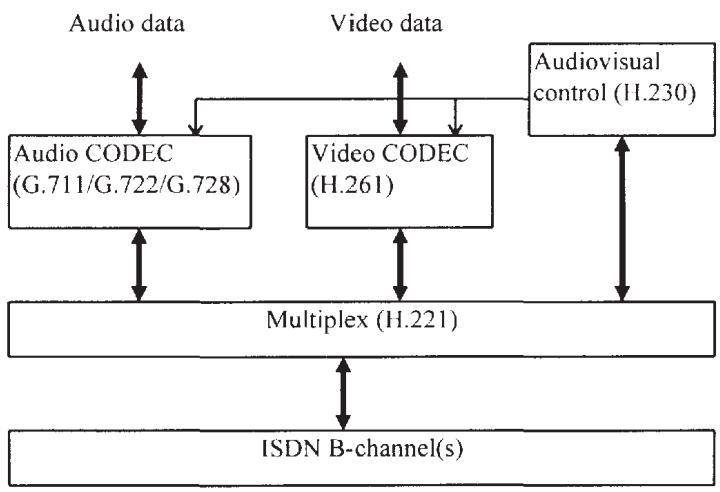

Figure 3.12 Interrelationship between H.320 standards (simplified).

cessed in *macroblocks*, where a macroblock consists of four 8×8 blocks of luminance samples, one 8×8 block of C_r samples, and one 8×8 block of C_b samples. There are two coding modes: intracoding (where no motion prediction is used) and intercoding (which uses motion prediction). Each macroblock is encoded using either intercoding or intracoding mode. All macroblocks in the first frame in an H.261 sequence must be intracoded as there are no previous frames to predict from.

The encoding process for intracoded macroblocks is similar to baseline JPEG encoding. Each 8×8 block of luminance or chrominance samples is transformed using the DCT. The coefficients are quantized and then variable-length encoded. In order to form predictions for future frames, the quantized coefficients are inverse-quantized (i.e., restored to their original magnitude) and inverse-transformed to provide a local copy of the frame, which will be decoded at the receiver. This copy is stored in a framestore.

Macroblocks in subsequent encoded frames are usually chosen to be intercoded. Each intracoded macroblock is motion-predicted from a nearby macroblock in the previous frame. The luminance samples in the current macroblock form a 16×16 block. This 16×16 block is compared with 16×16 blocks in the neighborhood within the previous frame (held in the framestore) and the closest matching block is chosen as the basis for motion prediction. The offset between the previous and current macroblocks is encoded as a motion vector for the macroblock. If the prediction error (i.e., the difference between the previous and current macroblocks) is less than a threshold value, no further information is encoded. Otherwise, the error macroblock is encoded using the DCT, quantization, and variable-length encoding steps described earlier. If the prediction error is above a second threshold value, then the motion compensation process does not give significant compression gain: in this case, the macroblock is intracoded. In addition, the predicted frame is reconstructed (by decoding the quantized coefficients at the encoder) and stored in the framestore for the next prediction.

The bitstream produced by an H.261 encoder has a hierarchical structure, shown in Figure 3.13. Blocks of variable-length coded coefficients are collected together to form macroblocks, as described above. Macroblocks are themselves collected to form a group of blocks (GOB) and a complete picture is made up of several GOBs.

H.261 is typically used to send coded data over a constant bit rate (CBR) channel, such as an ISDN channel. The encoder output bit rate (if left unconstrained) varies depending on the current activity in the scene. For example, a scene containing a lot of motion will generate much more data than a scene with little motion, since more macroblocks will need to be coded and transmitted. To map this varying bit rate onto the CBR channel, a rate control mechanism is required. The coded bitstream is buffered to smooth out any short-term

Snap - Exhibit 1010
Snap v. Nokia - IPR2025-01439
Page 46 of 78

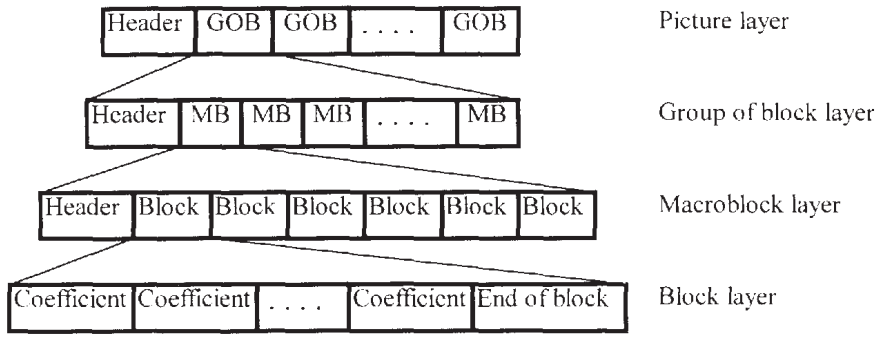

Figure 3.13 Structure of H.261 bitstream.

variations in bit rate and this buffer is emptied at the CBR of the channel. If this buffer fills up due to an increase in scene activity, the quantization scale in the encoder is increased. This increases the compression factor and acts to reduce the output bit rate. If the buffer starts to empty as a result of less scene activity, then the quantization scale is reduced in order to maintain a CBR output. This means that the compression, and hence the quality, can vary considerably depending on the amount of motion in the scene. A common characteristic of H.261 CODECs operating at bit rates of about 128 Kbps and below is that the decoded video signal degrades and "breaks up" if the motion suddenly increases.

H.261 provides a relatively straightforward system of encoding and decoding video information for two-way video communications. Its main limitation is the poor quality of decoded video, particularly at lower bit rates.

3.7 H.263: LOW BIT RATE VIDEO CODING

The majority of homes and small businesses have access only to the existing analog public switched telephone network (PSTN). Using the V.34 modem standard, a PSTN modem can now support a data rate of 28.8 Kbps. High-bandwidth links to the home (based on technologies such as optical fiber, coaxial cable, or asynchronous digital subscriber loop) are not likely to become widespread in the near term. There is therefore a need for video coding standards that can provide adequate-quality visual communications at bit rates of about 30 Kbps. This area is being addressed by the ITU-T.

The low bit rate coding group within ITU-T study group 15 is taking two approaches: firstly by developing a new video coding standard based on existing techniques in the near term and secondly by making use of fundamentally new coding techniques in the longer term. The near term standard, H.263, is now

technically "frozen" and is at the draft stage [13]. H.263 is largely based on H.261, with a number of improvements that can provide higher quality video at low bit rates. These improvements include

- Motion compensation with half-pixel prediction (as opposed to full-pixel prediction in H.261). If the motion vector points to half-pixel values, these are produced by interpolation between neighboring pixel values. This can give improved motion estimation and hence a lower prediction error.
- Unrestricted motion vectors (optional). Part of the macroblock indicated by a motion vector may be outside the picture. The pixel values outside the boundary are constructed by interpolating from the edge pixels, which can reduce the prediction error if movement occurs across the edges of the picture.
- Arithmetic coding rather than Huffman coding (optional).
- Advanced prediction mode (optional). Four motion vectors, each for one 8×8 luminance block within a macroblock, are used instead of one motion vector for the macroblock. The four vectors require more bits than a single vector. The prediction error is compared with the error for the equivalent single vector and if the error is significantly less, then the four motion vectors are used.
- PB frames mode (optional). Two frames are coded as one unit: the next P frame (forward predicted from the previous P frame) together with a B frame bidirectionally predicted between the previous and next P frames. This mode is based on the B pictures used in the MPEG standards (see Section 3.8.1).

The use of some or all of these features can lead to a significant improvement in decoded quality over H.261, particularly at low bit rates. They require a more complex CODEC; however, video coding and processing hardware is continually improving, so higher complexity can be justified if it leads to improved video quality.

The long-term initiative of the ITU-T, described as H.32P/long term, is likely to address low bit rate video communications in a wider sense. It is intended to make use of new techniques that provide a considerable improvement in performance over existing systems and to support video communications in other low bit rate environments such as mobile communication networks.

3.8 MPEG: MOTION VIDEO CODING FOR ENTERTAINMENT AND BROADCAST

The Moving Picture Experts Group (MPEG) is part of the International Standards Organization working group ISO-IEC/JTC1/SC2/WG11. The MPEG committee

started its activities in 1988. Since then, the efforts of this working group have led to two international standards, informally known as MPEG1 and MPEG2. Another standard, MPEG4, is currently under development. These standards address the encoding and decoding of video and audio information for a range of applications.

3.8.1 MPEG1 for CD Storage

The first of the standards to be released following the work of the MPEG committee was ISO 11172, popularly known as MPEG1 [14]. An overview of this standard is given in [15]. MPEG1 has three key parts. Part 1, "Systems," addresses synchronization and multiplexing issues for multiple compressed audio and video streams. Part 2, "Video," addresses video coding; and Part 3, "Audio," addresses the compression of digital audio information.

