

ABSTRACT

Title of Dissertation: DYNAMIC RESOURCE ALLOCATION
FOR MULTIUSER VIDEO STREAMING

Guan-Ming Su, Doctor of Philosophy, 2006

Dissertation directed by: Professor Min Wu
Department of Electrical and Computer Engineering

With the advancement of video compression technology and wide deployment of wired/wireless networks, there is an increasing demand of multiuser video communication services. A multiuser video transmission system should consider not only the reconstructed video quality in the individual-user level but also the service objectives among all users on the network level. There are many design challenges to support multiuser video communication services, such as fading channels, limited radio resources of wireless networks, heterogeneity of video content complexity, delay and decoding dependency constraints of video bitstreams, and mixed integer optimization. To overcome these challenges, a general strategy is to dynamically allocate resources according to the changing environments and requirements, so as to improve the overall system performance and ensure quality of service (QoS) for each user.

In this dissertation, we address the aforementioned design challenges from a resource-allocation point of view and two aspects of system and algorithm designs, namely, a cross-layer design that jointly optimizes resource utilization from physical layer to application layer, and multiuser diversity that explores the source and channel heterogeneity among different users. We also address the impacts on systems caused by dynamic environment along time domain and consider the time-heterogeneity of video sources and time-varying characteristics of channel conditions. To achieve the desired service objectives, a general resource allocation framework is formulated in terms of constrained optimization problems to dynamically allocate resources and control the quality of multiple video bitstreams.

Based on the design methodology of multiuser cross-layer optimization, we propose several systems to efficiently transmit multiple video streams, encoded by current and emerging video codecs, over major types of wireless networks such as 3G cellular system, Wireless Local Area Network, 4G cellular system, and future Wireless Metropolitan Area Networks. Owing to the integer nature of some system parameters, the formulated optimization problems are often integer or mixed integer programming problem and involve high computation to search the optimal solutions. Fast algorithms are proposed to provide real-time services. We demonstrate the advantages of dynamic and joint resource allocation for multiple video sources compared to static strategy. We also show the improvement of exploring diversity on frequency, time, and transmission path, and the benefits from multiuser cross-layer optimization.

DYNAMIC RESOURCE ALLOCATION
FOR MULTIUSER VIDEO STREAMING

by

Guan-Ming Su

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2006

Advisory Committee:

Professor Min Wu, Chair / Advisor
Professor K. J. Ray Liu
Professor Prakash Narayan
Professor Rama Chellappa
Professor Martin Loeb

©Copyright by
Guan-Ming Su
2006

DEDICATION

To my parents, my wife, Jing-Wen, and my daughter, Olivia.

ACKNOWLEDGEMENTS

On the way to attain advanced knowledge and develop independent problem-solving ability, two essential factors play major roles: creative environment filled with inspirational discussion and valuable help from generous people. Luckily, during the study in University of Maryland College Park, I benefit a lot from both of them and there are many people I would like to acknowledge.

First, I would like to express my sincere gratitude to my advisor, Prof. Min Wu, for her guidance and support during my study in University of Maryland. Under the nourish environment created by her and unlimited constructive advice, I learned to be open-minded to view things and to be collaborative with colleagues. With her support and encouragement, I have my own freedom to pursuit the research areas that I am passionate about. With her help, I developed the ability to conduct meaningful research.

I appreciate Prof. K. J. Ray Liu for generous support and priceless advice during my whole Ph.D. study. With his encouragement, I had the opportunity to contribute to a new Capstone course on multimedia signal processing, and gained experience related to multimedia. With his support, I obtained precious knowledge and incalculable help from his research group.

I thank Dr. Han Zhu for his endless suggestions and fruitful discussions when we collaborated on several projects. His professional knowledge and experiences enhanced the depth and width of the projects. With his tremendous help, our projects have advanced to a more concrete level.

I also would like to thank Prof. Rama Chellappa, who motivated me doing research in image/video related areas while I took his image processing related courses. I also would like to thank Prof. Prakash Narayan, who introduced the fantastic area in communication areas while I took his multiuser communication course. I also would like to thank Prof. Martin Loeb for serving in my dissertation committee and providing many useful suggestions for my dissertation.

I appreciate Dr. Peerapol Tinnakornsriruphap and Dr. Christopher Lott for providing me an opportunity to expand my vision through a collaborative project of algorithm development for video telephony over CDMA-2000 1xEV-DO during my internship at Qualcomm in summer 2005.

I owe my gratitude to my friends at University of Maryland. I would like to thank Dr. Andres Kwasinski, Dr. Yan Sun, Dr. Hong Zhao, Dr. Charles Pandana, Dr. Wipawee Siriwongpairat, Dr. Onur kaya, Yinian Mao, Meng Chen, Shan He, Hongmei Gou, Ashwin Swaminathan, Thanongsak Himsoon, Ahmed Sadek, Wei Yu, Zhu Ji, Tzung-Hsien Ho, Wei-Yen Chen, Huanfeng Ma, Seung-Jong Baek, and Kyechong Kim for providing many constructive suggestions for my research.

Last but not the least, I would like to thank my family for their support and understanding during my graduate study.

TABLE OF CONTENTS

List of Tables	viii
List of Figures	ix
1 Introduction	1
1.1 Motivations	1
1.2 Related Prior Work	2
1.3 Thesis Organization and Contributions	5
2 System Framework	7
2.1 System Overview	7
2.2 Wireless Networks	9
2.3 Video Coding	13
2.4 Service Objectives	15
2.5 Design Principles	16
3 Low-Delay Bandwidth Allocation	18
3.1 Introduction	18
3.2 FGS Rate-Distortion Model and Similarity	21
3.2.1 Intra-frame Rate-Distortion Model	22
3.2.2 Similarity in Inter-frame R-D Characteristics	23
3.3 Low-Delay Bandwidth Resource Allocation for Single User	24
3.3.1 System Constraints	24
3.3.2 Criteria for Visual Quality	28
3.3.3 Problem Formulation	29
3.3.4 The Proposed Resource Allocation Algorithm	31
3.4 Low-Delay Bandwidth Resource Allocation for Multiple Users	38
3.4.1 System Constraints	38
3.4.2 The Proposed Resource Allocation Algorithm	41
3.4.3 Differentiated Service	43
3.5 Simulation Results	44
3.5.1 Simulation Setup	44
3.5.2 Simulation Results for Single-User Case	46

3.5.3	Simulation Results for Multiuser Case	50
3.6	Chapter Summary	54
4	Video over Wireless Downlink	55
4.1	Introduction	55
4.2	System Description	56
4.2.1	Video Source Codec Subsystem	57
4.2.2	OFDMA Subsystem	60
4.2.3	EWV Video over OFDMA	65
4.3	Optimization in Resource Allocator: Focusing on Fairness	66
4.3.1	Formulation of the Fairness Problem	67
4.3.2	Proposed Algorithm for Fairness	68
4.4	Optimization in Resource Allocator: Focusing on Efficiency	74
4.4.1	Formulation of the Efficiency Problem	75
4.4.2	Proposed Algorithm for Efficiency	75
4.5	Network-Level Service Objectives	80
4.5.1	Tradeoff between Fairness and Efficiency	80
4.5.2	Unequal Error Protection	81
4.6	Simulation Results	82
4.6.1	Performance of the Fairness Algorithm	83
4.6.2	Performance of the $F_x E_y$ Algorithm Family	88
4.7	Video over Interference-limited Networks	93
4.7.1	System Description	94
4.7.2	Distortion Management Algorithm	95
4.7.3	Simulation Results	99
4.8	Chapter Summary	101
5	End-to-end Video Conferencing	103
5.1	Introduction	103
5.2	System Description	105
5.2.1	IEEE 802.11a	106
5.2.2	Application Layer FEC	109
5.2.3	Video over WLAN	112
5.3	Joint Uplink and Downlink Optimization for Single-Cell Case	115
5.3.1	Problem Formulation	115
5.3.2	Proposed Algorithm	118
5.4	Joint Uplink and Downlink Optimization for Multi-Cell Case	122
5.4.1	System Framework	123
5.4.2	Problem Formulation	124
5.4.3	Proposed Algorithm	125
5.5	Simulation Results	126
5.5.1	Simulation Setup	128

5.5.2	Performance Criteria	129
5.5.3	Single-Cell Case	130
5.5.4	Multiple-Cell Case	135
5.6	Chapter Summary	139
6	PDMA-Based Error Protection	140
6.1	Introduction	140
6.2	Multi-Source Multi-Path Transmission	143
6.2.1	System Overview	143
6.2.2	Cross-Path FEC for TDMA and PDMA Scheme	145
6.3	PDMA Error Protection for Video System	151
6.3.1	Multi-Stream Video System	151
6.3.2	Problem Formulation	153
6.3.3	Proposed Algorithm	155
6.3.4	Heterogeneous Channel Conditions Along Each Path	156
6.4	Simulation Results	158
6.4.1	Simulation Setup	159
6.4.2	Performance of Homogeneous Channel Conditions	160
6.4.3	Performance of Heterogeneous Channel Conditions	165
6.5	Chapter Summary	167
7	Conclusions and Perspectives	168
A	Proofs and Review	173
A.1	The Impact of Budget Factor on Perceptual Quality	173
A.2	Review of An Alternative Sliding-Window Algorithm	175
A.3	Performance Gain for PDMA over TDMA	177
A.4	Problem Formulation of TDMA-CP	179
	Bibliography	181

LIST OF TABLES

3.1	Summary of notations for low-delay bandwidth allocation	26
3.2	The proposed single-user resource allocation algorithm	37
3.3	The proposed multiuser resource allocation algorithm	45
4.1	SNR and transmission rate under different BER	64
4.2	GOF R-D used in each transmission interval	69
4.3	OFDMA resource allocation and feasibility check	71
4.4	Coding pass cluster assignment	73
4.5	Proposed algorithm to minimize overall distortion	79
4.6	Unequal error protection versus using equal error protection	93
5.1	Physical layer mode for 802.11a	107
5.2	Proposed algorithm for T-D function	121
5.3	Selected transmission modes for FGS layer	131
6.1	Proposed algorithm for cross-path PDMA-based scheme	157
6.2	Selected parameters of TDMA-EP	162
6.3	Selected parameters of TDMA-CP	162
6.4	Selected parameters of PDMA-CP	163

LIST OF FIGURES

2.1	Framework of multiuser cross-layer video transmission system over wireless networks	8
2.2	Illustration of multiuser resource allocation for 3G, wireless LAN, 4G, future Wireless LAN/MAN	10
3.1	Inter-frame similarity of FGS rate-distortion characteristics	23
3.2	Block diagram of a single-user video streaming system	25
3.3	Impact of β and w_p on visual quality	33
3.4	Selection of w_p according to the encoder buffer occupancy.	35
3.5	Block diagram of a multiuser video streaming system	39
3.6	Comparison of CBR, S-SWLF, and S-LDLF for single-user case	47
3.7	Frame-by-frame PSNR comparison of the CBR, S-SWLF, and the proposed S-LDLF approaches	48
3.8	Comparison of CBR, M-SWLF, M S-LDLF, and M-LDLF systems under uniform service for all users for multiuser case	49
3.9	Frame-by-frame PSNR results of the first and tenth user in the M S-LDLF, M-SWLF, and M-LDLF systems	52
3.10	Results of a 32-user system using M-LDLF with differentiated service	53
4.1	Illustration of the relationship among coding pass, subband, and quality layer.	60
4.2	System block diagram	64
4.3	Flowchart of the proposed algorithm	68
4.4	Comparison for the $F_{16}E_0$ system providing uniform quality and differentiated service.	84
4.5	Subcarrier assignment for $F_{16}E_0$ system in each transmission interval	85
4.6	Frame-by-frame PSNR for the $F_{16}E_0$ system with uniform quality and with differentiated service.	86
4.7	avePSNR and stdPSNR results of the F_xE_y algorithm family and TDM algorithm.	89
4.8	Frame-by-frame PSNR for different algorithms of a 4-user system.	91
4.9	Performance comparison for the worst quality received among all users using the proposed algorithm and TDM algorithm.	91

4.10	Power and distortion convergence track vs. the number of assigned codes	96
4.11	Code limited case: optimal solutions	99
4.12	Power limited case: close to performance upper bounds	100
4.13	Performance comparison of BCP and greedy algorithms	101
5.1	System block diagram for single-cell case.	106
5.2	Error protection scheme for application layer FEC	111
5.3	Flowchart of the proposed wireless video system	113
5.4	The Time-Distortion function	122
5.5	System block diagram for multi-cell case	123
5.6	Proposed algorithm for multi-cell case.	127
5.7	Frame-by-frame PSNR for User 1 to User 4.	133
5.8	PSNR performance results for different number of users for single-cell case.	134
5.9	PSNR results for different number of users for two-cell case	136
5.10	PSNR results for different number of intra-cell calls for two-cell case with 8-users.	138
6.1	Multiple sources over a multiple-path system	144
6.2	Illustration of TDMA and PDMA-based multi-stream error protection	146
6.3	Performance gain using PDMA-based scheme over TDMA-based scheme for two-stream case	150
6.4	Performance gain using PDMA-based scheme over TDMA-based scheme for different number of streams	150
6.5	Block diagram of multi-stream video system using cross-path PDMA-based error protection	152
6.6	Frame-by-frame PSNR for different schemes	161
6.7	Performance comparison for different packet loss rate	164
6.8	Performance comparison for different bandwidth	164
6.9	Performance comparison for heterogeneous channel conditions for different bandwidth	166

Chapter 1

Introduction

1.1 Motivations

Over the past few decades, wireless communications and networking have experienced an unprecedented growth. With the advancement in video coding technologies, transmitting real-time encoded video programs over wireless networks has become a promising service for such applications as video-on-demand and interactive video telephony. In these applications, multiple real-time encoded video programs are transmitted to multiple users simultaneously by sharing resource-limited communication networks.

The challenges for transmitting multiple compressed video programs over wireless networks in real time lie in several aspects. First, wireless channels are impaired by detrimental effects such as fading and co-channel interferences (CCI). Second, radio resources such as bandwidth and power are very limited in wireless networks and should be shared among multiple users. In addition, unlike the transmission of generic data and voice, the rates of compressed video programs can be highly bursty due to the differences in video contents and intra/inter coding modes, which

complicates the source coding rate allocation. Moreover, optimizations in individual layers often depend on other layers' parameters (including continuous and integer values), and the systems may be subject to non-linear or/and non-convex constraints. As such, searching for optimal solutions to the formulated problem is often *NP* hard. Further, handling multiple video streams over a wireless system involves several important service objectives, such as system efficiency and individual fairness, and there are inherent tradeoffs among these objectives.

1.2 Related Prior Work

The resource allocations for maximizing overall throughput of generic data for multiple users over Code Division Multiple Access (CDMA) networks [40, 45, 103] and over Orthogonal Frequency Division Multiplexing Access (OFDMA) networks [14, 79, 101, 104, 107] have been widely studied. However, the multiuser wireless resource allocation problem often has resources with integer values. Through Lagrangian relaxation, an algorithm satisfying users' minimal rate requirement and minimizing the overall transmission power was proposed in [113]. To alleviate the high computational complexity, several suboptimal but computationally efficient algorithms for transmitting generic data were proposed in [17, 28, 29, 48]. Unlike generic data, compressed video sources exhibit different characteristics from generic data, for example, there is highly bursty rate from frame to frame and different compression complexity from one scene to another scene. Furthermore, a real-time streaming video system has a strict delay constraint that belated video data is useless for its corresponding frame and will cause error propagation for the video frames encoded predictively from this frame. Therefore, the radio resource allocation problem for transmitting video is more difficult than the problem for

transmitting generic data. There is a need of a resource allocation mechanism designed specifically to support transmitting multiple wireless video programs in real time and optimize the received video quality.