The MPEG1 standard supports coding of video and associated audio at a bit rate of about 1.5 Mbps. The video coding techniques described in Part 2 can be applied at a range of different bit rates, but have been optimized for a coded bit rate of around 1.1 to 1.5 Mbps. The bit rate of a coded video stream together with its associated audio matches the data transfer rate of about 1.4 Mbps provided by a CD-ROM system. MPEG1 is well-suited to video coding for CD-based video entertainment, but can be also used for a range of other storage media as well as for data transmission.

The requirements of these applications are different than those of two-way communications, so there are a number of key differences between MPEG1 video and H.261. The MPEG1 standard does not specify a particular design of video encoder. Instead, it specifies the syntax of the coded bitstream and it describes a model decoder. Many of the details of the encoder design are left to the developer. However, in order to comply with the standard, the coded bitstream must follow the correct syntax and must be decodeable using the model decoder.

Frames of video are coded as *pictures*. Each frame is encoded in a progressive order (i.e., fields are not coded separately). The resolution of the source video information is not restricted. However, a resolution of 352 × 240 pixels in the luminance component, at about 30 frames per second, gives an encoded bit rate of around 1.2 Mbps. The coding algorithm is optimized for this approximate bit rate, so higher resolution sequences will not necessarily be encoded efficiently using MPEG1. A video sequence of resolution 352 × 240 pixels provides image quality that is broadly similar to that of VHS video. MPEG1 allows a significant amount of flexibility in the choice of encoding parameters. A specific set of parameters (the *constrained parameters*) are defined that should provide sufficient functionality for many applications.

There are three main types or classes of coded picture in MPEG1: I pictures, P pictures, and B pictures.

I pictures (intrapictures) are intraframe encoded without any temporal prediction using a coding technique similar to the intracoding mode of H.261. Blocks of pixel values are transformed using the DCT, quantized, reordered (in a zigzag scan such as that described in Section 3.5) and variable-length encoded. Coded blocks are grouped together in macroblocks, where each macroblock consists of four 8×8 luminance blocks, one 8×8 C_r block, and one 8×8 C_b block. The chrominance components have half the horizontal and vertical resolution of the luminance components (i.e., 4:2:0 sampling).

P pictures (forward predicted pictures) are interframe encoded using motion prediction from the previous I or P picture in the sequence. As in the H.261 standard, the luminance component of each macroblock is matched with a similar 16×16 sample region in the previous I or P picture. The difference macroblock, together with the motion vector, is encoded and transmitted. Macroblocks in P pictures may be optionally intracoded (without prediction), if, for example, the motion prediction is not effective.

B pictures (bidirectionally predicted pictures) are interframe encoded using interpolated motion prediction between the previous I or P picture and the *next* I or P picture in the sequence. Each macroblock is compared with the neighboring area in the previous and the next I or P picture. The prediction macroblock is chosen from the previous picture, from the next picture, or by calculating the average of both. The prediction mode that produces the smallest prediction error is chosen (again, intracoding may also be chosen). The forward and/or reverse motion vectors, together with the prediction error, are encoded and transmitted. B pictures are not used as a reference for further predicted pictures.

The three picture classes are grouped together in GOPs where a GOP consists of one I picture followed by a number of P and B pictures. An example of a GOP structure is shown in Figure 3.14, where each I or P picture is followed by two B pictures. The structure and size of each GOP is not specified in the standard and can be chosen to suit the application. I pictures have the lowest compression efficiency because they use only intraframe compression; P pictures have a higher compression efficiency due to the use of motion prediction; and B pictures have the best compression efficiency of the three due to the bidirectional motion estimation. In general, a large GOP size leads to efficient compression since few I pictures are coded. However, I pictures provide a useful access point to the sequence, which is particularly important for applications that require random access.

The encoded MPEG1 bitstream is arranged in a hierarchical structure, shown in Figure 3.15. The highest level is the *video sequence* level. The video sequence header describes the basic parameters of the coded sequence such as

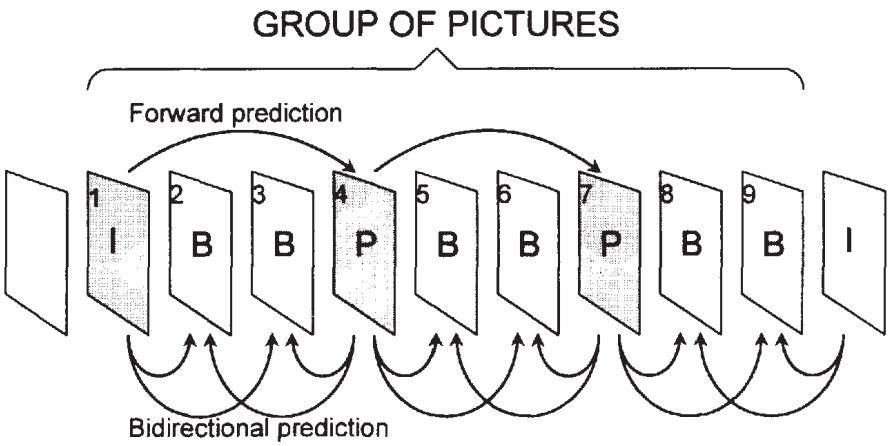

Figure 3.14 MPEG GOP structure.

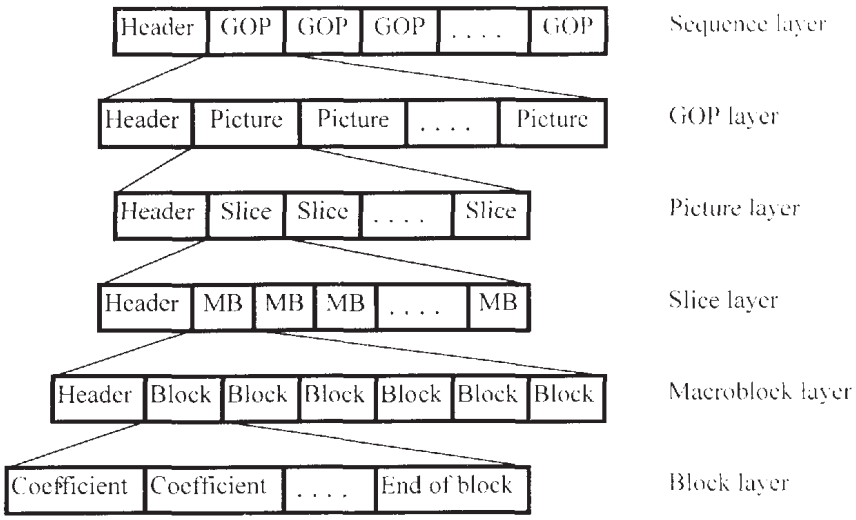

Figure 3.15 Structure of MPEG bitstream.

spatial and temporal resolution. Coded frames are grouped together in GOPs (as described above) at the *group of pictures* level. Within each GOP are a number of *pictures*. The picture header describes the class of coded picture (I, P, or B) and a temporal reference number that gives its position within the GOP. The next level is the *slice level*. Each slice consists of a continuous series of macroblocks. A complete coded picture is made up of one or more slices. The slice header is not a variable-length code and can be used by the decoder

to resynchronize with the coded bitstream if an error occurs. The remaining levels in the hierarchy are the *macroblock level* and the *block level.*

In most MPEG1 applications, the end user only needs to decode MPEG data (for example, prerecorded MPEG video on a CD-ROM). The decoding process is therefore made as simple as possible, at the expense of a more complex encoding process. The use of B pictures leads to a delay in the encoding and decoding processes, since both the previous and the next I/P picture must be received and stored before a B picture can be encoded or decoded. In order to minimize the delay at the decoder, the coded pictures are reordered by the encoder prior to transmission or storage, such that the I and/or P pictures required to decode each B picture are placed *before* the B picture. For example, the original order of the pictures in Figure 3.14 is as follows:

$I_1\ B_2\ B_3\ P_4\ B_5\ B_6\ P_7\ B_8\ B_9\ I_{10} \ldots$

The order in which the pictures are transmitted or stored is as follows:

$I_1\ P_4\ B_2\ B_3\ P_7\ B_5\ B_6\ I_{10}\ B_8\ B_9 \ldots$

The decoder receives and decodes I_1 and P_4. It can display I_1 since this is the first frame in display order. B_2 is then received and can immediately be decoded and displayed, since the past and future reference frames required to recreate the frame (I_1 and P_4) are already available at the decoder. B_3 is decoded and displayed next, P_4 can then be displayed, and so on. The decoder only needs to store two decoded frames at a time and the decoder can display each frame after at the most a one-frame delay.

The decoder processing requirements are relatively simple, since the motion information for each macroblock is already available. A real-time decoder can be implemented in hardware at a low cost. The encoder, however, has a much more demanding task as it must carry out block-matching operations for motion prediction. In addition, it must store more frames during encoding. MPEG1 encoding is often carried out "off line" (i.e., the video information is not encoded in real time). Real-time MPEG1 encoders are much more expensive than decoders.

3.8.2 MPEG2 for Broadcasting

It was quickly established that MPEG1 would not be ideal for coding of video and audio information for television applications. "Television quality" video information (e.g., CCIR-601 video) produces a higher encoded bit rate that requires different coding techniques than those provided by MPEG1. In addi-

tion, MPEG1 cannot efficiently code interlaced fields. For these reasons, the second MPEG standard was developed, known as MPEG2. The MPEG2 standard comes in three main parts: systems [16], video [17], and audio [18]. Parts 1 and 2 (systems and video) are defined as ITU-T standards as well as ISO standards. This reflects the close cooperation between the two organizations on the development of MPEG2.

MPEG2 extends the functions provided by MPEG1 to enable efficient encoding of video and associated audio at a wide range of resolutions and bit rates. Some of the bit rates and applications are summarized in Table 3.1.

Part 1 of the MPEG2 standard (Systems) specifies two types of multiplexed bitstreams: the *program stream* and the *transport stream*. The program stream is analogous to the systems part of the MPEG1 standard. It is produced by packetizing and multiplexing one or more streams from video/audio encoders such as MPEG2 video and audio encoders. It is designed for flexible processing of the multiplexed stream and for environments where there is a low probability of error. The transport stream is constructed in a different way than the program stream and includes a number of features that are designed to support video communications or storage in environments in which a significant number of errors may occur. The transport stream is discussed in more detail in Chapter 6.

MPEG2 video [17] is based on the MPEG1 video encoding techniques, but includes a number of enhancements and extensions. The constrained parameters subset of MPEG1 is incorporated into MPEG2 and the hierarchical structure of the video sequence is similar to the MPEG1 structure (see Figure 3.15). Some of the more flexible encoding features of MPEG1 are dropped and a number of other features are added. The extra features (as compared with MPEG1 video) include the following:

- Features to support interlaced video as well as progressive (noninterlaced) video. These include the ability to encode video directly in interlaced

Table 3.1
MPEG2 Bit Rates and Applications

Application	*Approx. Resolution*	*Approx. Coded Bit Rate*
Home entertainment video ("MPEG1" quality)	352 × 240, 30 Hz	1.5 Mbps
Digital television	720 × 486, 30 Hz	5–10 Mbps
Extended definition television	1,920 × 1,080, 30 Hz	30–40 Mbps

Note: The ISO MPEG study group originally proposed an MPEG3 standard to support encoding of video at higher resolutions than CCIR 601 digital television. However, it became clear that support for higher resolutions or definitions could be adequately provided by the different modes of MPEG2 video.

form as well as a number of new motion prediction modes. In addition to prediction from an equivalent macroblock in a previous or future frame, MPEG2 supports prediction from a previous or future field. This increases the efficiency of prediction.

- More chrominance sampling modes. A 4:2:2 sampling produces chrominance components with the same vertical resolution and half of the horizontal resolution of the luminance component. This provides better color resolution than 4:2:0 sampling. A higher resolution still is supported with 4:4:4 sampling, where the chrominance components are encoded at the same resolution as the luminance component.
- An alternative block scanning pattern to the zigzag scan discussed earlier. The alternative pattern can improve the encoding performance of interlaced video.
- Scalable coding modes. Each enables video information to be coded into two or more *layers*.

Scalable coding will be discussed in detail in Chapter 9. There are four scalable coding modes in MPEG2:

- *Spatial scalability* is analogous to the hierarchical coding mode in JPEG, where each frame is encoded at a range of resolutions that can be "built up" to the full resolution. For example, a standard decoder could decode a CCIR 601 resolution picture from the stream while a high definition television (HDTV) decoder could decode the full HDTV resolution from the same transmitted stream.
- *Data partitioning* enables the coded data to be separated into two streams. A high-priority stream contains essential information such as headers, motion vectors, and perhaps low-frequency DCT coefficients. A low-priority stream contains the remaining information. The encoder can choose which components of the syntax are placed in each stream.
- *Signal to noise ratio (SNR) scalability* is similar to the successive approximation mode of JPEG, where the picture is encoded in two layers, the lowest of which contains a "coarse" quality version of the picture and the highest of which contains enhancement information needed to decode the picture at its full quality.
- *Temporal scalability* is a hierarchical coding mode in the temporal domain. Here, the base layer is encoded at a lower frame rate. The intermediate frames required to give the full frame rate are interpolated between successive base layer frames. The difference between the interpolated frames and the actual intermediate frames (the prediction error) is encoded as a second layer.

MPEG2 describes a range of *profiles* and *levels* that provide encoding parameters for a range of applications. A *profile* is a subset of the full MPEG2 syntax that specifies a particular set of coding features. Each profile is a superset of the preceding profiles. Within each profile, one or more *levels* specify a subset of spatial and temporal resolutions that can be handled. The profiles defined in the standard are shown in Table 3.2.

Each level puts an upper limit on the spatial and temporal resolution of the sequence, as shown in Table 3.3.

Only a limited number of profile/level combinations are recommended in the standard. These are summarized in Table 3.4.

Table 3.2
MPEG2 Profiles

Profile	*Features*
Simple	4:2:0 sampling, I/P pictures only, no scalable coding
Main	As above, plus B pictures
SNR	As above, plus SNR scalability
Spatial	As above, plus spatial scalability
High	As above, plus 4:2:2 sampling

Table 3.3
MPEG2 Levels

Level	*Maximum Resolution*
Low	352 × 288 luminance samples, 30 Hz
Main	720 × 576 luminance samples, 30 Hz
High-1,440	1,440 × 1,152 luminance samples, 60 Hz
High	1,920 × 1,152 luminance samples, 60 Hz

Table 3.4
Recommended Profile/Level Combinations

	Level			
Profile	*Low*	*Main*	*High-1,440*	*High*
Simple		×		
Main	×	×	×	×
SNR	×	×		
Spatial			×	
High		×	×	×

Particular profile/level combinations are designed to support particular classes of applications. Simple profile, main level is suitable for conferencing applications, as no B pictures are transmitted (leading to a low encoding and decoding delay). Main profile, main level is considered suitable for most digital television applications and the majority of MPEG2 encoders and decoders currently available support main profile/main level coding. The two high levels are designed to support HDTV applications and can be used with either nonscalable coding (main profile) or with spatially scalable coding (spatial/high profiles). It should be emphasized that the profiles and levels are only recommendations and that other combinations of coding parameters are possible within the MPEG2 standard.

3.8.3 MPEG4 for Integrated Visual Communications

The MPEG4 working group was formed in 1993 as part of the ISO MPEG committee. The original aim of the MPEG4 effort was to produce a generic video coding standard for low bit rate applications (i.e., less than 64 Kbps). The aim was to complete the Committee Draft of the standard by 1997 and to incorporate fundamental advances in video coding technology that would support efficient coding at very low bit rates. Much research is still underway into very low bit rate coding, but it is not clear that the new techniques will be sufficiently developed in time to be incorporated into the standard. The focus of MPEG4 has therefore changed. The aim is now to address application areas that are not fully supported by existing standards, including audiovisual services that require mobile communications, interactivity, and integration with other multimedia data. MPEG4 will be a generic audiovisual coding standard that supports the wide range of applications emerging in these areas [19]. The functional areas that will be addressed by MPEG4 include

- Content-based coding and manipulation of multimedia data (i.e., coding objects or entities rather than just frames of video);
- Improved coding efficiency and coding of concurrent streams (e.g., multiple views);
- Robustness to errors;
- Content-based scalability.

The deadline for proposal submissions for MPEG4 was October 1995. The proposals have been evaluated and the successful submissions will eventually be incorporated into the ISO Committee Draft in late 1997. Interestingly, the coding algorithm chosen for the initial support of lower bit rate video coding is based on the H.263 algorithm. The structure of MPEG4 will be based around a "toolkit" of functions, such that specific applications should be able to select

Snap - Exhibit 1010
Snap v. Nokia - IPR2025-01439
Page 56 of 78

and download tools as required. This flexible structure should allow more advanced video coding techniques to be incorporated into the standard in the future.

3.9 DIGITAL AUDIO VISUAL COUNCIL (DAVIC)

DAVIC is an international forum whose goal is to promote global standards for digital audiovisual applications and services. The forum first met in 1994 and currently has a membership of over 200 organizations representing a wide range of interests in emerging audiovisual applications and broadband communications networks. The aim of the forum is to make use of elements of existing standards and to combine and, where necessary, augment these elements to provide specifications for complete digital audiovisual systems.