In the video communication literature, systems transmitting a single video program through wireless channels have been widely studied [11,15,23]. Traditionally, wireless data networks are designed in layers. By Shannon's Separation Theorem, source and channel coding can be designed separately while still achieving the optimality, if arbitrarily long delay is allowed. However, in most practical wireless networks, the assumptions in Shannon's Separation Theorem does not hold. For example, packets must be transmitted within delay constraints, especially for real-time video transmission. Therefore, streaming video systems can provide better end-to-end video quality by jointly considering source/channel rate adaptation and power allocation [44,55,85,86,117,126,127]. To improve the effective throughput, a wireless video streaming over wireless local area network (WLAN) was proposed in [58], using hybrid automatic repeat request (ARQ) with multiple descriptions in the application layer. Under the the current layered design network, system designers often perform cross-layer optimization to achieve system-wide optimality. A cross-layer design that jointly optimizes the resource allocation of all communication layers has been proposed to improve the video quality of single-user video streaming system [13,52,82,105].

Systems supporting multiple users, however, have more challenges than systems with single user. In a multiuser system, a limited amount of system resources is shared by multiple video streams for transmission. The transmission rates of different compressed video bitstreams vary among users and change over time. Such a system needs to efficiently allocate system resources to different video streams

to achieve the desired video quality. Because of different content complexity of video scenes, namely, the rates to achieve the same perceptual quality are content dependent, a multiuser video system may explore the content diversity among different video streams to efficiently utilize system resources and satisfy users' quality requirement. Joint multiuser video source coding over error-free channels subject to outbound bandwidth constraint has been proposed to leverage the diversity of video content to achieve more desired quality [9,108,114,124]. For communication over wireless networks, the channels are error-prone and the channel conditions experienced by different users are different and varying from time to time. Systems can exploit diversity, such as frequency, time, space, and multiuser, to take advantage of the random nature of radio propagation. System resources can be utilized in a more efficient way to improve the overall performance. In this dissertation, frameworks are proposed to jointly explore diversity in both source encoding and communication systems and allocate system resources dynamically. By doing so, we can improve the network performances and guarantee the quality of service (QoS) satisfactions for individual users.

A multiuser video transmission system should consider not only the reconstructed video quality of each individual user but also different perspectives from network-level point of view. In general, we can formulate the resource allocation problem as to optimize the network objective by allocating the resources across layers and among users subject to system constraints. Two essential network objectives, namely, *efficiency* and *fairness*, are considered in this dissertation. Efficiency concerns how to attain the highest overall video quality using the available system resources, and fairness concerns the video quality deviation among users who subscribe the same QoS level. There is a tradeoff between efficiency and fairness. We

will study how to attain the desired tradeoff in this dissertation.

Overall, this dissertation will address and overcome the aforementioned design issues and challenges via dynamic resource allocation. The proposed frameworks jointly adjust the system's parameters and utilize the limited system resources efficiently in the source coding and communication layers for multiple users. More specifically, we focus on two major aspects to optimize resource allocation, namely, *cross-layer design* and *multiuser diversity*, so as to accommodate a large number of users with acceptable received quality of service.

1.3 Thesis Organization and Contributions

This dissertation is organized as follows. Chapter 2 presents a generic framework for streaming multiple video programs over resource-limited networks. The available radio and video resources and the corresponding constraints are reviewed, and service objectives in both network level and individual-user level are discussed. This chapter lays out several general design principles for multiuser cross-layer video streaming system.

Based on the framework presented in Chapter 2, Chapter 3 studies the case of streaming multiple video programs over downlink error-prone channel. We perform resource allocation on video source coding rate and channel transmission rate to achieve the best tradeoff between the service objectives in the individual-user level. Experimental results in the chapter demonstrate the advantages of dynamically joint resource allocation for multiple video sources.

Chapter 4 extends the concept of dynamic resource allocation to a wireless video transmission system over downlink error-prone wireless networks. A framework with multiuser cross-layer optimization is proposed to capture the time-

heterogeneity of video sources and time-varying characteristics of channel conditions. We also study how to achieve tradeoffs between service objectives in the network level.

Chapter 5 addresses a system supporting interactive video conferencing by jointly optimizing resource allocation in both uplink and downlink in a single-cell WLAN. The proposed framework exploits one more dimension of diversity, namely, the heterogeneity of uplink and downlink channel conditions. We also extend the proposed algorithm to support systems with multiple WLAN cells.

Chapter 6 considers the scenario of transmitting multiple video streams over multiple error-prone transmission paths. We further explore two more dimensions of diversity in path and FEC coded packet. A novel cross-path Packet-Division Multiplexing Access (PDMA)-based error protection scheme is proposed to improve the overall system performance.

The dissertation is concluded in Chapter 7, with discussions on future perspectives.

Chapter 2

System Framework

In this chapter, we first present an overview of multiuser cross-layer system for transmitting multiple video streams over wireless networks. We then briefly review the communication subsystem and video source coding subsystem. For each subsystem, we discuss the available resources, study how to control parameters to achieve the desired goals, and analyze the corresponding constraints in practical implementation. At the end of this chapter, we study the design principles for multiuser cross-layer resource allocation. Based on the design principle, several new frameworks addressing different scenarios and requirements are proposed in the following chapters.

2.1 System Overview

Figure 2.1 depicts a generic framework for multiuser video transmission over wireless network with a total of N communication links. There are four major subsystems, namely, the video source coding subsystem, the communication subsystem, the receiver subsystem, and the resource allocator subsystem. The resource alloca-

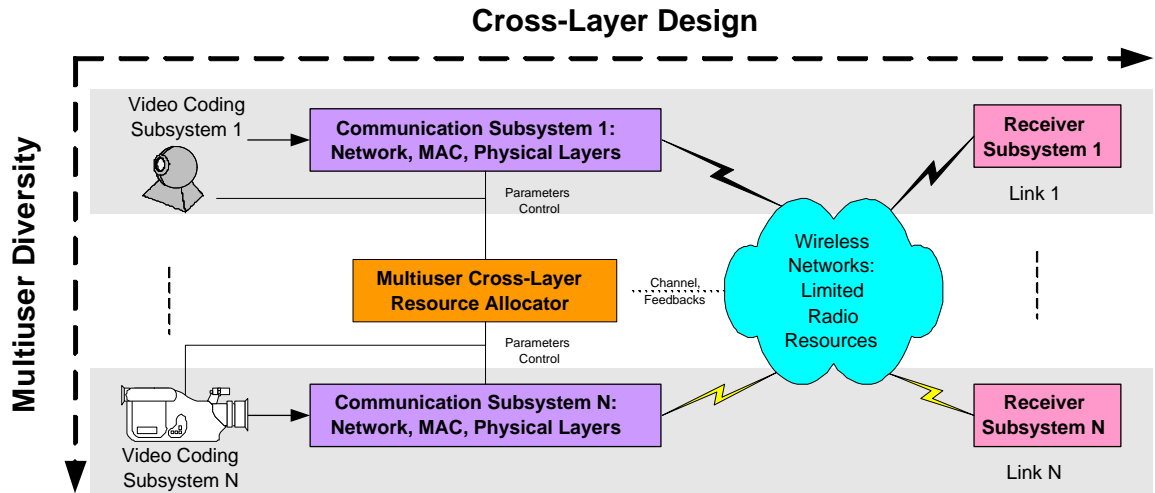


Figure 2.1: Framework of multiuser cross-layer video transmission system over wireless networks

For each subsystem, the necessary information is first collected from the video source coding subsystem, the communication subsystem, and the receiver subsystem. In the video source coding subsystem, each video program is encoded by an encoder in real time. These encoders compress the incoming video frames and send the corresponding rate and distortion (R-D) information to the resource allocator subsystem. The communication subsystem analyzes the available resources in the network layer, medium access control (MAC) layer, and physical (PHY) layer, and supplies the channel information obtained via feedback from the receivers. After gathering the information, the resource allocator subsystem executes optimization algorithms and allocates system resources to different links of different layers so as to achieve the system optimization objectives.

2.2 Wireless Networks

Since the available radio resources are limited, modern wireless networks often adopt the following resource allocation methods to adaptively improve spectrum utilization [25, 102].

1. **Adaptive Modulation and Coding (AMC):** In current and future wireless communications, adaptive modulation is applied to achieve better spectrum utilization. To combat different levels of channel errors, adaptive forward error coding (FEC) is widely used in wireless transceivers. Further, joint consideration of adaptive modulation and adaptive FEC provides each user with the ability to adjust the transmission rate and achieve the desired error protection level, thus facilitating the adaptation to various channel conditions.
2. **Power Control:** The gains of wireless channels generally fluctuate over time. To maintain the link quality, the signal-to-interference-and-noise ratio (SINR) should be dynamically controlled to meet a threshold known as the *minimum protection ratio*. This threshold depends on many factors such as the AMC rate and desired bit error rate (BER). The objective of power control is to guarantee certain link quality and reduce co-channel interferences.
3. **Channel Assignment:** The channel used here is a general concept representing the smallest unit of radio resources that a user can be assigned to transmit data, such as frequency band and time slot. With considering the different channel conditions and users' transmission requirements, dynamic channel assignment can improve the utilization of system resources by exploring the diversities over multiuser, time, and frequency.

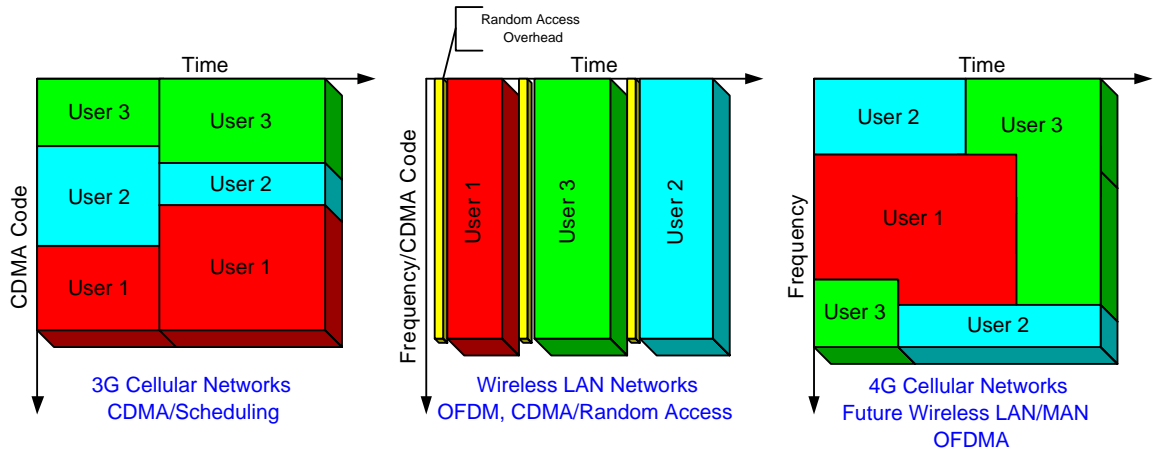


Figure 2.2: Illustration of multiuser resource allocation for 3G, wireless LAN, 4G, future Wireless LAN/MAN

Scheduling and random access are two special types of channel assignment schemes for multiple users to take turns to occupy the limited radio resources over times. Scheduling is a centralized control usually applied in cellular network to determine which user can transmit at a specific time. In contrast, random access can reduce transmission delay in lightly loaded networks such as WLAN and provide autonomous way to avoid conflict of resource usage.

In the following, we discuss the current and future broadband communication networks that can support real-time video transmission.

1. 3G Cellular Networks: CDMA/Scheduling

3G wireless communication systems employ Code Division Multiple Access (CDMA) [70, 75]. CDMA, or more generally, spread spectrum communications, has the following main characteristics: CDMA uses unique spreading codes to spread the baseband data before transmission. The signal occupies a bandwidth that is much broader than narrow-band transmission to

send the information. As the results, the signal is transmitted through the wireless channel with power density below noise level. The receiver then uses a correlator to despread the signal of interest, which is passed through a narrow-band bandpass filter. Unwanted interferences from other users or sources will not pass through the filter. This brings many benefits, such as immunity to narrow-band interference, jamming, and multiuser access.

In Figure 2.2 (a), we show the scheduling scheme for CDMA system to allocate resources to multiple users. Widely adopted in 3G networks, the scheduler allocates a different number of CDMA codes or uses CDMA codes with various spreading factors to users at different time according to the channel conditions, QoS types, bandwidth requirements, and buffer occupancies.

2. WLAN: OFDM, CDMA/Random Access

WLAN can provide higher transmission rate within local areas. There are two major current standards for WLAN, namely, IEEE 802.11b and IEEE 802.11g. IEEE 802.11b uses CDMA technology and supports up to 11Mbps; and IEEE 802.11g uses Orthogonal Frequency Division Multiplexing (OFDM) technology [25] and supports up to 54 Mbps. The basic idea of OFDM is to split a high-rate data stream into a number of lower-rate streams and transmit them over a number of frequency subcarriers simultaneously. Because the OFDM symbol duration of each subcarrier will be longer than that of a wideband signal, the relative dispersion caused by the multipath delay spread of wireless channel is decreased. Inter-symbol interference is eliminated almost completely in every OFDM symbol by introducing guard time which is longer than multipath delay spread. In the guard time, the OFDM symbol is cyclically extended to avoid inter-carrier interference.

Figure 2.2 (b) illustrates the time-frequency relation in the current IEEE 802.11 standard for multiple access. IEEE 802.11 MAC protocol supports two access methods: distributed coordination function (DCF) and point coordination function (PCF). The DCF is the basic access mechanism using carrier sense multiple access with collision avoidance (CSMA/CA), and must be implemented in all stations. In contrast, the PCF is optional and is based on polling controlled by a point coordinator. To reduce the overhead caused by collisions during random access, ready-to-send (RTS) / clear-to-send (CTS) is employed as random access schemes in current wireless LAN standard. At each time, only one user occupies all radio resources. At different time, users try to compete with each other for the next transmission period.

3. 4G Cellular Networks and Future Wireless LAN/MAN: OFDMA

In current OFDM systems, all subcarriers are assigned to a single user at each moment, and multiple users are supported through time division. However, for a given subcarrier, different users experience different channel conditions and the probability for all users to have deep fades in the same subcarrier is very low. Orthogonal Frequency Division Multiplexing Access (OFDMA) allows multiple users to transmit simultaneously on the different subcarriers, while each subcarrier is assigned to the user who is experiencing a good channel condition.

In Figure 2.2 (c), we show the multiuser resource allocation strategy for OFDMA system. We can see that the users' transmission can be allocated to different time-frequency slots. By doing this, the multiuser, time, and frequency diversity can be fully explored to improve the system performance.

2.3 Video Coding

Owing to the perceptual characteristics of human vision, the received video can tolerate a certain level of quality degradation. Trading in lossless reconstructed quality by lossy compression can substantially reduce the required bit rate and still maintain acceptable visual quality. Most current standardized video codecs, such as H.261/3/4 and MPEG-1/2/4, adopt block-based motion compensated prediction and block discrete cosine transform (DCT) coding with quantization to remove temporal and spatial redundancy [84]. Beyond the currently standardized video codecs, researchers have been exploring the 3-D wavelet coding to simultaneously remove the spatiotemporal redundancies [65].

The potential application for video technology has evolved from pre-compressed files in pre-distributed storage, such as DVD, to real-time compressed bitstreams over wireless networks, such as video conferencing. However, there are still many remaining design challenges for real-time video compression and transmission. We summarize them as follows:

1. **Perceptual Quality Control:** Unlike throughput as a major concern in a data transmission system, a video system concerns video quality in terms of either subjective quality assessment or objective distortion measurement such as mean-squared error (MSE) or peak signal-to-noise ratio (PSNR) [123]. We must determine how to control the source coding parameters, such as quantization step size, intra/inter coding mode, and the search range of the motion vectors, to obtain acceptable video quality. In general, a video encoded in variable bit rate (VBR) bitstream gives better perceptual quality than in constant bit rate (CBR) bitstream due to the variation of the scene complexity [50, 98]. To transmit VBR bitstream, the communication module

needs to either dynamically distribute bandwidth for each video frame, or allocate buffers to smoothen traffic over several video frames. The latter approach introduces additional delay.