One of the key targets of DAVIC is "to establish technical specifications required for the implementation of a wide range of audiovisual services of applications with the main goal of providing end-to-end interoperability across countries and applications/services" [20]. To this end, the DAVIC forum has issued calls for proposals to invite technical solutions to the particular requirements of these applications. The responses to the first calls were used to develop the first issue of the DAVIC specifications (published in December 1995 [21]). These specifications define a complete end-to-end digital audiovisual system that is capable of providing a wide range of services. All of the necessary system components are defined in the specifications in terms of "toolkits." A specific application will choose a subset of tools. Particular groupings of tools, known as *profiles*, are recommended in the specifications. The use of profiles should limit the number of possible combinations of tools.

The DAVIC specifications draw on elements of a number of existing standards and protocols, including several components of MPEG2 and protocols related to ATM networks.

3.10 FUTURE DEVELOPMENTS AND RESEARCH AREAS

The past five years have seen the standardization of techniques for image and video coding and decoding. A very wide range of algorithms for coding and compression of still and moving images have been developed and new algorithms and variations are being devised all the time. There is, however, limited benefit in adopting a nonstandard technique unless its performance is significantly superior to that provided by the international standards. Each of the standards is the result of a number of years of effort by many individuals and organizations and the techniques incorporated into the standards have been exhaustively tested and refined.

There are still, however, areas in which further work needs to be carried out and in which further developments could be beneficial. Firstly, there is a requirement for coding algorithms that provides a significant improvement over the existing standards. Video coding techniques are continuing to advance and hardware implementations are becoming faster and cheaper. The need for significant improvements in coding performance, particularly at lower coded bit rates, is being addressed by the development of the H.263 and MPEG4 standards. Secondly, many of the implementation details for systems that comply with the standards are left to the developer. There is scope for optimization and improvement in this area and for efficient implementation of the standards in hardware and in software. Thirdly, many of the issues relating to the development of effective video communication systems are not addressed by the existing standards. These issues, among others, are currently being tackled by the DAVIC forum.

3.10.1 Optimization and Application of Existing Standards

Many problems and issues need to be addressed in order to apply the international standards in practical video communication systems. The MPEG standards, for example, are restricted in their scope and in the components and techniques that they specify. In both MPEG1 and MPEG2, the standard specifies the syntax of the coded bitstream, but does not specify the design of the encoder or the details of the decoder. Many of the further issues relating to the transmission of coded video through communication networks are not addressed in the standards. There is therefore considerable scope for optimizing video communication systems while remaining within the standards.

The performance of the encoder is affected by a number of factors. Some of the critical areas include block matching (i.e., searching for the best match for motion-compensated prediction), implementation of the discrete cosine transform, and output rate control. The speed of the encoder and the quality of the decoded video can be affected significantly by these design factors. The choice of hardware implementation of the encoder and decoder is another very important factor that influences performance.

The practical characteristics of the network used to transmit coded video have a further influence on performance. Coded video data has a number of demanding transmission requirements. It generally occupies a very large transmission bandwidth. It can be very sensitive to transmission errors and, because it is real-time data, it is affected by network transmission delay and by delay variation. These requirements can be difficult to meet in a practical system. These "quality of service" issues will be discussed in detail in Chapter 6.

Particular video communication applications have their own distinct requirements. For example, a two-way videoconferencing system requires a

low transmission delay. In this type of system, it is important that the encoder and decoder are not prohibitively expensive. A broadcast system such as digital television has different requirements. Here, transmission delay is not so significant since there is no two-way communication. The encoder can be relatively sophisticated since encoding will be performed by the broadcaster. The decoder must be inexpensive in order to appeal to consumers. More advanced video communication systems such as video databases, video on demand systems, and medical applications have further requirements and constraints that must be met.

There is currently a great deal of research and development in progress that is attempting to address these issues. Improvements in algorithms and in hardware and software implementations are continually producing advances in digital video communications. There is still much scope for improvement and development within the existing video coding standards.

3.11 SUMMARY

In this chapter we have described the techniques and standards that are enabling digital video applications. Analog video can be represented in a digital form. This has a number of advantages, including the ability to process and transmit the information using digital computers, processors, and networks. However, the very high bit rates of digital video signals mean that the information must be compressed before it can be processed and transmitted over practical communication networks.

Many different techniques for compressing (coding) digital video have been developed in recent years. All of these techniques exploit the inherent spatial and temporal redundancy of a video sequence in order to achieve compression. We described the main steps involved in image and video compression, concentrating in particular on the techniques adopted within the international standards for video coding.

Several international standards have been issued in the last five years that provide the means of encoding still and moving video information. The JPEG standard is designed for still image coding and has become a popular means of compressing and storing digitized photographic images. The H.261 standard is part of the H.320 videoconferencing standards and is optimized for two-way videoconferencing applications. A system based on H.261 can support video communications of a basic quality over an ISDN connection. The MPEG1 and MPEG2 standards have been developed to support a wide range of digital video applications, including home entertainment and broadcast applications. They are designed to support video coding at bit rates above 1 Mbps at a comparable quality to existing analog video and television. The overall requirements of video communication systems are being addressed by the DAVIC forum, which

will release guidelines in the form of specifications that should provide a framework for many of the emerging digital video applications.

Video coding is still an important area for research and development. Low bit rate applications are not supported by the existing standards and the ITU-T and ISO are developing standards in this area. The H.263 standard (currently at its draft stage) is based on H.261 and provides higher quality video communications at bit rates below about 64 Kbps. The MPEG4 standard that is currently under development is likely to provide a range of tools to support more advanced video communication applications. Research is underway to develop ways of optimizing the existing video coding standards and to integrate these standards into practical video communication systems.

In later chapters, we will examine some of the problems associated with transmitting coded video data through practical communication networks. We will discuss the quality of service required by coded video and introduce some techniques that might be used to meet this requirement.

References

[1] "Encoding parameters of digital television for studios," CCIR Recommendation 601, XI, pp. 95–104.

[2] Netravali, A. N., and B. G. Haskell, *Digital Pictures: Representation, Compression and Standards*, 2nd Ed., Chapter 3, Plenum Press, 1995.

[3] Jayant, N. S., and P. Noll, *Digital Coding of Waveforms*, Englewood Cliffs, NJ: Prentice-Hall, 1984.

[4] Ahmed, N., T. Natarajan, and K. R. Rao, "Discrete Cosine Transform," *IEEE Trans. on Computers*, Jan. 1974, pp. 90–93.

[5] Chen, W. H., C. Harrison Smith, and S. C. Fralick, "A fast computational algorithm for the Discrete Cosine Transform," *IEEE Trans. on Communications*, Sept. 1977, pp. 1004–1009.

[6] Dinstein, I., K. Rose, and A. Heiman, "Variable block-size transform image coder," *IEEE Trans. on Communications*, Nov. 1990, pp. 2073–2078.

[7] Clarke, R. J., *Digital Compression of Still Images and Video*, Academic Press, 1995.

[8] ISO/IEC IS 10918, "Digital compression and coding of continuous-tone still images."

[9] Wallace, G. K., "The JPEG still picture compression standard," *Communications of the ACM*, Vol. 34/4, April 1991, pp. 31–44.

[10] ITU-T Recommendation H.320, "Narrow-band visual telephone systems and terminal equipment," 1993.

[11] ITU-T Recommendation H.261, "CODEC for audio-visual services at p × 64 Kbps," 1993.

[12] Liou, M. L., "H.261 overview," *Communications of the ACM*, Vol. 34/4, April 1991, pp. 60–63.

[13] ITU-T Draft Recommendation H.263, "Video coding for low bit rate communication," July 1995.

[14] ISO/IEC 11172, "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbps," 1993.

[15] Le Gall, D., "MPEG: A video compression standard for multimedia applications," *Communications of the ACM*, Vol. 34/4, April 1991, pp. 47–58.

[16] ISO/IEC 13818-1, ITU-T Recommendation H.222.0, "Generic coding of moving pictures and associated audio, Part 1: Systems."

Snap - Exhibit 1010
Snap v. Nokia - IPR2025-01439
Page 60 of 78

[17] ISO/IEC 13818-2, ITU-T Recommendation H.262, "Generic coding of moving pictures and associated audio, Part 2: Video."
[18] ISO/IEC 13818-3, "Generic coding of moving pictures and associated audio, Part 3: Audio."
[19] Pereira, F., "MPEG4: A new challenge for the representation of audio-visual information," *Proc. Int. Picture Coding Symposium*, Melbourne, March 1996, pp. 7–16.
[20] Digital Audio Visual Council, DAVIC/212, "Third DAVIC call for proposals," Sept. 1995.
[21] Digital Audio Visual Council, "DAVIC 1.0 Specifications," Dec. 1995.

The Internet 4

4.1 INTRODUCTION

The Internet is a global communications network for interconnecting computers that allows computers to communicate even when they reside on different types of subnetworks. It links universities, research organizations, companies, government establishments, and homes. It provides a de facto standard set of protocols that allows all participating computers to interwork. A large number of application services exist, including file transfer, remote login, electronic mail, news groups, and the World Wide Web. These allow users to access services and data on computers throughout the Internet.

The Internet was designed to support traditional computer data traffic. However, in recent years work has started on the communication of real-time traffic such as video and audio. Although currently the requirements of real-time traffic are not supported to a satisfactory level, the Internet is evolving to meet the needs of real-time traffic. In this chapter, we review the history and architecture of the Internet. Then, we consider the transmission of video and describe some of the emerging work that will allow the successful transmission of real-time traffic.

The Internet grew out of a research network called ARPANET created by the U.S. Department of Defense Advanced Research Projects Agency (ARPA) in 1969. Originally, ARPANET used point-to-point leased line connections and not packet switching. ARPA funded research into packet switching and between 1980 and 1983 the ARPANET converted to using packet switching with the now famous TCP/IP protocols, which was the start of true internetworking technology. TCP/IP protocols rapidly spread throughout the academic establishment as part of the UNIX operating system available at low cost to university researchers from the University of California at Berkeley. In the mid 1980s, the U.S. Department of Defense moved its military operations to the separate MILNET, which left the original ARPANET to the researchers and academics.

Snap - Exhibit 1010
Snap v. Nokia - IPR2025-01439
Page 62 of 78

In 1990, the ARPANET was dissolved and management of the Internet was passed to the U.S. National Science Foundation, and ARPANET became NSFNET.

No single organization manages the whole Internet, but rather the Internet consists of many interconnected networks managed by different organizations throughout the world, including

- NORDUNET in Sweden;
- INRIA in France and Switzerland;
- EASInet/CERN in Switzerland;
- DNF in Germany;
- JANET (Joint Academic Network) in the United Kingdom.

One group that has a great deal of influence over the development of the Internet is the Internet Society (ISOC). This is a nonprofit professional society that promotes research into multiprotocol internetworking. Research and development on the Internet is performed in a distributed manner using the Internet itself to coordinate activities. The Internet Activities Board (IAB) provides a focus for this research. Each member of the IAB chairs an Internet engineering task force (IETF), responsible for investigating a particular issue. Members of the IAB recruit volunteers to participate in the research. Research reports, protocol proposals, and protocol standards are published on the Internet as technical reports referred to as requests for comments (RFCs). The network information center (NIC) distributes RFCs electronically using the Internet. These RFCs can be accessed most easily using a worldwide web browser. The RFCs can be located by using one of the many search engines and searching for the keyword RFC.

4.2 THE INTERNET ARCHITECTURE

The Internet is not a new kind of physical network, but rather the technology to interconnect existing subnetworks. A number of interconnected subnetworks are shown in Figure 4.1. These subnetworks may be local area networks (LANs) such as Ethernet (ISO 8802/3) or the token ring (ISO 8802/5) and/or wide area networks (WANs) such as X.25, leased lines, satellite links, or the part of the original ARPANET. Separate subnetworks are interconnected by gateways or routers. These gateways route packets based on their destination network rather than their destination host address.

4.2.1 Gateways or Routers

In terms of the International Standards Organization (ISO) reference model for open systems interconnection (OSI), gateways perform routing at the network

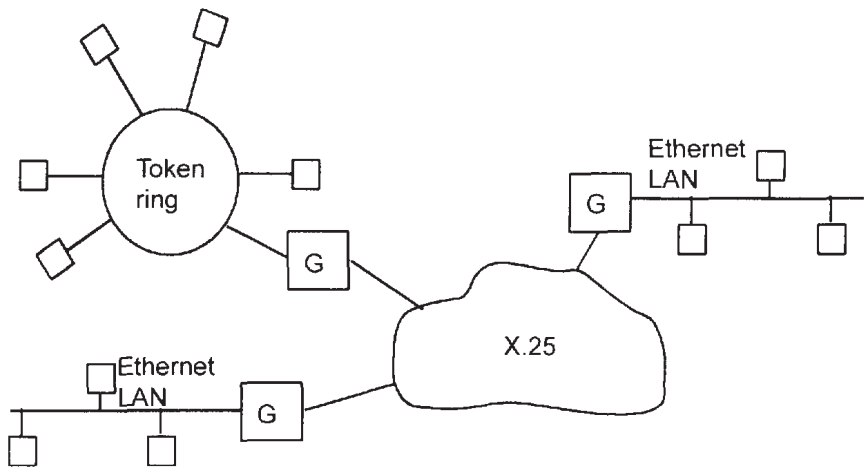

Figure 4.1 Part of the Internet.

layer. However, strictly speaking the Internet technology predates the ISO OSI model, so Internet protocols are not international standards. However, the Internet protocols are defacto standards and since the ISO OSI model provides a convenient framework for classifying the protocols, it is common to discuss Internet protocols in the context of the ISO OSI model. Therefore, an Internet address can be considered a network-level address. Gateways are the devices that are used to interconnect separate subnetworks. Descriptions of the operation of the TCP/IP protocols and gateways are provided in [1] and [2].

4.2.2 Internet Addresses

In order to be able to send data across the Internet, it is necessary to provide a global method of addressing. Each computer (or host) is assigned a unique 32-bit Internet (IP) address. Each address consists of both a network identifier and a host identifier. The network identifier indicates a particular subnetwork and the host identifier indicates a particular host on that subnetwork. There are three classes of addresses, referred to as class A, B, and C and illustrated in Figure 4.2.

The NIC (a central authority) assigns the *netid* part of the Internet addresses. Gateways use the *netid* field to route data packets. That is, they route using only the destination subnetwork. A *netid* of all 0s is used to indicate *this* network. Internet addresses are listed in dotted decimal form. Each number represents one byte of the 32-bit address. The different classes of address are assigned according to availability and the size of the subnetwork. A large subnet-

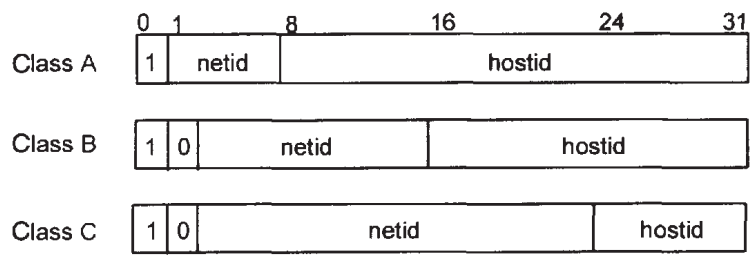

Figure 4.2 IP Addresses.

work with many host machines may be allocated a class A address, whereas a small subnetwork may be allocated a class C address.

4.2.3 The Domain Name System

Users prefer to use meaningful symbolic names rather than numeric IP addresses. These symbolic names are referred to as high-level names and must be translated into the low-level IP addresses. The original Internet used a flat name space where symbolic names were assigned by the Internet NIC. A flat name space was found to be not suitable for a large network and a hierarchical name space has been adopted as a replacement.

The hierarchical name space adopted in the Internet is partitioned into separate domains, and authority for assigning names in each partition is passed to a separate designated authority. Each partition is responsible for its own name space. Authority may be subdivided down to the level of individual sites where a site authority is responsible for assigning names of specific machines on a particular site.

An example of an Internet domain name is

eee.rgu.ac.uk

This is the address of the electrical and electronic engineering department (eee), the Robert Gordon University (rgu), Academia (ac), United Kingdom (UK).

In order for the IP protocol to operate, domain names must be translated to IP addresses. This is carried out automatically using the services provided by computers that act as domain name servers (DNS). This is performed by first asking a local server to look up the address. There are three possibilities:

1. The local server knows the address because the address is in the local server's database. For example, a local server in the School of Electrical

and Electronic Engineering knows the addresses of all machines in this department.

2. The local server knows the address for this particular name because someone else has recently asked for this address. The local server maintains a cache of previously resolved name and address pairs.
3. The local server is not able to translate the name, but it knows the address of the server at the next level of the addressing hierarchy and so it asks this server. This process repeats until a server is able to resolve the address.

4.3 INTERNET PROTOCOLS

Figure 4.3 illustrates the loose relationship between the Internet protocols and the ISO OSI reference model. However, it is important to realize that as the TCP/IP protocol suite predates ISO OSI, it is not strictly speaking conforming to the reference model. The physical and data link layers are part of the particular subnetworks. For example, the carrier sense multiple access/collision detection (CSMA/CD) Ethernet protocol forms the media access control (MAC) sublayer of the data link layer.

LAYER 7	Application	X. Windows, SNMP, TELNET FTP, World wide web ...
LAYER 6	Presentation	
LAYER 5	Session	
LAYER 4	Transport	Transmission Control Protocol (TCP) or User Datagram Protocol (UDP)
LAYER 3	Network	Internet Protocol (IP)
LAYER 2	LLC MAC	ISO 8802/2 LLC ISO 8802/3 CSMA/CD
LAYER 1	Physical	ISO 8802/3
	ISO OSI	Internet Host on 'Ethernet'

Figure 4.3 Relating the Internet protocols to ISO OSI.