2. **Rate/Delay Control:** The second challenge concerns how to adjust the source coding parameters to control the data rate such that the bit stream can arrive at the receiver and be decoded in time. Compressed video bitstreams have decoding dependency on the previous coded bitstreams, owing to the spatial and temporal prediction. Therefore, transmitting video streams in real time has a strict delay constraint that belated video data is useless for its corresponding frame and will cause error propagation for the video data that are predictively encoded using that frame as reference.
3. **Error Control:** Owing to the use of variable length coding (VLC) and predictive coding, the compressed bit stream is vulnerable to bit error, as bit error may cause the following VLC codes to be decoded incorrectly. A wireless video system needs to consider the channel introduced error, such as bit error or packet loss [88]. Many error resilient tools, such as synchronization marker, data partitioning, and reversible variable length code, have been proposed and adopted in the latest video standard to increase the robustness of video transmission [109]. Error concealment schemes, which utilize the received video data to conceal the damaged coded video data, are also useful tools to alleviate bit error and packet loss [111]. Those tools can be integrated together to improve the end-to-end video quality.
4. **Scalability:** Unlike the traditional video coding, the next generation of video codec provides a new coding paradigm with scalability, whereby the video

is encoded once and can be transmitted and decoded in many targeted rate according to the channel conditions or users' needs. Thus, a scalable video codec provides flexibility and convenience in reaching the desired visual quality and/or the desired bit rate. Several technologies, such as the MPEG-4 Fine Granularity Scalability (FGS) coding [38, 53, 74], Fine Granular Scalability Temporal (FGST) coding [106], and MPEG-4 Part-10 Scalable Video Coding (SVC) [39], have been proposed to provide high scalability, such as spatial scalability, temporal scalability, and quality scalability.

2.4 Service Objectives

Resource allocation strategies are tied to the system's service objective. There are two different levels of service objective in a multiuser video transmission system, namely, individual level and network level. We discuss these two aspects as follows.

1. Individual Level:

There are two types of visual quality concern in the individual-user level. The most common concern is the average visual quality, often measured in terms of the average mean-squared error of all video frames, or the corresponding PSNR [34, 54, 121]. The other important concern is the quality fluctuation, as substantial quality differences between nearby frames can bring annoying flickering and other artifacts to viewers even when the average PSNR is satisfactory. In many systems that employ a set of frames as an encoding unit (known as a group of pictures/frames (GOP/GOF)), severe quality fluctuation may also appear at the boundaries between groups of frames [116]. The quality fluctuation can be measured by the mean absolute difference of the

MSE between adjacent frames [54, 121, 125].

2. Network Level:

A multiuser video transmission system should consider not only the reconstructed video quality of each individual user but also different perspectives from network-level point of view. We consider two essential network-level service objectives, namely, the fairness and efficiency. The first objective is efficiency, namely, how to achieve the highest overall users' received video qualities with a limited amount of system resources. The second objective regards whether the received video qualities are fair or not for the users who subscribe the same level of video quality. If the users pay the same price for a certain video quality, the received qualities for these users should be similar.

2.5 Design Principles

For video communication systems designed in layers, different layers have their own resources as mentioned in Chapter 2.2 and 2.3. These available resources have practical constraints such as the feasible ranges or finite sets of discrete values that are associated with the resource parameters [93]. For a system with cross-layer design, the allocation of system resources is constrained *vertically across layers*. For example, the bandwidth consumption for use in the application layer should not exceed the achievable capacity by the physical layer. Unlike in a single-user system, network resources are shared by multiple users in a multiuser wireless video system. Allocating these resources to one user would affect the performances of the other users due to the limited amount of resources or interference of simultaneous usage. In other words, the allocation of system resources is further constrained *horizontally among users*. Owing to time-heterogeneity of video source and time-varying

characteristics of channel condition, the allocation of system resources should be performed *dynamically along time*. Moreover, real-time video transmission has additional delay constraints such as playback deadline.

The resource allocation problem often has to deal with resources having both continuous and integer valued parameters. Systems may also have non-linear or/and non-convex constraints and many local optima may exist in the feasible range. Thus, obtaining the optimal solution is often *NP* hard. General approaches to solve the problem are to reduce the search space by some bounds or to adaptively find the solution close to optimum. Some engineering heuristics can be employed for certain network scenarios. To allocate system resources, the resource allocators require some level of up-to-date information for currently available resources in both communication and source subsystem. For systems with powerful nodes, such as base stations in cellular networks, a centralized algorithm can be implemented in the resource allocators located at those powerful nodes to gather all users' information and perform global optimization. On the other hand, for systems with high communication cost to exchange information or without central authority, such as ad-hoc networks, the resource allocators can be located within each individual node, and a distributed solution can be adopted by utilizing only local information. In general, resource allocators with more information can have better performance but require more communication overhead or signaling for accurate and updated information. Thus, there is a design tradeoff between centralized and distributed algorithms.

Based on these design principles, we present in the next chapters several major design methodologies for various video transmission applications over different types of communication networks.

Chapter 3

Low-Delay Bandwidth Allocation

3.1 Introduction

In this chapter, we study a multiuser video transmission system where the transmission channels are error-free. An rate-control algorithm for a single user is proposed to jointly determine source coding rate and channel transmission rate. The proposed algorithm can satisfy low delay requirement and achieve an excellent trade-off in the individual-user level of service objective, namely, the average visual distortion and the quality fluctuation. We then extend to the multiuser case and propose a dynamic resource allocation algorithm with low delay and low computational complexity. By exploring the variations in the scene complexity of video programs as well as dynamically and jointly distributing the available system resources among users in a cross-layer fashion, our proposed algorithm provides low fluctuation of quality for each user, and can support both consistent and differentiated quality among all users from the network-level point of view.

Rate control for single user can be considered as a special case of multiuser resource allocation. As mentioned in Chapter 2.3, a video encoded in VBR bitstream

gives better perceptual quality than in CBR bitstream. Most applications deliver a VBR video bitstream through a channel with a fixed amount of bandwidth (known as a CBR channel) because of its predictable traffic pattern as well as simple network management. However, VBR transmission has been shown to provide better source quality and network utilization [98]. To smoothen the traffic and alleviate the jitter caused by VBR coding and transmission, the system allocates buffers on both the transmitter side and the receiver side. The dynamics of the buffer is subject to two constraints to maintain the QoS. When the buffer overflows, we will start to lose data, which degrades the received visual quality; and when the decoder buffer underflows, the decoder has no data to keep up the decoding, which causes jitters. Therefore, a rate control algorithm must be applied to prevent the buffers from overflowing and underflowing [77].

For systems employing MPEG-1/2/4, or H261/3/4, the encoding rate is often changed by adjusting the quantization step size [19, 77]. To achieve high overall perceptual quality in the single-user scenario, rate control was formulated as an optimization problem in [34, 56, 76, 128]. These approaches are suitable for off-line applications where the entire video content is known to the transmitter. The computation cost for handling a long video sequence is high due to the nature of integer and dynamic programming. To facilitate solving the rate control problems, several R-D models of existing video codecs have been exploited in the literature. An R-D based approach was proposed in [78] under the assumption that the DCT coefficients of a motion-compensated residue frame are uncorrelated and Laplacian distributed. A rate-distortion model using intra-frame approximation and inter-frame dependency within one GOP was proposed in [54] to meet the perceptual requirement. A quadratic R-D model and rate control for MPEG-4 was studied

in [51, 62]. A rate control algorithm employing a linear correlation model was proposed in [33], whereby the correlation between the rate and the percentage of zeros among the quantized transform coefficients was explored. To simplify the selection of encoding and channel rates, wavelet-based embedded codecs were considered in the rate control problems of [121] and [72].

Most prior work targeted at optimizing either average visual distortion or the quality fluctuation in the individual level of service objectives. If the rate-distortion characteristics of all video frames are identical, the bit rate allocated to each frame will be equal, leading to identical perceptual quality between frames and the above two measures can be simultaneously optimized [124]. In reality, however, a video has varying R-D characteristics, making it difficult to optimize the average quality and the quality fluctuation at the same time. In this chapter, we aim at reaching an excellent trade-off between these two quality criteria through a real-time low-delay algorithm [96, 97].

Sliding window is a general approach that can be used to keep track and allocate system resources. The work in [18] took advantage of the fine granularity of the MPEG-4 FGS codec and proposed a variable-size sliding window scheme to control how much FGS layer data is sent under different channel conditions. An R-D based rate control scheme for pre-stored video was studied in [125] using a three-level bit allocation for the base layer and employing a sliding window for the FGS layer rate control. An online algorithm using a look-ahead sliding window to achieve constant perceptual quality was proposed in [124]. To apply this scheme for transmitting real-time encoded video, we need to allocate extra storage to store several frames ahead, and perform bit allocation for the current frame by solving such an optimization problem that all frames within a look-ahead window

have consistent and the highest possible perceptual quality subject to a given rate budget. Our studies show that to obtain a low fluctuation of quality, the window size should be no smaller than the size of half to one GOP, which leads to a nontrivial amount of delay that is often too long for real-time interactive applications. We will investigate in this chapter how to overcome the problems of long delay and extra storage associated with the sliding window approach.

Several works on joint rate control for multiple video programs employed MPEG-1/2 codecs [9, 108, 114]. And the extension of the sliding window approach to multiple MPEG-4 FGS video programs was proposed in [124], employing a 2-D window to address the multiuser problem. However, the computational complexity and extra storage for the look-ahead frames of the sliding window approach go up with the increase of the window size and the number of users. As the number of users increases in the system, the required computational resources to achieve a low fluctuation of quality become formidable. In this chapter, we also study how to overcome the problem of high computational complexity of the 2-D sliding window approach and improve the system scalability to accommodate many users.

3.2 FGS Rate-Distortion Model and Similarity

Existing rate control schemes for a single-layer video stream often employ an intra-frame rate-distortion model. Laplacian and Gaussian distribution are typical approximations of DCT coefficients, leading to the frequent use of an exponential or a polynomial rate-distortion model [78, 124]. In contrast to single-layer codecs, FGS codec is a two-layer embedded scheme with an enhancement layer encoded bit plane by bit plane. There is a need to model the statistical distribution of DCT bit planes and their rate-distortion characteristics. Furthermore, due to the nature

of the temporal redundancy in video, the predicatively encoded frames within one scene have highly similar rate-distortion characteristics. In this section, we present rate-distortion models for intra-frame and inter-frame of a FGS layer, which will be used in our work.

3.2.1 Intra-frame Rate-Distortion Model

MPEG-4 FGS standard employs bit-plane coding of the DCT residue between the original frame and the base layer. For a given bit plane in a frame, if the video is spatially stationary so that the length of the entropy encoded FGS symbols in all blocks is similar to each other, the decoded bit rate and the corresponding amount of reduced distortion will have an approximately linear relationship over the bit rate range of this bit plane. Previous studies in [124, 125] and our experiments show that a piecewise linear line is a good approximation to the operational rate-distortion curve of FGS video in the frame level. This piecewise linear line model can be described as

$$D_j(r_j) = \left(\frac{MSE_j^{k+1} - MSE_j^k}{Rate_j^{k+1} - Rate_j^k} \right) (r_j - Rate_j^k) + MSE_j^k, \quad (3.1)$$

$$\text{for } Rate_j^k \leq r_j \leq Rate_j^{k+1}, \text{ and } k = 0, \dots, p - 1.$$

Here, $D_j(r_j)$ represents the MSE between the j^{th} original frame and the decoded frame with rate r_j , MSE_j^k the distortion of the j^{th} frame measured in mean square error after completely decoding the first k DCT bit planes, $Rate_j^k$ the corresponding bit rate, and p the total number of bit planes. We use MSE_j^0 and $Rate_j^0$ to represent the distortion and rate of the base layer, respectively. Since DCT is a unitary transform, measuring the mean square error between an original frame and its partially decoded version from FGS encoded stream is equivalent to calculating

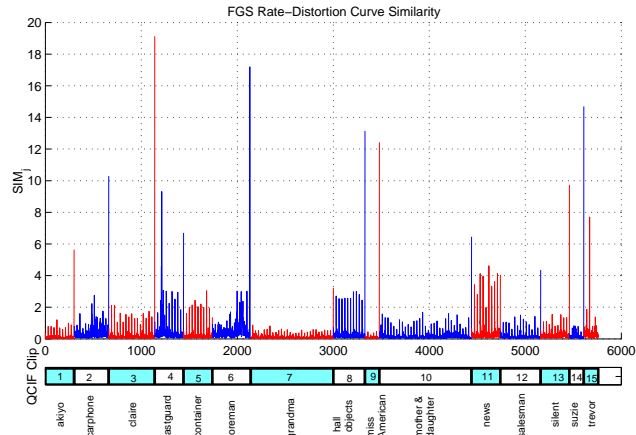


Figure 3.1: Inter-frame similarity of FGS rate-distortion characteristics. The results for the odd and even scenes are presented in alternating colors.

the average energy of the un-decoded DCT bit planes in the FGS data stream, along with the residue between the original frame and the complete FGS data. Thus, all MSE_j^k 's and $Rate_j^k$'s can be obtained during the encoding process.

3.2.2 Similarity in Inter-frame R-D Characteristics

Another important characteristic of FGS video is that the rate-distortion curves of FGS layer between two consecutive frames are similar when they are within the same scene. The rationale is as follows: for a video segment within a scene, the energy of the motion compensation residues between two adjacent frames are comparable. As the base layer is generated using a set of large quantization steps, it leaves most motion residues to be coded by the FGS layer. Therefore after FGS encoding, the overall R-D characteristics between two adjacent frames are similar.

We quantify the similarity of the R-D characteristics between frame j and $j + 1$

using

$$SIM_j = \frac{1}{T} \sum_{t=1}^T |D_j(t \cdot c) - D_{j+1}(t \cdot c)|, \quad (3.2)$$

where c is a bit rate sampling interval, $T = \lfloor \frac{\min\{rf_j^{max}, rf_{j+1}^{max}\}}{c} \rfloor$, and rf_j^{max} is the maximal available amount of FGS data for the j^{th} frame. A low value of SIM_j implies high similarity in the R-D characteristics of the j^{th} and $(j + 1)^{th}$ frames. Figure 3.1 shows the SIM_j for a long video sequence consisting of 15 different standard QCIF clips. As we can see, the R-D models within each clip show a strong similarity. The similarity measure becomes large when transiting from one clip to another.

3.3 Low-Delay Bandwidth Resource Allocation for Single User

To facilitate the investigation of the resource allocation problem in a multiuser system, we first study in this section a special case that concerns only a single user in the system. We begin with a discussion on the mechanism of a single-user FGS streaming video system and the corresponding constraints. We formulate this system as a resource allocation problem with two perceptual objectives subject to the system constraints. An online bandwidth resource allocation algorithm with low delay and low fluctuation of quality is then proposed to achieve a trade-off point between these two perceptual criteria.