4.3.1 Internetworking Protocol (IP)

The Internetwork protocol (IP) [3] resides in the network layer. The IP is responsible for routing IP datagrams (packets) through the Internet. The function of an IP router is illustrated in Figure 4.4. The IP provides a connectionless datagram delivery service. This implies that IP does not create end-to-end connections, but rather treats every IP packet independently. This is analogous to the way the postal system routes letters. Each IP datagram contains the full IP destination address. When being carried across subnetworks, IP datagrams are carried within the payload of the subnetwork packets. This is referred to as *encapsulation*. The format of the IP datagram is shown in Figure 4.5.

A router implementing the IP protocol is capable of receiving IP datagrams, examining their addresses, and choosing a suitable path or route over which to forward the packets. Routing can be either direct or indirect. Direct routing requires delivery of an IP datagram between two machines on a single physical network and does not involve gateways; the sender encapsulates the datagram in a subnetwork frame and sends it directly to the destination. To determine

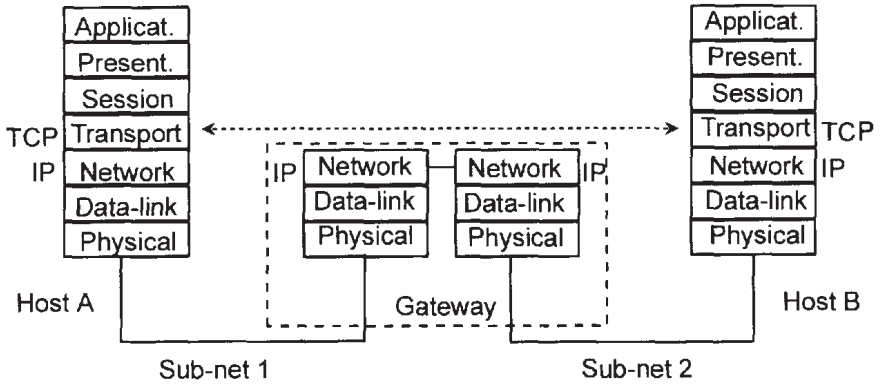

Figure 4.4 Function of a router.

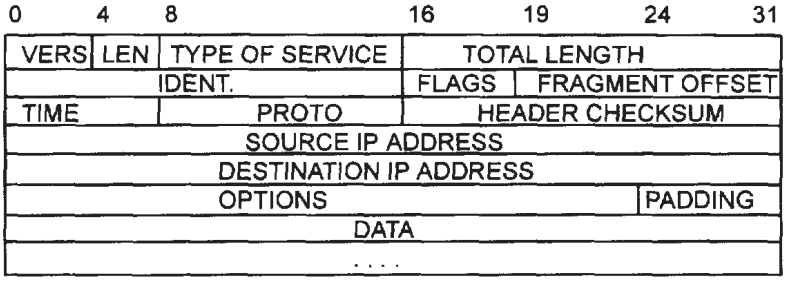

Figure 4.5 IP datagram format.

if direct routing is appropriate, the sender extracts the network portion of the Internet address and compares it to its own network address. Indirect routing requires that datagrams be sent to a gateway on the subnetwork and then on from gateway to gateway until they reach a gateway that can deliver the datagram directly.

Each machine contains an Internet routing table that contains information about possible destinations. The machine extracts the network portion from an IP datagram in order to make routing decisions. The routing table contains a list of pairs (N,G) where G represents the gateway through which destination network N can be reached. One technique that is used to keep routing tables small is to consolidate many entries into a default case. An example internetwork joined by gateways is shown in Figure 4.6. This shows three subnetworks that are joined by gateways labeled as A, B, and C. The corresponding routing table for gateway B is shown in Table 4.1.

In order to prevent continuously circulating datagrams, which would occur if routing tables are incorrectly configured, each IP datagram has a time-to-live field that decrements each time the datagram is forwarded. If this field reaches zero, the datagram is deleted.

4.3.2 Mapping Internet Addresses to Physical Addresses

A problem that still exists for routing IP datagrams across subnetworks concerns the translation of an IP address into the appropriate physical address. The

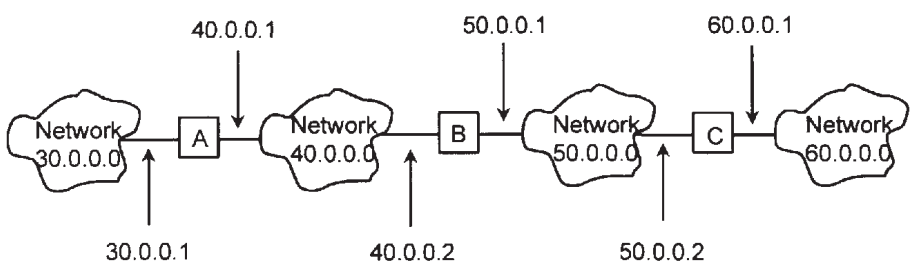

Figure 4.6 Routing with routing tables (gateway G).

Table 4.1
IP Routing

To Reach Hosts on Network	*Route to . . .*
40.0.0.0	Direct
50.0.0.0	Direct
30.0.0.0	40.0.0.1
60.0.0.0	50.0.0.2

physical address is specific to a particular network card on the subnetwork. In the case of ISO 8802/3 *Ethernet* LAN, the address resolution protocol (ARP) is used to translate IP addresses into physical addresses. The relationship of ARP to the other Internet protocols is illustrated in Figure 4.7. The IP address is represented by the "*" and the physical layer *Ethernet* address is represented by "@." The translation is maintained with a lookup table, called an ARP table. It should be noted that there is no way of automatically translating between these two address types since IP addresses are assigned arbitrarily to particular machines. If a machine is moved to a different part of the Internet then its IP address may change.

The ARP protocol must update a table referred to as an ARP table. An example of an ARP table is Table 4.2, which relates IP addresses to the physical *Ethernet* addresses.

Initially, this table is empty. When a particular destination IP address does not exist in the ARP table, the following occurs:

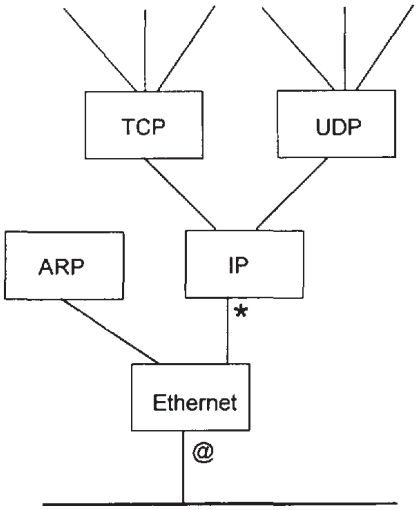

Figure 4.7 Internet protocols ("Ethernet" host).

Table 4.2
ARP Cache

IP Address	*Ethernet Address*
227.1.2.1	04-00-39-00-2F-C3
227.1.2.3	04-00-5C-19-A1-22
227.1.2.4	04-00-12-99-BC-41

1. The outgoing IP packet is queued.
2. The ARP protocol on the sending machine generates an ARP request that contains the destination IP address.
3. This ARP request is sent as a broadcast Ethernet packet to all stations on the subnetwork.
4. The ARP protocols in all stations receive this packet and the appropriate station recognizes its own address and responds with an ARP response that contains the appropriate Ethernet address.
5. The Ethernet address in the ARP response is used to update the ARP table.
6. The Ethernet address in the ARP response is used to send the queued IP packet.

4.3.3 TCP and UDP

The transport protocol uses the services provided by the network layer (located in the IP protocol) to provide the users with an end to end service with the required quality of service.

The user datagram protocol (UDP) [4] extends the unreliable datagram service to the user. This provides users with little more than access to the underlying IP datagram service, with little of the functionality that is usually associated with a transport protocol. Connections cannot be set up, but rather each packet is treated as an independent datagram. If a datagram is lost in transit, no action is taken and error recovery (if needed) is left to the application program, which is run over UDP.

The transmission control protocol [5] (TCP) enhances the unreliable datagram service provided by the IP in order to produce a reliable end-to-end stream delivery of packets. In order to do this, TCP does the following:

1. Opens a connection between the two communicating machines across the network;
2. Uses a technique referred to as positive acknowledgment with retransmission (PAR) to overcome the problem of lost packets.

The TCP performs most of the functions expected of an ISO OSI transport protocol, although TCP is not an international standard. The TCP is much more widely used than the official ISO protocol and there is now considerable doubt that the official ISO protocol will replace TCP. The TCP deals with the transmission of a stream of bytes from the transmitter to the receiver. The byte stream is delivered without errors at the receiver in the same order that the bytes were transmitted. In order to perform reliable stream delivery, the TCP shields the

user from dealing with packet boundaries and automatically recovers from packet loss. Since the IP provides only an unreliable datagram service and TCP packets are carried within the payload of IP datagrams, the possibility exists for TCP packets to be lost. The TCP detects this packet loss and recovers the lost packets using a PAR strategy. When a packet is lost, a timeout followed by a retransmission is used, as shown in Figure 4.8.

In order to make this scheme work well, a good estimation of the round-trip delay time is needed. However, this varies depending on the path taken and the buffering delays incurred within the network. The TCP makes an estimate of the round-trip delay by keeping a weighted average (W_aver) of the round-trip delay times and updating this average slowly with a contribution from the latest measurement of the round trip delay time (New_Round_Trip_Time):

$$W_aver = (q \times W_aver) + ((1 - q) \times \text{New_Round_Trip_Time}) \qquad (4.1)$$

where q is the weighting factor, and varying q causes the weighted average to change more quickly or more slowly. The value of q is usually chosen (close to 1) such that the protocol is relatively immune to occasional IP packets that incur long delays. The TCP uses a sliding window algorithm for the purposes of efficient transmission and flow control. A sender's sliding window is imple-

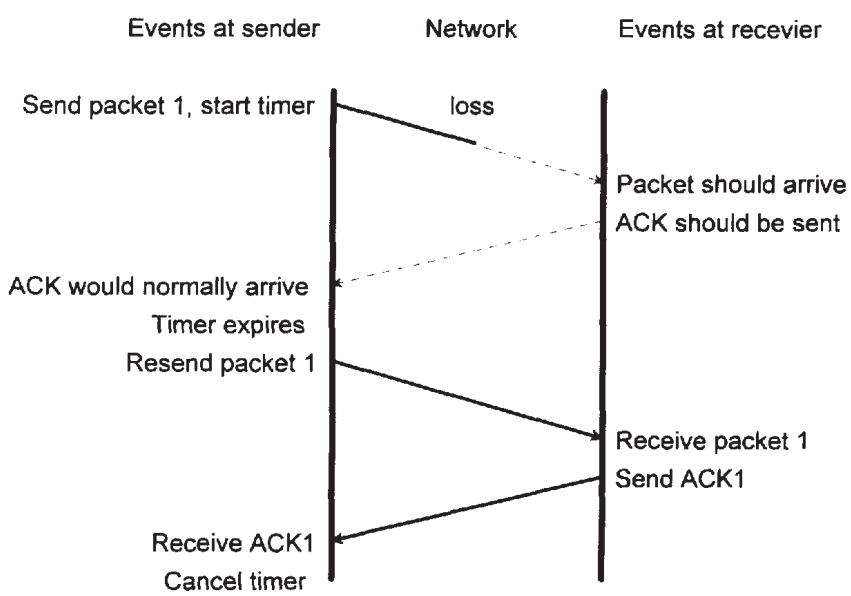

Figure 4.8 TCP PAR operation.

Snap - Exhibit 1010
Snap v. Nokia - IPR2025-01439
Page 71 of 78

mented to allow the sender to transmit a number of packets without waiting for acknowledgments. The window size determines the number of packets that can be transmitted with outstanding acknowledgments. For example, in Figure 4.9 a sliding window of size 8 is illustrated by the dark box. Figure 4.9(a) illustrates that packets with sequence numbers 1 to 8 have been transmitted but not yet acknowledged. Since 8 is the maximum window size, no further packets can be transmitted until an acknowledgment has been received. In Figure 4.9(b), an acknowledgment has been received for packet 1 allowing the sliding window to be advanced and packet 9 to be transmitted. The sliding window allows the transmission of packets to be pipelined as shown in Figure 4.10. For efficient transmission, the sliding window should be large enough to hold all the packets that can be transmitted during one round-trip delay time. In order to identify individual packets and associate acknowledgments with the appropriate packet, sequence numbers are assigned to all packets. Using the sliding window algorithm, the network can be kept saturated with packets, allowing the maximum possible throughput to be achieved.

Flow control is the name given to the process of controlling the flow of data so that a fast sender does not overrun a slow receiver. TCP also uses the sliding window to implement flow control, but the simple scheme described above is extended by allowing the size of the sender's window to be varied. Each acknowledgment sent back from the receiver contains a window advertisement as well as specifying the number of bytes that have been received. The window advertisement specifies how many additional bytes the receiver is prepared to accept. This is controlled by the amount of space left in the receiver's buffer for incoming data. In response to an increased window advertisement, the sender increases the size of its window and proceeds to transmit more data to the receiver. In response to a decreased window advertisement, the receiver decreases the size of its window. A receiver should not contradict previous advertisements by shrinking the size of the window past previously acceptable positions in the byte stream.

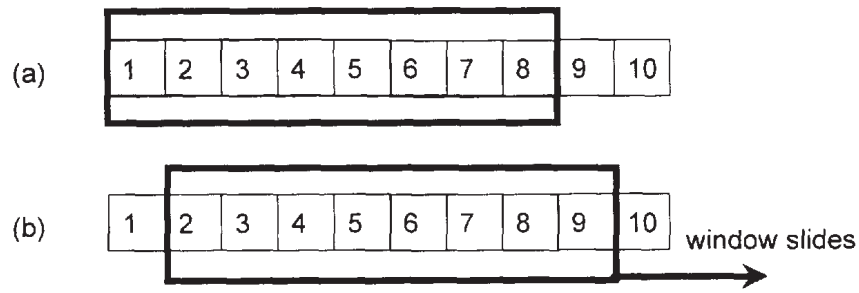

Figure 4.9 The sliding window algorithm.

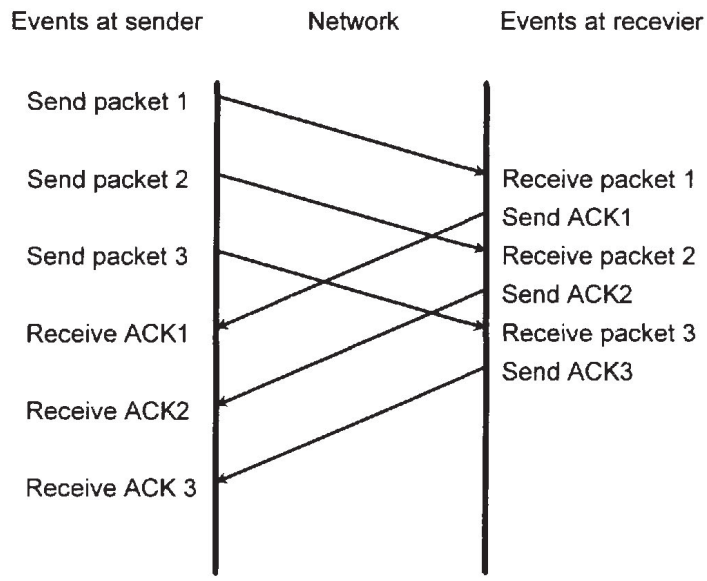

Figure 4.10 Pipelining packet transmission.

4.4 CONGESTION IN THE INTERNET

The Internet operation is based on datagram delivery. Therefore, there is no bandwidth reserved for specific connections, but rather there is a free-for-all. The capacity of limited bandwidth connections within the Internet may therefore be exceeded. If this happens, queues within the network will grow, increasing the round-trip delay through the network until eventually the queues are full, at which point packets will be dropped by the network. The loss of packets causes further retransmissions and additional traffic that may compound the situation further, causing congestion collapse. In a congestion collapse situation, all connections will experience a very low (almost zero) effective throughput. The causes of congestion are thoroughly discussed in [6].

In order to avoid congestion collapse, special additional algorithms have been built into the TCP [7]. Under equilibrium conditions, an additional packet will only be put into the network in response to an acknowledgment indicating that a packet has been removed. This is referred to as the principle of packet conservation. Congestion will only occur if packet conservation does not occur. The circumstances when this occurs are discussed at length in [7].

4.5 THE INTERNET AND REAL-TIME TRAFFIC

In recent years it has been the aim of many in the Internet community to develop the Internet to support real-time traffic. The IETF has conducted initial large-

scale experiments by broadcasting their meetings across the Internet. These broadcasts include real-time video and audio traffic. These broadcasts rely on the additional capability of IP multicasting, which is not (yet) generally available in commercial routers. This multicasting capability is provided using the M-BONE, which is described in more detail below. In order to receive these broadcasts, workstations equipped with multimedia hardware and running multimedia application software are required. Both the software and multimedia hardware are now relatively inexpensive. These initial experiments demonstrated that real-time applications do not always work well over the Internet. At times of congestion, variable queuing delays and packet losses can produce very low quality video and audio. This result is not surprising since the traditional Internet built on top of the IP only provides a best effort service.