3.3.1 System Constraints

Illustrated in Figure 3.2 is a typical streaming video system. There are two sub-components in the encoder. One is the base layer encoder and the other is the

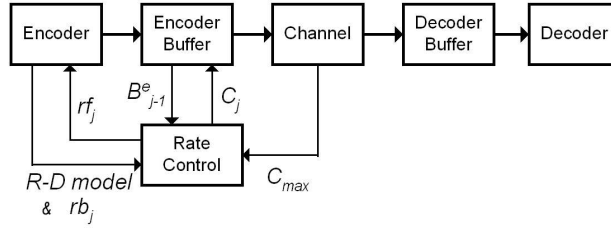


Figure 3.2: Block diagram of a single-user video streaming system

FGS layer encoder. We discretize the time line by dividing one second into F time slots, where F is the video frame rate. For the simplicity in system design and providing a primitive quality with low fluctuation, we set a large fixed quantization step for all frames in the base layer codec and only perform the rate control for the FGS layer. We denote the base layer rate as rb_j , i.e. a total of rb_j bits must be sent at the j^{th} time slot to ensure the baseline quality. The FGS encoder encodes the bit planes of the residue. Both encoders analyze the R-D characteristics of the incoming video frame and pass the necessary information, such as the R-D pairs $(Rate_j^k, MSE_j^k)$, to the rate control module. After the rate control module determines the amount of FGS data to be transmitted, the encoded base layer and the truncated FGS layer bitstream are moved to the encoder buffer, where we denote the FGS data rate at the j^{th} time slot as rf_j . The channel then delivers video bitstream from the encoder buffer to the decoder buffer. Here we assume that the channel has a maximum rate for reliable transmission, C_{max} , although it is not necessarily in its full load all the time. The amount of channel transmission rate at the j^{th} time slot, denoted as C_j , is also determined by the rate control module. For simplicity, we assume that the transmission delay of every packet is fixed at d^c time slots [34, 56]: if a packet is sent from the encoder buffer at the j^{th} time slot, it will arrive at the decoder buffer at the $(j + d^c)^{th}$ time slot. The decoder

Table 3.1: Summary of notations for low-delay bandwidth allocation

B_{max}^e	maximal size of encoder buffer
B_j^e	encoder buffer occupancy at the j^{th} time slot
B_{max}^d	maximal size of decoder buffer
B_j^d	decoder buffer occupancy at the j^{th} time slot
C_{max}	maximal channel capacity
C_j	channel transmission rate at the j^{th} time slot
rb_j	base layer rate of the j^{th} frame
rf_j^{max}	maximal available FGS layer rate of the j^{th} frame
rf_j	transmitted FGS layer rate of the j^{th} frame
r_j	total effective encoding rate of the j^{th} frame, $r_j = rb_j + rf_j$
d^c	transmission delay (in unit of frames)
d^d	number of pre-stored frames in the decoder buffer at the beginning of the service
$D_j(\cdot)$	rate-distortion function of the j^{th} frame
R_j	upper bound of FGS rate budget at the j^{th} time slot
R_j^f	fractional FGS rate budget at the j^{th} time slot
R_j^p	required rate of j^{th} frame to keep the same distortion as the previous frame
β	FGS budget factor for determining R_j^f
w_p	weighting factor for combining R_j^f and R_j^p

fetches data from the decoder buffer, decodes it, and displays each decompressed video frame at its desired instant. Therefore, the major task of the rate control module is to determine rf_j and C_j . To ease the discussion, we summarize the notations in Table 3.1.

There are three constraints imposed in this system, as studied in the literature [77, 98]. The first constraint is to prevent the encoder buffer of a limited size from overflow. At the j^{th} time slot, a data segment of size C_j is taken from an encoder buffer and sent through the channel, and then a newly encoded frame with

size $r_j = rb_j + rf_j$ is added to the encoder. The dynamics of the encoder buffer can thus be expressed as

$$B_j^e = \max\{B_{j-1}^e + rb_j + rf_j - C_j, 0\} \leq B_{max}^e, \quad (3.3)$$

where B_j^e is non-negative and describes the occupancy of encoder buffer, and B_{max}^e is the maximal size of encoder buffer. In addition, the FGS rate rf_j should be non-negative. For a given C_j , we can rearrange inequality (3.3) as a constraint for FGS rate rf_j :

$$0 \leq rf_j \leq B_{max}^e + C_j - B_{j-1}^e - rb_j. \quad (3.4)$$

The second constraint is on the channel transmission rate, C_j . It is non-negative and cannot exceed the maximal channel capacity, C_{max} . That is,

$$0 \leq C_j \leq C_{max}. \quad (3.5)$$

The third constraint is on the occupancy of the decoder buffer, which should neither overflow nor underflow. We assume that the decoder fetches all the data that belongs to the next frame from the decoder buffer and decodes it within one time slot. In addition, we assume playback buffering of d^d frames, i.e., the first d^d frames are received and stored in the decoder buffer before the playback is started. The total end-to-end delay from the encoder buffer through the channel and decoder buffer to the decoder is thus $d^c + d^d$ frames delay. The decision on how much data is sent into the channel at the j^{th} time slot will directly affect the decoder buffer occupancy at the $(j + d^c)^{th}$ time slot. To meet the constraint imposed on the decoder buffer occupancy at the $(j + d^c)^{th}$ time slot, there is a corresponding limit on how much data can be sent through the channel at the j^{th} time slot. Denote the decoder buffer occupancy as B_j^d , and the maximal size of

decoder buffer as B_{max}^d . We summarize the constraint on the decoder buffer at the $(j + d^c)^{th}$ time slot as

$$B_{j+d^c}^d = B_{j+d^c-1}^d + C_j - r_{j-d^d} \in [0, B_{max}^d]. \quad (3.6)$$

Combining (3.5) and (3.6), we arrive at the following constraint for the channel transmission rate C_j

$$\begin{aligned} \max\{r_{j-d^d} - B_{j+d^c-1}^d, 0\} &\leq C_j \leq \\ \min\{B_{max}^d - B_{j+d^c-1}^d + r_{j-d^d}, C_{max}\}. \end{aligned} \quad (3.7)$$

In summary, inequalities (3.4) and (3.7) are the fundamental constraints for a single-user FGS video streaming system.

3.3.2 Criteria for Visual Quality

As mentioned in Chapter 2.4, there are two different levels of service objective. In this single-user system, we consider the individual-user level of service objectives. We adopt two visual quality criteria for video sequences to measure the average distortion and the quality fluctuation. More specifically, the average received quality is measured by the average mean square error (*aveMSE*) of all M frames in a video sequence:

$$aveMSE = \frac{1}{M} \sum_{j=1}^M D_j(r_j), \quad (3.8)$$

where $D_j(r_j)$ represents the MSE between the j^{th} original frame and the decoded frame with rate r_j . To account for the fluctuation of quality between consecutive frames, a large value of which can be objectionable to viewers, we use the mean absolute difference of consecutive frames' mean square error (*madMSE*) to measure

the perceptual fluctuation:

$$madMSE = \frac{1}{M-1} \sum_{j=2}^M |D_j(r_j) - D_{j-1}(r_{j-1})|. \quad (3.9)$$

The higher the $madMSE$ is, the larger the perceptual fluctuation is. We also define the corresponding peak signal-to-noise ratio (PSNR) version of these two criteria and denote as $avePSNR$ and $madPSNR$, respectively.

3.3.3 Problem Formulation

Our objective is to design a rate control strategy to achieve both low $aveMSE$ (high $avePSNR$) and low $madMSE$ (low $madPSNR$) subject to the constraints of (3.4) and (3.7). For offline applications where the entire video content is readily available before the transmission, all rate-distortion information is known and we can formulate this system as the following optimization problem:

$$\min_{\{rf_j, C_j\}} f(aveMSE, madMSE)$$

subject to

$$\begin{cases} 0 \leq rf_j \leq B_{max}^e - B_{j-1}^e - rb_j + C_j, \forall j, \\ \max\{r_{j-d^d} - B_{j+d^c-1}^d, 0\} \leq C_j \leq \min\{B_{max}^d - B_{j+d^c-1}^d + r_{j-d^d}, C_{max}\}, \forall j, \\ B_j^e = \max\{B_{j-1}^e + rb_j + rf_j - C_j, 0\} \leq B_{max}^e, \forall j, \\ 0 \leq B_{j+d^c}^d = B_{j+d^c-1}^d + C_j - r_{j-d^d} \leq B_{max}^d, \forall j. \end{cases}$$

In this formulation, $f(\cdot, \cdot)$ is a function reflecting the importance and relevance of the average distortion and the quality fluctuation in the human perceptual system. For example, a linear combination function of $aveMSE$ and $madMSE$ is a simple choice of $f(\cdot, \cdot)$.

An optimal solution can be found for the above offline problem using standard nonlinear programming with penalty functions. The complexity for searching for

optimal solution would, however, be formidable except for short video clips. In addition, the offline solution is not applicable to online applications where the video content is not entirely available beforehand. If the variations of the R-D characteristics of video sources can be well captured by a finite-state Markovian chain, we can model this system using stochastic three-machine flowshop with finite buffers [80] and obtain an optimal rate control policy using dynamic programming techniques. However, it has been shown that a compressed video sequence trace has long-range dependence [8], which is different from the short-range dependence such as a Markovian process and cannot be handled well using existing solutions. Thus in this chapter, we focus on a sequential resource allocation solution that has a moderate amount of computational complexity and can accommodate online video applications.

The strategy of choosing the effective encoding rate for the FGS layer, $\{rf_j\}$, and the channel transmission rate, $\{C_j\}$, closely depends on the relative weights of the average distortion and the perceptual fluctuation in the objective function. To achieve low *aveMSE* alone, one may employ a greedy strategy to make the encoder buffer as full as possible all the time and make full use of the available channel bandwidth. This may lead to the desire to select C_j at the upper bound in (3.7), namely,

$$C_j = \min\{B_{max}^d - B_{j+d^e-1}^d + r_{j-d^d}, C_{max}\}, \quad (3.10)$$

and to set the FGS rate at the upper bound in (3.4), which is denoted as R_j and defined as:

$$R_j \triangleq B_{max}^e + C_j - B_{j-1}^e - rb_j. \quad (3.11)$$

When the encoder buffer is always full, the amount of incoming data cannot exceed the maximal amount of data allowed to be sent through the channel at each time

slot. This is equivalent to assigning the same bandwidth resource for transmitting each frame. When encountering intra-coded frames (or I frames), which have a larger amount of data at the base layer than predictively coded frames, we will have very limited budget left for sending their associated FGS enhancement layers. The MSE of I frames will thus be larger than the MSE of the other types of frames. This leads to a potential increase in *madMSE*.

On the other hand, low *madMSE* may be achieved by assigning each frame a rate that corresponds to the same distortion, MSE_s . To do so, we extract the rate-distortion pairs from the FGS encoder, approximate the rate-distortion curve for each frame, and assign the rate for the FGS enhancement layer as $rf_j = D_j^{-1}(MSE_s)$. To prevent encoder buffer from overflowing when encountering I frames or a new complex scene, we would have to allocate a small amount of data rate for the FGS layers of these I frames. To keep the lowest *madMSE*, other frames will also have a small amount of FGS-layer data. As a result, this second approach would not give a low *aveMSE*. Next, we present a new resource allocation algorithm that can achieve an improved trade-off between the average distortion and the quality fluctuation.

3.3.4 The Proposed Resource Allocation Algorithm

We introduce two weight factors in our proposed resource allocation algorithm to solve the above-mentioned problems. To overcome the quality fluctuation problem in the lowest-*aveMSE* scheme, we propose to use a fraction of the maximally allowed FGS data rate (determined by the buffer constraints) as the effective FGS encoding rate, i.e. $R_j^f \triangleq \beta R_j$, where $\beta \in [0, 1]$ is a budget factor. Compared to adopting the full budget R_j , the fractional budget can keep the encoder buffer

occupancy low to accommodate future I frames and other complex frames. As such, the rate budget available to the incoming I-frames will be close to the maximal encoder buffer size plus the full channel bandwidth, allowing for more FGS data of the I frames to be sent to avoid a high increase in the *madMSE*.

To overcome the problem of low overall perceptual quality as in the lowest-*madMSE* scheme, we relax the requirement of zero *madMSE* fluctuation by taking partial consideration of both the rate that maintains zero *madMSE* and the current occupancy of the encoder buffer. We quantify this strategy using a weighting factor $w_p \in [0, 1]$ and allocate the FGS rate for the j^{th} frame as

$$\begin{aligned} rf_j &= \min\{w_p R_j^p + (1 - w_p) R_j^f, R_j\} \\ &= \min\{w_p R_j^p + (1 - w_p) \beta R_j, R_j\}, \end{aligned} \quad (3.12)$$

where R_j^p is the amount of FGS data needed to achieve the same perceptual quality as the previous frame and can be determined by

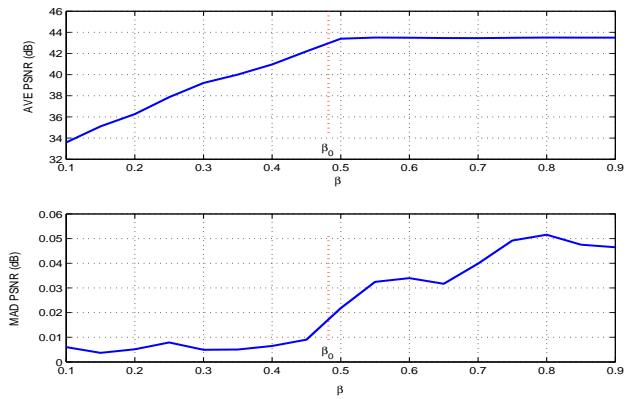
$$R_j^p \triangleq D_j^{-1}(D_{j-1}(r_{j-1})) - rb_j. \quad (3.13)$$

As we can see, the allocated FGS rate rf_j is determined using the two factors β and w_p . The lowest *aveMSE* scheme and *madMSE* scheme are two special cases of this new strategy: when $w_p = 0$ and $\beta = 1$, (3.12) becomes the lowest-*aveMSE* scheme; and when $w_p = 1$, (3.12) becomes the lowest-*madMSE* scheme.

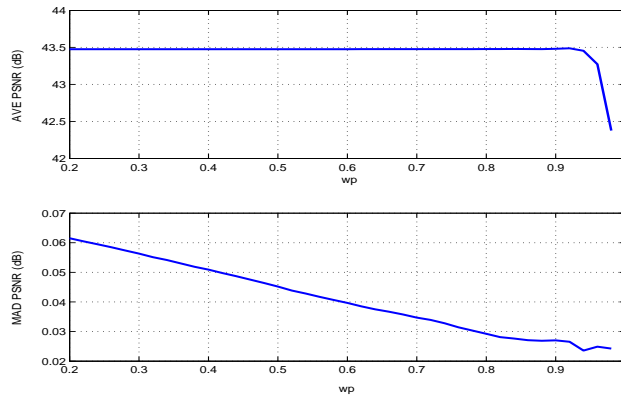
We now examine how to select appropriate β and w_p to achieve a good trade-off between low *aveMSE* and low *madMSE*.

Selection of β

We first fix w_p and study the impact of β on *avePSNR* and *madPSNR* when consecutive video frames have similar R-D characteristics. In this situation and



(a) avePSNR and madPSNR for different β



(b) avePSNR and madPSNR for different w_p

Figure 3.3: Impact of β and w_p on visual quality for the first 200 frames of the *grandmother* sequence

with a fixed w_p , when β becomes larger, both the *madPSNR* and the *avePSNR* will increase. However, after β passes a specific value, β_0 , the improvement of *avePSNR* is dramatically reduced while the quality fluctuation becomes more significant. This phenomenon is demonstrated in Figure 3.3(a), where we use the first 200 frames from the QCIF video clip of the *grandmother* as an example and set the w_p factor at 0.95. Given such trends of *avePSNR* and *madPSNR* for different β , the β_0 value, indicated by a vertical line in Figure 3.3(a), provides a good trade-off between *avePSNR* and *madPSNR*. As shown in Appendix A.1, β_0 can be expressed as

$$\beta_0 = \frac{C_{max} - \overline{rb}}{B_{max}^e + C_{max} - \overline{rb}}, \quad (3.14)$$

where \overline{rb} represents the average rate of the base layer, which can be approximated using a moving average of the bit rate statistics of the past L frames. We can see that β_0 is an equilibratory operating point to keep the encoder buffer near empty and the channel utilization near full.

We should notice that in reality, the consecutive video frames do not have exactly the same R-D characteristics. So if β is set to be exactly β_0 , the system is on the verge between stable and unstable operation: the encoder buffer is nearly empty, and as the video content fluctuates, the buffer may underflow. Thus to ensure a high utilization of channel bandwidth and high *avePSNR*, we should select a β that is slightly above β_0 such that $\beta = \beta_0 + \Delta\beta$, where $\Delta\beta$ is a small positive constant.

Selection of w_p

In general, a system with a high value of w_p has low fluctuation of visual quality. When consecutive frames within a video segment have similar R-D characteristics,

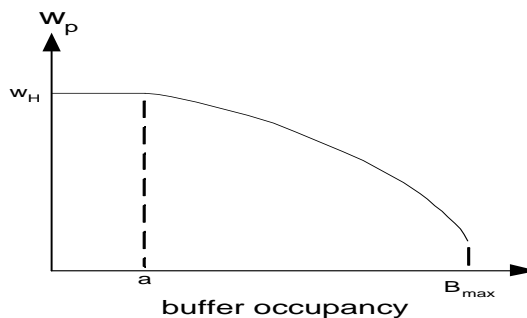


Figure 3.4: Selection of w_p according to the encoder buffer occupancy.

increasing w_p affects only the *madPSNR* while the *avePSNR* has little decrease until w_p is close to one. To achieve low fluctuation of quality, high value of w_p is preferred. This trend is illustrated in Figure 3.3(b), where we again use the above-mentioned *grandmother* video clip as an example and set β to a fixed value of 0.5.

When two adjacent frames exhibit significant difference in R-D characteristics such as when arriving at scene boundary, we need to make adjustment to this system to handle the following frames. We consider two cases here. The first case is that the video sequence enters a new segment with more complex R-D characteristics than that of the previous segment, whereby the FGS rate required to maintain the same PSNR level as before is higher than the FGS rate for the previous sequence. To balance between the need of preventing the encoder buffer from overflowing and controlling the fluctuation of perceptual quality, we dynamically adjust the weighting factor w_p with respect to the encoder buffer occupancy. When the encoder buffer occupancy is lower than a threshold a , we set w_p at a high value w_H to keep the distortion similar to that of the previous frame. When the encoder buffer occupancy is higher than threshold a , we try to drain out the data from the

buffer quickly by choosing w_p as a concave and decreasing function of the buffer occupancy as shown in Figure 3.4, so that the higher buffer occupancy is, the lower w_p is. As an example, the overall selection of w_p can be chosen as:

$$\begin{aligned} w_p &= W(B_{j-1}^e) \\ &= w_L + (w_H - w_L) \left(1 - u(B_{j-1}^e - a) \frac{B_{j-1}^e - a}{B_{max}^e} \right)^b, \end{aligned} \quad (3.15)$$

where $u(\cdot)$ is the step function, and a and b are positive constants.

The second case is that the video sequence enters a new segment with simpler R-D characteristics than that of the past segment, whereby the FGS rate required to maintain the same PSNR level as before is lower than the FGS rate for the previous sequence. To balance between fully utilizing the available channel bandwidth resource and maintaining low fluctuation of quality, after detecting a change in R-D characteristics, we immediately adjust w_p to a low value w_L to utilize more available bandwidth and maintain this value for the following P_T frames. As the scene transition is complete and the channel bandwidth becomes highly utilized again, we can adjust w_p back to a high value to maintain constant quality.

In summary, we adjust w_p dynamically according to the encoder buffer occupancy and the detection of significant change in rate-distortion characteristics. The changes in R-D characteristics can be identified by calculating the relative rate change between R_j^p and R_{j-1}^p , i.e. we check whether $|R_j^p - R_{j-1}^p|/R_{j-1}^p$ is greater than a threshold S_T . The β parameter will be chosen to be right above β_0 as in (3.14) and the channel transmission rate according to (3.10). We present the detailed algorithm in Table 3.2.

Table 3.2: The proposed single-user resource allocation algorithm (S-LDLF)

<p>1. Initialization:</p> $j=RDSC=1, B_0^e = B_1^d = \dots = B_{d^c}^d = 0, R_0^p = rb_1,$ $w_p = w_L, D_0(rb_0) = D_1(rb_1).$ <p>2. While the last frame of this video is not reached:</p> <p>a) Calculate the budget factor and weighting factor</p> $\bar{rb} = \frac{1}{j-\max(j-L,0)} \sum_{k=\max(j-L,0)+1}^j rb_k,$ $\beta = \frac{C_{max} - \bar{rb}}{B_{max}^e + C_{max} - \bar{rb}} + \Delta\beta, R_j^p = D_j^{-1}(D_{j-1}(r_{j-1})) - rb_j,$ <p>If $R_j^p - R_{j-1}^p /R_{j-1}^p \geq S_T$, then $RDSC=j$.</p> <p>If $j \in [RDSC, RDSC + P_T]$, then $w_p = w_L$,</p> <p>else $w_p = W(B_{j-1}^e)$.</p> <p>b) Select the channel transmission rate</p> $C_j = \min\{B_{max}^d - B_{j+d^c-1}^d + r_{j-d^d}, C_{max}\}.$ <p>c) Select FGS rate</p> $R_j = B_{max}^e + C_j - B_{j-1}^e - rb_j,$ $rf_j = \min\{w_p R_j^p + (1 - w_p)\beta R_j, R_j\}.$ <p>d) Update the encoder buffer occupancy information:</p> $B_j^e = \max\{B_{j-1}^e + rb_j + rf_j - C_j, 0\}.$ <p>If $B_j^e = 0$, then $C_j = B_{j-1}^e + rb_j + rf_j$.</p> <p>e) Update the decoder buffer occupancy information:</p> $B_{j+d^c}^d = B_{j+d^c-1}^d + C_j - r_{j-d^d}.$ <p>f) $j = j + 1$.</p>

3.4 Low-Delay Bandwidth Resource Allocation for Multiple Users

In this section, we extend the proposed bandwidth resource allocation algorithm from handling single user to multiple users. A simple way to deal with multiple users/sequences is to allocate a fixed amount of resource, including various buffers and channel bandwidth, to each user, and apply our proposed single-user approach to each individual user. We shall call this strategy *multiple single-user approach*. A more sophisticated approach allows for dynamically allocating resource among users and has the potential to improve the utilization of critical resources. Multiple users share the total channel bandwidth and buffer capacity, and a central resource allocation system dynamically distribute these system resources to handle the transmission of the video sequences from all users. We shall call this class of strategies *dynamic multiuser approaches*. We will focus on the dynamic multiuser approach and aim at achieving high average visual quality and low fluctuation of quality for each user. We will examine the scenarios of uniform quality of service among all users versus differentiated service from the network-level point of view. The performance of the dynamic multiuser strategy will be compared with the multiple single-user strategy through simulations in Section 3.5.

3.4.1 System Constraints

An N -user system is depicted in Figure 3.5. At the server side, each user has his/her own video encoder to encode a different video program in real time. For the i^{th} user, the corresponding encoder sends the measured parameters of the rate-distortion model of the current j^{th} frame to the resource allocation module. The

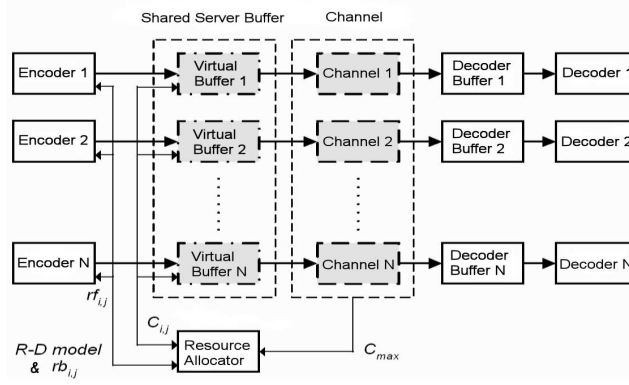


Figure 3.5: Block diagram of a multiuser video streaming system

parameters are in the form of $(Rate_{i,j}^k, MSE_{i,j}^k)$ for the first k^{th} bitplane. Using the R-D model, the resource allocation module determines the amount of FGS data to be transmitted. The encoder of each user then moves both the base layer data at the rate of $rb_{i,j}$ and the FGS layer bitstream truncated at the allocated rate $rf_{i,j}$ to the shared server buffer whose maximal capacity is B_{max}^e . Denote the occupancy of the shared server buffer at the time slot j as B_j^e and the amount of data left by the i^{th} user in the server buffer as $B_{i,j}^e$. We can treat $B_{i,j}^e$ as a virtual encoder buffer for the i^{th} user, and the sum of all virtual encoder buffers' occupancy equals to B_j^e . All users also share a channel whose maximal outbound capacity is C_{max} . The resource allocation module needs to determine the channel transmission rate allocated for each user's data at the time slot j , which we denote as $C_{i,j}$ for the i^{th} user. Upon receiving the data packets of the video program intended for him/her, each end-user first stores them temporarily in the decoder buffer, then decodes and renders each frame on time. In summary, similar to the single user case, the duty of the multiuser resource allocation module is to determine $rf_{i,j}$ and $C_{i,j}$ jointly for all users.

In parallel to the single-user case, there are three sets of system constraints for

multiuser resource allocation. The first set of constraints is on the server buffer, which should not overflow. In particular, the sum of all virtual encoder buffers should not exceed the capacity of the server buffer. The dynamics of the buffer occupancy can be extended from the single-user case. The constraints can be described as

$$\begin{aligned} B_{i,j}^e &= \max\{B_{i,j-1}^e + rb_{i,j} + rf_{i,j} - C_{i,j}, 0\} \leq B_{max}^e, \forall i \\ B_j^e &= \sum_{i=1}^N B_{i,j}^e \leq B_{max}^e. \end{aligned} \quad (3.16)$$

The constraints of the FGS layer rates for each user can be extended from the single-user problem and described as

$$0 \leq rf_{i,j} \leq B_{max}^e - B_{i,j-1}^e - rb_{i,j} + C_{i,j}, \quad \forall i, \quad (3.17)$$

and

$$0 \leq \sum_{i=1}^N rf_{i,j} \leq B_{max}^e - B_{j-1}^e - \sum_{i=1}^N rb_{i,j} + \sum_{i=1}^N C_{i,j}. \quad (3.18)$$

Since all users share the overall bandwidth, both the individual and the aggregated channel transmission rate should be non-negative and not exceed the maximal capacity. These channel transmission rate constraints can be described as

$$0 \leq C_{i,j} \leq C_{max}, \quad \forall i, \quad (3.19)$$

$$0 \leq \sum_{i=1}^N C_{i,j} \leq C_{max}. \quad (3.20)$$

The constraints of the decoder buffer is the same as the single user case:

$$B_{i,j+d_i^c}^d = B_{i,j+d_i^c-1}^d + C_{i,j} - r_{i,j-d_i^d} \in [0, B_{i,max}^d], \quad (3.21)$$

where $B_{i,max}^d$ is the maximal size of the i^{th} decoder buffer, d_i^c the channel transmission delay for user i , and d_i^d the pre-stored frame delay in the decoder buffer

for user i . Rearranging and combining (3.19) and (3.21), we obtain a simplified constraint for the individual channel transmission rate:

$$C_{i,j}^L \leq C_{i,j} \leq C_{i,j}^U, \quad \forall i, \quad (3.22)$$

where

$$\begin{aligned} C_{i,j}^U &\triangleq \min\{B_{i,max}^d - B_{i,j+d_i^c-1}^d + r_{i,j-d_i^d}, C_{max}\}, \\ C_{i,j}^L &\triangleq \max\{r_{i,j-d_i^d} - B_{i,j+d_i^c-1}^d, 0\}. \end{aligned}$$

Inequalities (3.17), (3.18), (3.20), and (3.22) are fundamental constraints in a multiuser system. Under these constraints, we determine the rate of the FGS data and the channel transmission rate for each user in the system to achieve low fluctuation of perceptual quality of each program as well as the desired uniform or differentiated perceptual quality among all programs.

3.4.2 The Proposed Resource Allocation Algorithm

Our proposed multiuser resource allocation algorithm first allocates the channel transmission rate for each user subject to (3.20) and (3.22). With a selected channel transmission rate, we extend rate control strategy that we have proposed for the single-user case to the multiuser case to determine the feasible range for FGS layer data of each user according to (3.17) and (3.18). Specifically, we use two weight factors β and w_p to achieve a trade-off between average perceptual quality and quality fluctuations.

Selection of Channel Transmission Rate

As all users share the overall channel bandwidth in multiuser system, we need to dynamically adjust the transmission rate allocated for each user. Our strategy con-

sists of two steps: first, we assign each user a lower bound of channel transmission rate, $C_{i,j}^L$, to prevent all decoder buffers from underflowing. Second, to help drain out the virtual encoder buffers, we distribute the rest of the available bandwidth, $C_{max} - \sum_{i=1}^N C_{i,j}^L$, to each user proportional to his/her previous encoding rate $r_{i,j-1}$. Thus, when a video program encounters an I-frame and leaves a large amount of data in its virtual encoder buffer at the previous time slot, our strategy will assign the corresponding user a high channel transmission rate to drain his/her virtual encoder buffer at the current time slot.

Selection of FGS Rate

As in the single-user strategy proposed in Chapter 3.3.4, to balance between low fluctuation of quality and high average quality, we introduce two weight factors to our multiuser algorithm, namely, β and w_p .

We first take an aggregated view on how much total bit rate are spent in the base layer for all users ($\sum_i r b_{i,j}$), and on what the upper bound on total FGS rate is at the j^{th} time slot (R_j) according to (3.18). This is as if the aggregated rates are applied to a single “super-user”. The β factor is applied to R_j to obtain a fractional FGS rate budget R_j^f that helps overcome the quality fluctuation.

Next, we distribute R_j^f to each user. For applications that desire uniform quality among users, the fractional rate budget for each user, $\{R_{i,j}^f\}$, is determined through the following optimization formulation:

$$\min_{R_{1,j}^f, \dots, R_{N,j}^f} D_j \tag{3.23}$$

subject to

$$\begin{cases} D_j \triangleq D_{1,j}(R_{1,j}^f) = \dots = D_{N,j}(R_{N,j}^f), \\ \sum_{i=1}^N R_{i,j}^f \leq \beta R_j. \end{cases}$$

Since the rate-distortion functions are monotonically decreasing, this optimization problem with equality constraints can be easily solved using bi-section search. The search algorithm calculates the total required rates to achieve a target distortion, and then increases the target distortion at the next iteration if the total required rates is higher than the rate constraint and vice versa.

Finally, we determine the allocated FGS rate for each user, $rf_{i,j}$, using a similar linear combination as in (3.12):

$$rf_{i,j} = \min\{w_p R_{i,j}^p + (1 - w_p) R_{i,j}^f, R_{i,j}\} \quad (3.24)$$

where $R_{i,j}^p$ represents the FGS rate for the i^{th} user in the j^{th} frame (time slot) in order to maintain the same quality as the previous frame, and $R_{i,j}$ is the upper bound in (3.17).

3.4.3 Differentiated Service

Differentiated service refers to a service in which each user receives different quality according to his/her service agreement with the server. We consider a scenario that at the beginning of the service, each user submits his/her priority request, quantified by $DS_i \in (0, 1]$, such that the average distortion received by each user normalized by DS_i is constant:

$$\frac{aveMSE_1}{DS_1} = \frac{aveMSE_2}{DS_2} = \dots = \frac{aveMSE_N}{DS_N}. \quad (3.25)$$

In other words, a user who specifies a smaller value of DS_i (and possibly pays a premium fee in return) will receive a higher overall perceptual quality. This can be achieved by modifying the optimization problem in (3.23) as follows:

$$\min_{R_{1,j}^f, \dots, R_{N,j}^f} \overline{D_j} \quad (3.26)$$

subject to

$$\begin{cases} \overline{D}_j \triangleq \frac{D_{1,j}(R_{1,j}^f)}{DS_1} = \dots = \frac{D_{N,j}(R_{N,j}^f)}{DS_N}, \\ \sum_{i=1}^N R_{i,j}^f \leq \beta R_j. \end{cases}$$

The uniform quality problem of (3.23) is a special case of (3.26) when all DS_i 's are the same. This generalized optimization problem can also be solved using bi-section search. We present the complete multiuser algorithm in Table 3.3.