To make the Internet suitable for real-time traffic, a number of changes to the underlying infrastructure are required. As well as a multicast capability, it is necessary to provide a real-time traffic stream with a guaranteed quality of service (QoS) in the form of bandwidth, packet loss, and end-to-end delay guarantees. A useful overview of proposed changes to the Internet infrastructure is described in [8]. As well as providing bandwidth guarantees for real-time traffic, the ability to control the allocation of bandwidth between different traffic classes is useful to many network operators.

The proposed changes to the Internet affect the fundamental service model of the Internet, which has until now been based on a best effort datagram delivery service using IP. It is proposed to minimize the effect of the proposed changes by supplementing rather than replacing the basic service model that is currently in place. The model proposed includes two service classes for real-time traffic, referred to as guaranteed and predictive service, in addition to the existing best effort service class. The model utilizes controlled link sharing and includes multicast as well as unicast connections. In order to manage the link sharing, it is necessary to provide mechanisms for resource reservation and admission control.

The need to implement resource reservation implies a fundamental change to the Internet in that routers are then required to store flow-specific state information for specific user packet streams or flows. Much of the current robustness of the Internet comes from the absence of flow-specific state information in the routers and one of the challenges of supporting integrated services over the Internet is to maintain this robustness. To provide these extra facilities, the reference implementation framework proposed in [8] includes four additional components, referred to as the packet scheduler, the packet classifier, the admission controller, and the reservation setup protocol.

Traditionally, the IP protocol treats all packets equally using a strict first in, first out (FIFO) queuing discipline. The packet scheduler is implemented at the point where packets are queued for transmission and makes decisions

regarding when packets from each service class should be forwarded. Each incoming packet must be mapped into a service class. All packets in the same service class are treated identically by the scheduler. The classifier has the job of identifying the class to which a packet belongs. This can be achieved by examining the existing packet headers or, alternatively, by using some additional (yet to be defined) classification number. For example, packets belonging to a particular connection may be identified using the source and destination address pair.

The admission controller makes the decisions as to whether additional traffic flows can be supported by the network without impacting previous QoS guarantees. A traffic source negotiates a traffic contract with the network and the network either accepts or rejects the connection based on the stated QoS requirements. Once a connection has been accepted, the network will police each traffic flow to ensure that the traffic contract is not broken by the source. It is postulated that this function should be performed by the traffic scheduler.

In order to create and maintain state information relevant to each flow in the hosts and in the routers along the traffic path, a protocol is required. The RSVP protocol, which we discuss in Section 4.5.2, is being developed for this purpose. This protocol is required to communicate a set of parameters specifying the desired QoS and referred to as the *flowspec* from the host to the admission control. The *flowspec* is then used by the packet scheduler if the flow is accepted.

Figure 4.11 illustrates how a router could be designed to include the additional components necessary for supporting integrated services.

4.5.1 M-BONE (IP Multicast)

Transmitting traffic from one source to many destinations is referred to as multicasting. To multicast efficiently over the Internet, a method is needed whereby IP datagrams are replicated at appropriate points in the network. This functionality is not part of the existing IP version 4 protocol. It is planned to build a multicasting capability along with a larger addressing capability (larger than the 32 bits currently allocated for addressing) into the emerging version 6 of the IP protocol. In the meantime, a temporary extension to IP, referred to as the M-BONE, will be used to support multicasting.

The M-BONE is an extension to IP that originated from the first two IETF experiments in the transmission of audio and video. The M-BONE is an experimental IP multicast testbed operated on a cooperative volunteer basis. The M-BONE consists of islands that can, by running special multicasting software, directly support multicasting joined by point-to-point links called tunnels, which may consist of a number of standard IP routers that do not support multicasting. IP multicast packets are encapsulated into standard IP

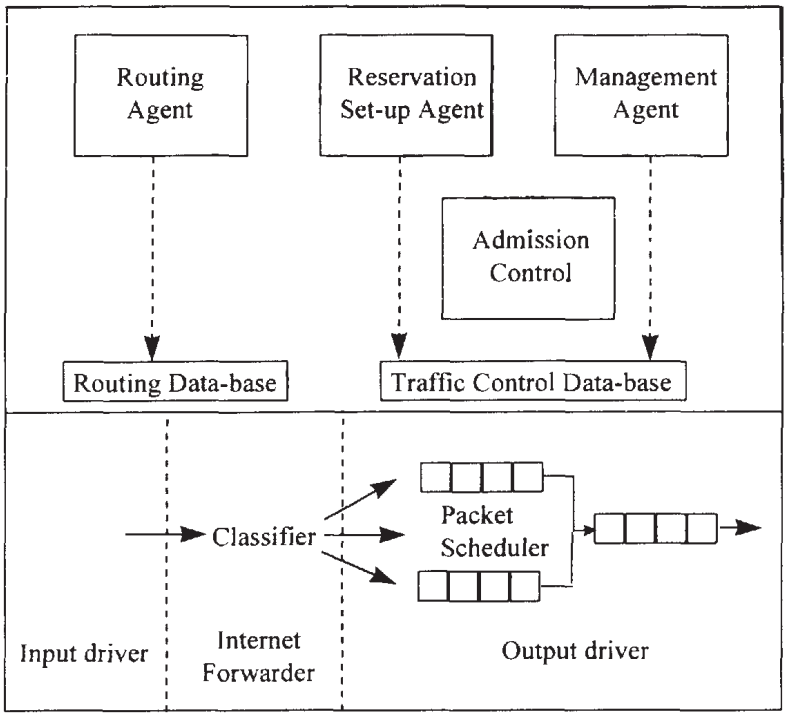

Figure 4.11 Proposed architecture for a router supporting integrated services.

packets for transmission through the tunnels. At either end of the tunnel will be an IP router with multicasting capability. To send a multicast packet through a tunnel, the multicast router will encapsulate the multicast packet by prepending an additional IP header onto the multicast packet. The additional IP header contains the address of the multicast router at the other end of the tunnel as a destination address and sets the IP protocol field to 4 to indicate that the next protocol is IP. Upon receiving an encapsulated IP multicast packet, a multicast router will strip off the additional header and forward the multicast packet as appropriate.

IETF meetings using audio and video are regularly broadcast over the M-BONE. The traffic generated is typically less than 500 Kbps during the conference. Audio is coded at 32 or 64 Kbps, and video at up to 128 Kbps. Other services tend to use much less bandwidth.

4.5.2 RSVP

RSVP forms a part of the proposed future integrated services Internet and allows applications to reserve resources along the data paths between the traffic source

and destinations. RSVP is implemented in the end systems and in each router along the path. The end systems talk to each of the routers along the path using RSVP to reserve network resources such as buffer space and bandwidth in order to ensure that the network can meet an agreed QoS. RSVP supports both multicast and unicast delivery of data. It supports multicast efficiently by treating dataflows as fundamentally unidirectional rather than bidirectional. RSVP also supports heterogeneous receivers, each with a different capability and therefore needing a different QoS. In order to support these heterogeneous receivers efficiently, RSVP makes the receivers responsible for requesting and reserving resources so that each receiver reserves a QoS appropriate to its particular requirements.

4.6 SUMMARY

The current Internet makes use of the IP protocol to provide a best effort, unreliable datagram delivery network. This makes the Internet an extremely robust network since state information does not have to be held in the routers. However, this best effort service is not suitable for real-time traffic such as video and audio, which requires bandwidth and delay guarantees for satisfactory operation. For this reason, a number of developments are underway to adapt the existing Internet to provide these capabilities. These developments are designed to be compatible with and work alongside the existing Internet protocols.

The key additional capabilities being developed are, firstly, an IP multicast capability to allow a single traffic source to transmit to many destinations simultaneously with packet replication occurring at the most appropriate points in the network and, secondly, a resource reservation protocol that will allow network resources to be reserved prior to data transmission, thereby allowing a requested QoS to be met by the network.

This new emerging integrated services Internet is capable of supporting real-time audio and video and providing a networking infrastructure capable of allowing effective multimedia communication between PCs or workstations to support applications such as video broadcasting, desktop videoconferencing, and video on demand.

References

[1] Comer, D., *Internetworking with TCP/IP*, Englewood Cliffs, NJ: Prentice-Hall, 1995.
[2] Socolofsky, Kale, "A TCP/IP Tutorial," *Internet Request For Comments 1180.*
[3] "Internet Protocol," *Internet Request For Comments 791 and 760.*
[4] "User Datagram Protocol," *Internet Request For Comments 768.*
[5] "Transmission Control Protocol," *Internet Request For Comments 793, 761 and 675.*

[6] Jain, R., "Congestion Control in Computer Networks: Issues and Trends," *IEEE Network Magazine*, Vol. 4, No. 3, May 1990, pp. 24–30.
[7] Jacobson, V., "Congestion Avoidance and Control," *Proc. of. ACM SIGCOMM '88*, Stanford, CA, Aug. 1988, pp. 314–329.
[8] Braden, R., D. Clark, and S. Shenker, "Integrated Services in the Internet Architecture: An Overview," *Internet Request For Comments 1633*.