3.5 Simulation Results

In this section, we examine the performance of the proposed low-delay resource allocation algorithm with low-fluctuation of quality (*LDLF*), and compare it with two alternatives. The first alternative is the constant-bitrate (CBR) approach, which assigns a constant bit rate to each frame. The second alternative is a look-ahead sliding-window algorithm (*SWLF*) with buffer constraints adapted from [124], the details of which are given in Appendix A.2. Three statistics are used to evaluate the proposed algorithm and the two alternatives: the average PSNR (*avePSNR*), the mean of absolute difference of PSNR (*madPSNR*), and the overall channel utilization (*ChUtiliz*).

3.5.1 Simulation Setup

We concatenate 15 QCIF (176×144) video sequences to form one testing video sequence of 5760 frames. The 15 sequences are 300-frame *Akiyo*, 360-frame *car-phone*, 480-frame *Claire*, 300-frame *coastguard*, 300-frame *container*, 390-frame *foreman*, 870-frame *grandmother*, 330-frame *hall objects*, 150-frame *Miss American*, 960-frame *mother and daughter*, 300-frame *MPEG4 news*, 420-frame *salesman*, 300-frame *silent*, 150-frame *Suzie*, and 150-frame *Trevor*. The base layer is gen-

Table 3.3: The proposed multiuser resource allocation algorithm (M-LDLF)

<p>1. Initialization:</p> <p>$j = 1$, Set $RDSC=1$, $B_0^e = 0$, $B_{i,0}^e = 0$, $B_{i,1}^d = \dots = B_{i,d_i}^d = 0$, $R_{i,0}^p = rb_{i,1}$, $rb_{i,0} = rf_{i,0} = 0$, $w_p = w_L$, and $MSE_{i,0} = D_{i,1}(rb_{i,1})$, $\forall i$.</p> <p>2. While the last frame of this video is not reached:</p> <p>a) Calculate the budget factor and weighting factor</p> $\bar{r}b = \frac{1}{j - \max(j-L, 0)} \sum_{i=1}^N \sum_{k=\max(j-L, 0)+1}^j rb_{i,k},$ $\beta = \frac{C_{max} - \bar{r}b}{B_{max}^e + C_{max} - \bar{r}b} + \Delta\beta,$ $R_{i,j}^p = D_{i,j}^{-1}(MSE_{i,j-1}) - rb_{i,j}, \forall i.$ <p>If $\frac{1}{N} \sum_{i=1}^N (R_{i,j}^p - R_{i,j-1}^p / R_{i,j-1}^p) \geq S_T$, then $RDSC=j$. If $j \in [RDSC, RDSC + P_T]$, then $w_p = w_L$, else $w_p = W(B_{j-1}^e)$.</p> <p>b) Select the channel transmission rate</p> $C_{i,j}^U = B_{i,max}^d - B_{i,j+d_i^e-1}^d + r_{i,j-d_i^d} \forall i,$ $C_{i,j}^L = \max\{-B_{i,j+d_i^e-1}^d + r_{i,j-d_i^d}, 0\} \forall i,$ <p>If $\sum_{i=1}^N C_{i,j}^U \geq C_{max}$, then</p> $C_{i,j} = \min\{C_{i,j}^L + (C_{max} - \sum_{i=1}^N C_{i,j}^L) \frac{r_{i,j-1}}{\sum_{i=1}^N (r_{i,j-1})}, C_{i,j}^U\},$ <p>Else $C_{i,j} = C_{i,j}^U$.</p> <p>c) Select FGS rate</p> <p>i) $R_{i,j} = B_{max}^e + C_{i,j} - B_{i,j-1}^e - rb_{i,j}, \forall i$.</p> <p>ii) $R_j = B_{max}^e + \sum_{i=1}^N C_{i,j} - \sum_{i=1}^N B_{i,j-1}^e - \sum_{i=1}^N rb_{i,j}$,</p> <p>iii) Solve the optimization problem:</p> $\min_{R_{1,j}^f, \dots, R_{N,j}^f} \bar{D}_j, \quad \text{subject to}$ <p>Quality Constraint $\bar{D}_j \triangleq \frac{D_{1,j}(R_{1,j}^f)}{DS_1} = \dots = \frac{D_{N,j}(R_{N,j}^f)}{DS_N}$;</p> <p>Rate Constraint $\sum_{i=1}^N R_{i,j}^f \leq \beta R_j$.</p> <p>iv) $rf_{i,j} = \min\{w_p R_{i,j}^p + (1 - w_p) R_{i,j}^f, R_{i,j}\}, \forall i$.</p> <p>d) Update the encoder buffer occupancy information:</p> $B_{i,j}^e = \max\{B_{i,j-1}^e + rb_{i,j} + rf_{i,j} - C_{i,j}, 0\}.$ <p>If $B_{i,j}^e = 0$, then $C_{i,j} = B_{i,j-1}^e + rb_{i,j} + rf_{i,j}, \forall i$.</p> <p>e) Update the decoder buffer occupancy information:</p> $B_{i,j+d_i^e}^d = B_{i,j+d_i^e-1}^d + C_{i,j} - r_{i,j-d_i^d} \forall i.$ <p>f) $j = j + 1$.</p>
--

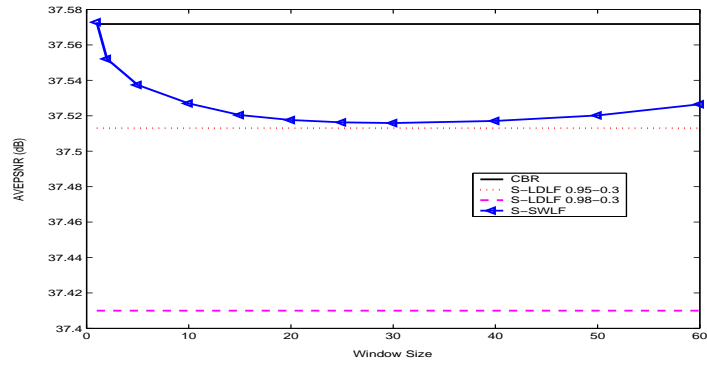
erated by MPEG-4 encoder with a fixed quantization step of 30 and the GOP pattern is 29 P frames after one I frame. All frames of FGS layer have up to six bit planes. For N users in this system, we allocate $N * 80K$ bits for the server buffer and the shared maximal channel capacity is $N * 960$ Kbps. Each user has a small decoder buffer of $400K$ bits. For each user, the transmission delay, d_i^c , is 3 frames and initial playback delay, d_i^d , is 3 frames. The parameters $(a, b, P_T, S_T, \Delta\beta, L)$ used in the *LDLF* algorithm are set to $(B_{max}^e/4, 0.75, 3, 0.3, 0.01, 30)$.

3.5.2 Simulation Results for Single-User Case

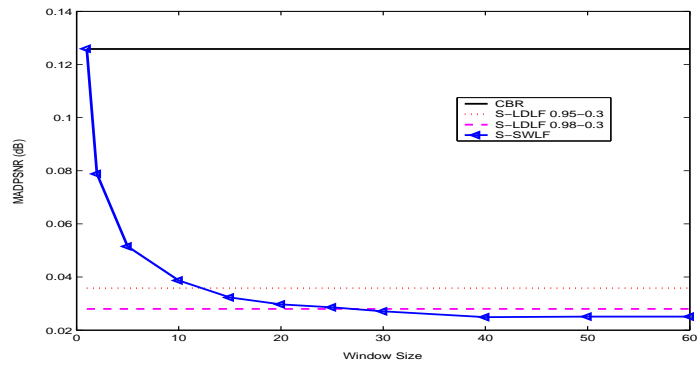
For the single-user system, the video content is picked from frame 301 to 2100, corresponding to the video sequences of *carphone*, *Claire*, *coastguard*, *container*, and *foreman*. Figure 3.6 shows the *avePSNR*, *madPSNR*, and *ChUtiliz* using the three different algorithms.

The solid line with triangle makers represents the results of single-user *SWLF* algorithm (*S-SWLF*) with different window sizes. The solid line represents the results of CBR approach, which encodes each frame at a constant bit rate of $32K$ bits. The dotted and dashed lines represent the results of the proposed single-user *LDLF* algorithm (*S-LDLF*) with $(w_H, w_L) = (0.95, 0.3)$ and $(0.98, 0.3)$, respectively. As the CBR and *S-LDLF* algorithms do not have the window parameter, we plot their results as horizontal lines to allow for the performance comparison among CBR, *S-LDLF*, and *S-SWLF* of different window sizes.

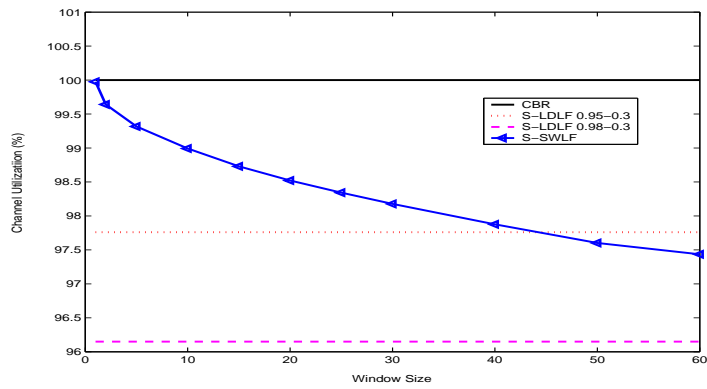
As we can see from Figure 3.6, CBR approach provides the highest average perceptual quality and channel utilization. However, CBR has the worst fluctuation of visual quality. In contrast, our experiment shows that variable rate control for video, such as *S-LDLF* and *S-SWLF*, can provide more consistent quality. For the



(a) avePSNR



(b) madPSNR



(c) Channel Utilization

Figure 3.6: Comparison of CBR, S-SWLF, and the proposed S-LDLF rate control algorithms for video frame 301~2100 from the testing sequence.

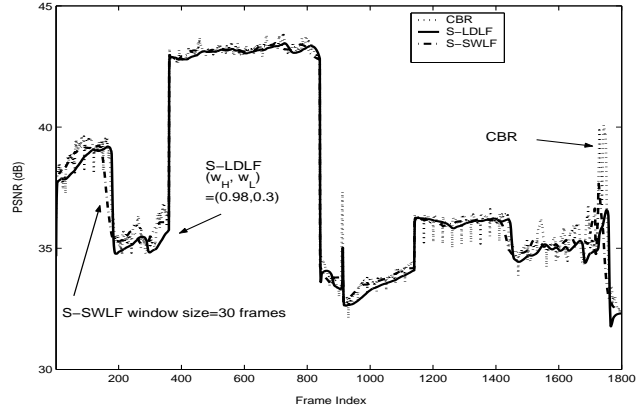
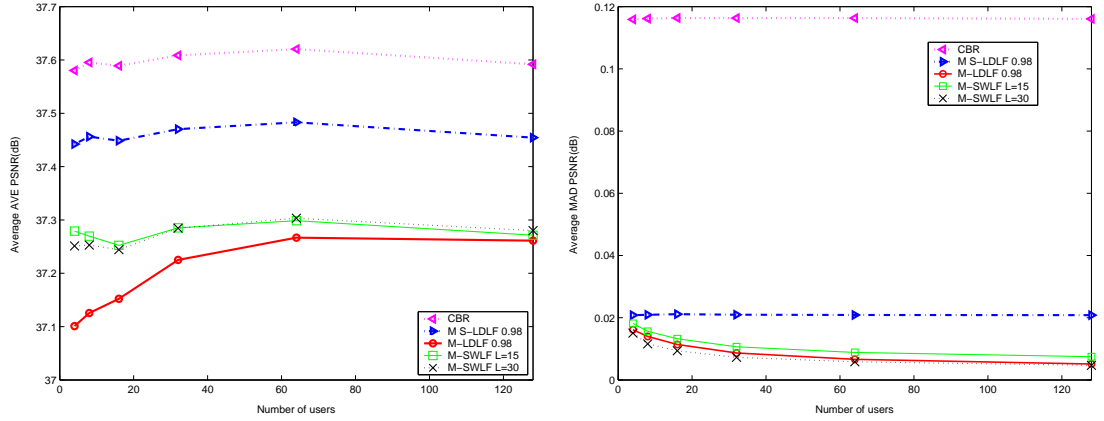


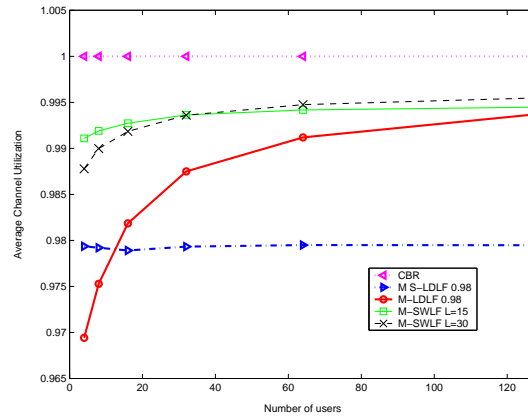
Figure 3.7: Frame-by-frame PSNR comparison of the CBR, S-SWLF, and the proposed S-LDLF approaches

proposed *S-LDLF* algorithm, a higher value of w_H gives smaller $madPSNR$ and a little lower $avePSNR$ as expected. We also observe that $madPSNR$ decreases when the window size increases in *S-SWLF* algorithm. To provide sufficient smoothening, the window size L of the *S-SWLF* algorithm needs to be at least the size of half to one GOP to cover an I-frame of high data rate, which is 15~30 frames in our experiment. We compare *S-SWLF* with window size 30 frames with *S-LDLF* with $(w_H, w_L) = (0.98, 0.3)$, and present the PSNR results for each frame in Figure 3.7. We can see that *S-SWLF* and *S-LDLF* have similar performance in terms of visual quality. However, to achieve this comparable performance, *S-SWLF* requires about one-second more delay ($L = 30$ frames) and a corresponding large amount of extra storage for the look-ahead frames on the encoder side, while the proposed *S-LDLF* algorithm requires no extra delay and storage on the encoder side.



(a) Average *avePSNR* among all users

(b) Average *madPSNR* among all users



(c) Average *ChUtiliz* among all users

Figure 3.8: Comparison of CBR, M-SWLF, M S-LDLF, and M-LDLF systems under uniform service for all users. Shown here are three statistics averaged among all users.

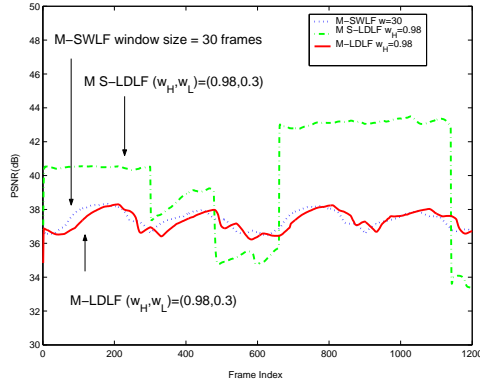
3.5.3 Simulation Results for Multiuser Case

For the multiuser system, the content program for each user is 1200 frames long and starts from a randomly selected I frame of the testing video source. If the length of this video source is not long enough, we loop from the beginning of the testing sequence. We repeat the simulation multiple times for a total of about 25000 user cases to obtain the averaged results for the systems with 4, 8, 16, 32, 64 and 128 users.

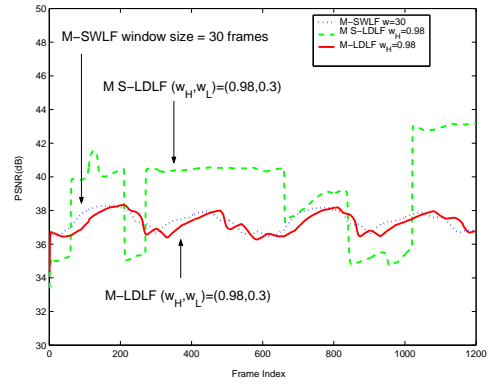
We first demonstrate the performance when all users request the same level of visual quality. Figure 3.8 shows the average of all users' $avePSNR$, $madPSNR$, and $ChUtiliz$ for different number of users using four algorithms, namely, the CBR algorithm (CBR), the multiuser $SWLF$ algorithm ($M-SWLF$) with different window sizes, the multiple single-user approach using the above $S-LDLF$ ($M S-LDLF$) with $(w_H, w_L) = (0.98, 0.3)$, and the proposed multiuser $LDLF$ ($M-LDLF$) approach with $(w_H, w_L) = (0.98, 0.3)$. The CBR approach assigns each user a fixed encoding rate, 32K bits per frame. The $M S-LDLF$ system provides individual encoder buffer (80K bits) and channel bandwidth (960Kbps) to each user using $S-LDLF$ algorithm. As we can see from Figure 3.8, the CBR approach suffers from much higher fluctuation of visual quality than any other approaches, suggesting once again the need of variable rate control. Among the three approaches providing VBR video, the two joint resource allocation approaches for multiuser system, $M-SWLF$ and $M-LDLF$, provide smaller fluctuation of visual quality than the individual resource allocation scheme ($M S-LDLF$). The more users a system has, the higher possibility we can take the advantage of the variations of video content similar to those in multiplexing [114] and offer desired quality to each user through dynamic bandwidth allocation. Comparing these two dynamic multiuser

algorithms, the fluctuation of visual quality of *M-LDLF* algorithm is between the quality fluctuations of the *M-SWLF* algorithm with window sizes of 15 and 30 frames; and the performance of both visual quality measurements, *avePSNR* and *madPSNR*, of the *M-LDLF* algorithm approaches the results of *M-SWLF* with window size 30 frames when the number of users increases.

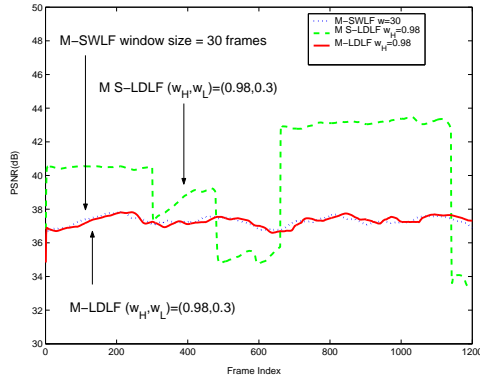
To compare the frame-by-frame PSNR of the three algorithms, *M S-LDLF*, *M-SWLF*, and *M-LDLF*, we simulate the scenario in which the content program for user i is 1200 frames long and starts from frame $600 \times (i - 1) + 1$ of the testing video source. Figure 3.9 shows the frame-by-frame PSNR of the first and tenth users in the *M S-LDLF*, *M-SWLF*, and *M-LDLF* systems when there are 16 and 32 users in the systems, respectively. Again, we see that the dynamic multiuser approaches (*M-LDLF* and *M-SWLF*) can provide more uniform quality than the multiple single-user approach both within a scene and when crossing scene boundaries. When the number of users increases, the gain from joint resource allocation is more significant, providing more uniform quality and less quality fluctuation. Between the two dynamic multiuser approaches, our proposed *M-LDLF* approach can achieve similar perceptual quality to that of *M-SWLF* approach with large window size (30 frames); however, similar to the single-user case, the prior work *M-SWLF* needs a longer delay (1 second) and a substantially larger storage than the propose approach. This additional storage is for keeping the look-ahead data of all users. In the example illustrated above, *M-SWLF* needs an extra storage of $30 \text{ frames/user} \times 32 \text{ users} = 960 \text{ frames}$. As a result, the proposed approach has higher system scalability than the *M-SWLF* approach. This makes the proposed scheme an attractive choice for building a large system to accommodate many users.



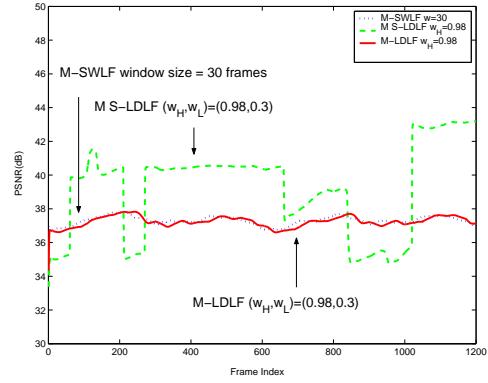
(a) 1st user in a 16-user system



(b) 10th user in a 16-user system

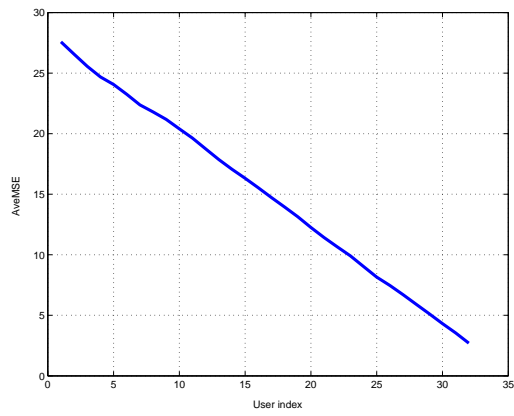


(c) 1st user in a 32-user system

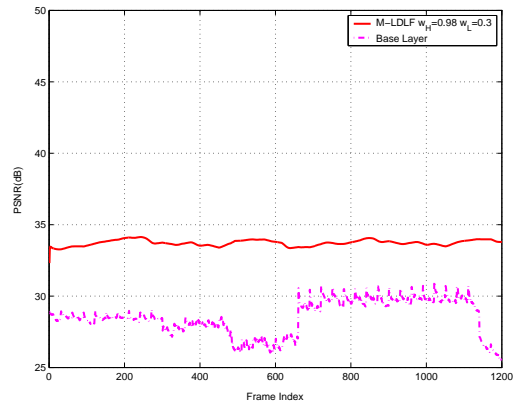


(d) 10th user in a 32-user system

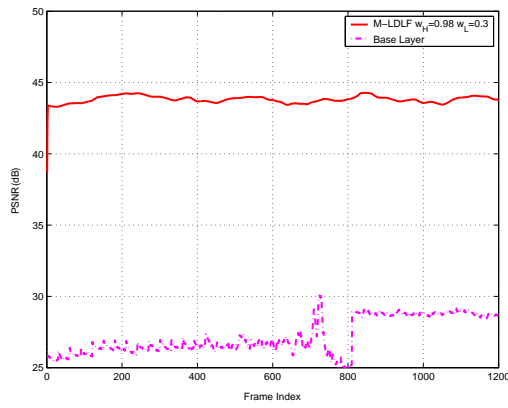
Figure 3.9: Frame-by-frame PSNR results of the first and tenth user in the M S -LDLF, M -SWLF, and M -LDLF systems



(a) AveMSE for each user



(b) PSNR for the 1st user



(c) PSNR for the 32nd user

Figure 3.10: Results of a 32-user system using M-LDLF with differentiated service

We now use Figure 3.10 to demonstrate the differentiated service of a 32-user system by keeping the same system settings as in Figure 3.9 except setting the differentiated service priority for the i -th user as $DS_i = 1 - 0.9(i - 1)/31$. Figure 3.10(a) illustrates the average MSE for each user. As we can see, the proposed algorithm can achieve the required differentiated service priority, which is almost a linear line as we have designed for. Figure 3.10(b) and 3.10(c) highlight the received PSNR for the first and the last user in this system, who request the lowest and the highest video quality, respectively.

3.6 Chapter Summary

In summary, we have proposed an efficient bandwidth resource allocation algorithm for streaming multiple MPEG-4 FGS video sequences. By exploring the intra- and inter-frame R-D characteristics of MPEG-4 FGS codec, we present a control policy to achieve an excellent trade-off between the average quality and quality fluctuation criteria. We demonstrate that multiuser systems with dynamic joint resource allocation provide more consistent quality than the multiple single-user approaches that do not dynamically share resources. Evaluating the video quality in terms of the average distortion and the quality fluctuation, our algorithm gives excellent performance comparable to those by the general look-ahead sliding-window approach. But compared to the existing approaches, our algorithm has higher system scalability, as it does not need a delay of dozens of frames' long and does not require extra storage proportional to the number of users. Therefore, the proposed multiuser resource allocation algorithm with low delay and low fluctuation of quality can serve as an efficient and effective building block for real-time multiuser broadband communications.

Chapter 4

Video over Wireless Downlink

4.1 Introduction

In Chapter 3, we have demonstrated the advantages of dynamical and joint resource allocation in a multiuser video system. In this chapter, we will extend this concept to a video system that transmits multiple compressed video programs over band-limited wireless fading channels. The challenge to support such services is how to effectively allocate radio and video resources to each video stream. To facilitate resource management, a system with highly adjustable radio and video resources is preferred. For the radio resource, the wireless communication system should provide high data rates to accommodate multimedia transmission and equip multi-dimensional diversity so that radio resources can be dynamically distributed according to users' needs and channel conditions. For the video source coding, the video codec should have high scalability to aid rate adaptation to achieve the required quality. The promising development of next generation video source codecs and wireless networks can provide such flexibility as discussed in Chapter 2.

In this chapter, we propose a framework to provide multiple video streams to

different users using dynamic distortion control based on the next generation of video and wireless systems. The proposed framework has the following features. First, the system dynamically gathers the information of system resources from different components to capture the time-heterogeneity of video sources and time-varying characteristics of channel conditions. Subject to delay constraint, the system explores multi-dimensional diversity among users and across layers, performs joint multiuser cross-layer resource allocation optimization, and then distributes the system resources to each user. The benefit for such joint consideration is the higher utilization of system resources. We will first study a scenario in which we transmit multiple scalable video bitstreams over OFDMA networks to demonstrate the proposed strategy [92,95]. At the end of this chapter, based on this principle, we will study another scenario in which we transmit video bitstreams over CDMA networks [90].

4.2 System Description

There are three major subsystems in the proposed wireless video system, namely, the video source codec subsystem, the OFDMA subsystem, and the resource allocator subsystem. We first review the video source codec subsystem along with the corresponding R-D characteristics, and describe the OFDMA subsystem with adaptive modulation and adaptive channel coding. Then, we present the proposed framework for transmitting multiple scalable video bitstreams over OFDMA networks.

4.2.1 Video Source Codec Subsystem

To transmit video programs over wireless networks, a system should be able to adjust the video source bit rates according to the varying channel conditions. A highly scalable video codec is desired since it provides flexibility and convenience in reaching the desired visual quality or the desired bit rate. Although current MPEG-4 FGS and FGST can provide high flexibility; however, their overall qualities are worse than the non-scalable coding results, and there remains a non-scalable base layer. The development of 3-D subband video coding [16, 35, 46, 64, 69, 99] provides an alternative to compress video with full scalability. The current MPEG-4 Part-10 Scalable Video Coding [39] has called for proposals based on 3-D subband coding to achieve spatial scalability, temporal scalability, SNR scalability, and multiple adaptations. Unlike the motion compensated video codec based on block matching (such as H.263 and MPEG-4), the 3-D subband coding explores the spatiotemporal redundancies via a 3-D subband transform. Extending the bit allocation ideas from the EBCOT algorithm for image compression [100], the 3-D embedded wavelet video codec (EWW) [35] outperforms MPEG-4 for sequences with low or moderate motion and has comparable performance to MPEG-4 for most high-motion sequences. Moreover, the rate-distortion information can be predicted during the encoding procedure and provide a one-to-one mapping between rate and distortion such that we can achieve the desired perceptual quality or the targeted bit rate. Thus, we adopt the EWW codec in the proposed framework as an example. We can easily incorporate other codecs with similar coding strategy into the proposed scheme.

The EWW encoder consists of four stages [35], namely, 3-D wavelet transform, quantization, bit plane arithmetic coding, and rate-distortion optimization. At

the first stage, we collect a group of frames (GOF) as an encoding unit and apply 1-D dyadic temporal decomposition to obtain temporal subbands. The 2-D spatial dyadic decomposition is applied in each temporal subband to obtain wavelet spatiotemporal subbands (or “subbands” for short). At the second stage, a uniform quantizer is used for all wavelet coefficients in all subbands. At the third stage, fractional bit plane arithmetic coding is applied to each subband. Except that the most significant bit plane (MSB) has only one coding pass, every bit plane is encoded into three coding passes, namely, significance propagation pass, magnitude refinement pass, and normalization pass. Each coding pass can be treated as a candidate truncation point and the EWV decoder can decode the truncated bitstream containing an integer number of coding passes in each subband. The more consecutive coding passes of each subband a receiver receives, the higher decoded video quality we have. The coding passes among all subbands can be further grouped into several quality layers such that the received video quality can be refined progressively by receiving more layers. At the last stage, the encoder determines which coding passes are included in the output bit stream subject to quality or rate constraint.

To maintain the coding efficiency, the R-D curve in each subband should be convex [100]. Some coding passes in a subband cannot serve as feasible truncation points to maintain the convexity and they will be pruned from the truncation point list. To facilitate the discussion, we call all the coding passes between two truncation points as a coding pass cluster.

Consider now there are a total of B subbands for the k^{th} user and the subband b has $T_k^{b,max}$ coding pass clusters. We can measure the rate and the corresponding decrease in normalized mean squared distortion of the t^{th} coding pass cluster in

subband b for the k^{th} user [100], and denote them as $\Delta r_{t,b,k}$ and $\Delta d_{t,b,k}$, respectively. We divide the whole duration for transmitting a total of L quality layers into L transmission intervals with equal length. The l^{th} quality layer is transmitted at the l^{th} transmission interval. The received distortion D_k^l and rate R_k^l for quality layers 0 to l can be expressed as:

$$D_k^l = D_k^{max} - \sum_{b=0}^{B-1} \sum_{t=0}^{T_k^{b,l}-1} \Delta d_{t,b,k}, \quad (4.1)$$

$$R_k^l = \sum_{q=0}^l \Delta R_k^q. \quad (4.2)$$

Here D_k^{max} is the distortion without decoding any coding pass cluster,

$$\Delta R_k^l = \sum_{b=0}^{B-1} \sum_{t=T_k^{b,l-1}}^{T_k^{b,l}-1} \Delta r_{t,b,k}, \quad (4.3)$$

and $T_k^{b,l}$ is the total number of coding pass clusters of subband b in the quality layers 0 to l , which satisfies:

$$0 \leq T_k^{b,l-1} \leq T_k^{b,l} \leq T_k^{b,max}, \quad \forall b \text{ and } 0 < l < L. \quad (4.4)$$

Define the number of coding pass clusters for subband b in quality layer l as $\Delta T_k^{b,l} = T_k^{b,l} - T_k^{b,l-1}$ and for all subbands

$$\Delta \mathbf{T}_k^l = [\Delta T_k^{0,l}, \Delta T_k^{1,l}, \dots, \Delta T_k^{B-1,l}]. \quad (4.5)$$

We also define a matrix $\Delta \mathbf{T}^l$ whose k^{th} row is $\Delta \mathbf{T}_k^l$. Thus, in each transmission interval l , the source coding part of our system determines the coding pass cluster assignment $\Delta \mathbf{T}_k^l$ and packetizes them as a quality layer for each user. We use Figure 4.1 to illustrate the relationship among coding pass, subband, and quality layer. Note that owing to different content complexities and motion activities shown in video sources, the R-D information should be evaluated for each GOF of each user to capture the characteristics of the corresponding bitstream.

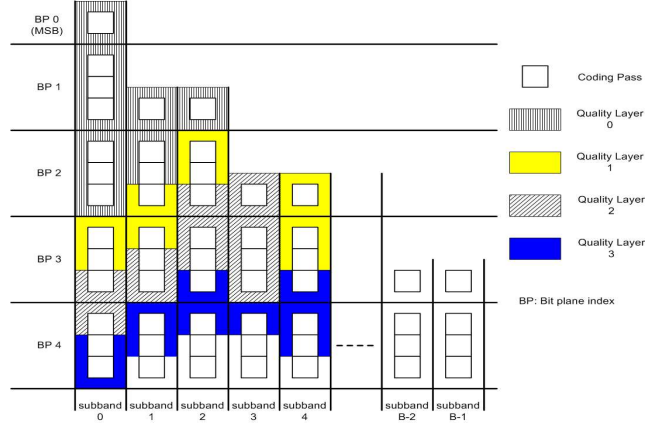


Figure 4.1: Illustration of the relationship among coding pass, subband, and quality layer.

4.2.2 OFDMA Subsystem

To provide high data transmission rate, OFDM system is a promising modulation scheme and has been adopted in the current technology, such as Digital Audio Broadcasting (DAB), Digital Video Broadcasting (DVB), WLAN standard (IEEE 802.11 a/g), and Wireless Metropolitan Area Networks (WMAN) standard (IEEE 802.16a). Compared to the traditional OFDM system, an OFDMA system has higher adjustability for dynamic allocation of resources such as subcarrier, rate, and transmission power; as reviewed in Chapter 2.2. Therefore, an OFDMA system can explore time, frequency, and multiuser diversities to improve system performances,

We consider a downlink scenario of a single-cell OFDMA system in which there are K users randomly located. The system has N subcarriers and each subcarrier has bandwidth of W . We use an indicator $a_{kn} \in \{0, 1\}$ to represent whether the n^{th} subcarrier is assigned to user k . Note that in a single-cell OFDMA system, each subcarrier can be assigned to at most one user, i.e., $\sum_{k=0}^{K-1} a_{kn} \in \{0, 1\}, \forall n$. The overall subcarrier-to-user assignment can be represented as a matrix \mathbf{A} with

$[\mathbf{A}]_{kn} = a_{kn}$. Let r_{kn} be the k^{th} user's transmission rate at the n^{th} subcarrier and the total rate for the k^{th} user can be expressed as $\sum_{n=0}^{N-1} a_{kn}r_{kn}$. The overall rate allocation can also be represented as a matrix \mathbf{R} with $[\mathbf{R}]_{kn} = r_{kn}$.

In mobile wireless communication systems, signal transmission suffers from various impairments such as frequency-selective fading due to multipath delay [75]. The continuous complex baseband representation of user k 's wireless channel impulse response is expressed as

$$g_k(\mathbf{t}, \tau) = \sum_i v_{k,i}(\mathbf{t})\delta(\tau - \tau_{k,i}), \quad (4.6)$$

where $v_{k,i}(\mathbf{t})$ and $\tau_{k,i}$ are the gain and the delay of path i for user k , respectively. In Rayleigh fading, the sequence $v_{k,i}(\mathbf{t})$ is modelled as a zero-mean circular symmetric complex Gaussian random variable with variance $\sigma_{v_{k,i}}^2$ proportional to $d^{-\alpha}$, where d is the distance and α is the propagation loss factor. All $v_{k,i}(t)$ are assumed to be independent for different paths. The root-mean-square (RMS) delay spread is the square root of the second central moment of the power delay profile:

$$\sigma_{k,\tau} = \sqrt{\overline{\tau_k^2} - (\bar{\tau}_k)^2}, \quad (4.7)$$

where $\overline{\tau_k^2} = \frac{\sum_i \sigma_{v_{k,i}}^2 \tau_{k,i}^2}{\sum_i \sigma_{v_{k,i}}^2}$ and $\bar{\tau}_k = \frac{\sum_i \sigma_{v_{k,i}}^2 \tau_{k,i}}{\sum_i \sigma_{v_{k,i}}^2}$.

After sampling at the receiver, the channel gain of OFDMA subcarriers can be approximated by the discrete samples of the continuous channel frequency response as

$$G_{kn}^h = \int_{-\infty}^{\infty} g_k(\mathbf{t}, \tau) e^{-j2\pi f\tau} d\tau \Big|_{f=nW, \mathbf{t}=hT_f}, \quad (4.8)$$

where T_f is the duration of an OFDM symbol and h is the sampling index. This approximation does not consider the effect of the smoothing filter at the transmitter and the front-end filter at the receiver.

We assume a slow fading channel where the channel gain is stable within each transmission interval.¹ The resource allocation procedure will be performed in each transmission interval. To facilitate the presentation, we omit h in the channel gain notation. The channel parameters from different subcarrier of different users are assumed perfectly estimated, and the channel information is reliably fed back from mobiles users to the base station in time for use in the corresponding transmission interval. Denote Γ_{kn} as the k^{th} user's signal to noise ratio (SNR) at the n^{th} subcarrier as:

$$\Gamma_{kn} = P_{kn}G_{kn}/\sigma^2, \quad (4.9)$$

where P_{kn} is the transmission power for the k^{th} user at the n^{th} subcarrier and σ^2 is the thermal noise power that is assumed to be the same for each subcarrier of different users. Further, let $[\mathbf{G}]_{kn} = G_{kn}$ be the channel gain matrix and $[\mathbf{P}]_{kn} = P_{kn}$ the power allocation matrix. For downlink system, because of the practical constraints in implementation, such as the limitation of power amplifier and consideration of co-channel interferences to other cells, the overall power is bounded by P_{max} , i.e., $\sum_{k=0}^{K-1} \sum_{n=0}^{N-1} a_{kn}P_{kn} \leq P_{max}$.

The goal of the proposed framework is to provide good subjective video quality of the reconstructed video. Since the distortion introduced by channel error is typically more annoying than the distortion introduced by source lossy compression, the system should keep the channel-induced distortion at a negligible portion of the end-to-end distortion so that the video quality is controllable by the source coding subsystem. This can be achieved when we apply an appropriate amount of channel coding to keep the BER after the channel coding below some targeted BER

¹In practice, the duration of a transmission interval can be adjusted shorter enough so that the channel gain is stable within a transmission interval.

threshold, which is 10^{-6} in our system and achievable in most 3G/4G systems [21]. In addition, joint consideration of adaptive modulation, adaptive channel coding, and power control can provide each user with the ability to adjust each subcarrier's data transmission rate r_{kn} to control video quality while meeting the required BER.

We focus our attention on MQAM modulation and convolutional codes with bit interleaved coded modulation (BICM) as they provide high spectrum efficiency and strong forward error protection, respectively. We list the required SNRs' and the adopted modulation with convolutional coding rates to achieve different supported transmission rates under different BER requirement in Table 4.1 based on the results in [1]. Given a targeted BER, there is a one-to-one mapping between the selected transmission rate and the chosen modulation scheme with convolutional coding rate when the required SNR is satisfied. In this case, determining r_{kn} is equal to determine the modulation and channel coding rate. For each rate allocation $[\mathbf{R}]_{kn}$, the corresponding power allocation $[\mathbf{P}]_{kn}$ should maintain the SNR in (4.9) larger than the corresponding value listed in Table 4.1 to achieve the BER requirement. To facilitate our discussion, we define the feasible set of the transmission rate in Table 4.1 as $\nu = \{\nu_0, \nu_1, \nu_2, \dots, \nu_Q\}$ and the corresponding set of the required SNR for $\text{BER} \leq 10^{-6}$ as $\rho = \{\rho_0, \rho_1, \rho_2, \dots, \rho_Q\}$. Here, $\nu_0 = 0$ and $\rho_0 = 0$, and Q represents the number of combinations for different modulation with convolutional coding rates, which is 11 in our case. All transmission rate r_{kn} 's should be selected from the feasible rate set ν .

Table 4.1: Required SNR and transmission rate using adaptive modulation and convolutional coding rates [1]

k	Rate ν_k	Modulation	Convolutional Coding Rate	SNR ρ_k (dB) for BER $\leq 10^{-6}$	SNR (dB) for BER $\leq 10^{-5}$
1	1W	QPSK	1/2	4.65	4.09
2	1.33W	QPSK	2/3	6.49	5.86
3	1.5W	QPSK	3/4	7.45	6.84
4	1.75W	QPSK	7/8	9.05	8.44
5	2W	16QAM	1/2	10.93	10.04
6	2.66W	16QAM	2/3	12.71	12.13
7	3W	16QAM	3/4	14.02	13.29
8	3.5W	16QAM	7/8	15.74	15.01
9	4W	64QAM	2/3	18.50	17.70
10	4.5W	64QAM	3/4	19.88	18.99
11	5.25W	64QAM	7/8	21.94	21.06

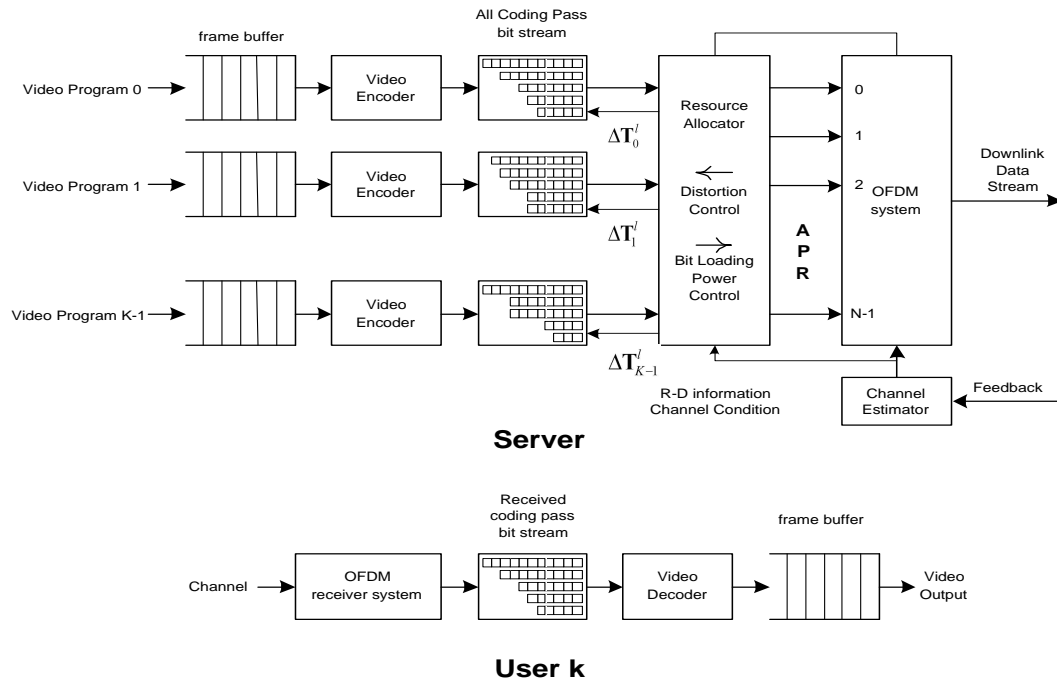


Figure 4.2: System block diagram

4.2.3 EWV Video over OFDMA

The block diagram of the proposed wireless video system is shown in Figure 4.2. The upper and lower parts show the modules located in the server side and the mobile user side, respectively. For the server side, the server buffers each user's incoming video frames in the user's frame buffer. After collecting a GOF with H frames for each user, the server moves those raw video frames to a wavelet video encoder for compression as a coding pass bit stream. The selected coding pass clusters will be transmitted during the next GOF transmission time of H/F second long, where F is the video frame refreshing rate. To capture the varying channel conditions and video content characteristics, the resource allocator should obtain the channel information for each transmission interval from the channel estimator and the R-D information of each GOF from the video coder. With the estimated channel conditions, the resource allocator can predict how many data rates with $\text{BER} \leq 10^{-6}$ the wireless networks can support in the next transmission interval. With the R-D information, the resource allocator can estimate the qualities of the reconstructed videos after decoding at each mobile terminal. By jointly considering the R-D information and the estimated channel conditions, the resource allocator performs resource optimization and distributes video and radio resources to each video stream in each transmission interval. According to the allocated resources, the source coding subsystem will group the selected coding pass clusters into a quality layer for each user and pass them to the transmission system; and the OFDMA subsystem will load the video data to be transmitted to different subcarriers at a controlled amount of power. On the mobile user side, an OFDMA receiver buffers the received data until the end of the current GOF transmission time. Then, those received data are moved to a wavelet video de-

coder for decoding and the decoded frames are sent for display during the next GOF transmission time.

Since we only know the channel conditions provided by the channel estimator in the near future within the next transmission interval and the GOF bitstreams are transmitted across L transmission intervals, it is necessary to break down the optimization problem into a sequential optimization problem and solve each problem in each transmission interval. There are two different objectives we want to achieve in each transmission interval: fairness and efficiency. To ensure the fairness among all users, we formulate the problem as a min-max optimization problem to minimize the maximal (weighted) end-to-end distortion among all users. Maintaining short-term fairness in each transmission interval ensures the long-term fairness for GOFs [49]. To achieve high resource-allocation efficiency in terms of a high overall video quality, we formulate the problem as an optimization problem to minimize the overall end-to-end distortion among all users. We will discuss the fairness and efficiency problems in Chapter 4.3 and Chapter 4.4, respectively. The tradeoff between efficiency and fairness will be addressed in Chapter 4.5.

4.3 Optimization in Resource Allocator: Focusing on Fairness

In this section, we consider how to achieve fair video quality among all users in a transmission interval and formulate this problem as a min-max problem. Given the integer programming nature of the problem, we propose a three-stage suboptimal algorithm to solve the optimization problem in real time.

4.3.1 Formulation of the Fairness Problem

At the beginning of the l^{th} transmission interval, according to the channel information and subject to the transmission delay constraint as one transmission interval long, the resource allocator minimizes the maximal distortion received among all users as follows:

$$\min_{\mathbf{A}, \mathbf{R}, \Delta \mathbf{T}^l} \max_k w_k \cdot f(D_k^l) \quad (4.10)$$

$$\text{subject to } \left\{ \begin{array}{l} \text{Subcarrier Assignment: } \sum_{k=0}^{K-1} a_{kn} \leq 1, a_{kn} \in \{0, 1\}, \forall n; \\ \text{Subcarrier Rate: } r_{kn} \in \nu, \quad \forall k, n; \\ \text{User Rate: } 0 \leq \Delta R_k^l \leq \sum_{n=0}^{N-1} a_{kn} r_{kn}, \forall k; \\ \text{Power: } \sum_{k=0}^{K-1} \sum_{n=0}^{N-1} a_{kn} P_{kn} \leq P_{max}; \end{array} \right.$$

where w_k is the quality weighting factor and $f(\cdot)$ the perceptual distortion function. We solve this optimization problem by selecting the values of subcarrier assignment matrix \mathbf{A} , rate assignment matrix \mathbf{R} , and coding pass cluster assignment $\Delta \mathbf{T}^l$ subject to four constraints: the first constraint is on subcarrier assignment that a subcarrier can be assigned to at most one user; the second one restricts the subcarrier rate to be selected only from the feasible rate set ν ; the third one is that the user's overall assigned rate in (4.3) should be no larger than the overall assigned subcarrier rate; and the fourth one is on the maximal power available for transmission. Note that the system can provide differentiated service by setting $\{w_k\}$ to different values according to the quality levels requested by each user. As a proof of concept, we consider the case of $w_k = 1, \forall k$ and $f(D_k^l) = D_k^l$ for providing uniform quality among all users here. The proposed solution can be easily extended to other quality weighting factors and quality functions, and we will demonstrate the ability for providing differentiated service in Chapter 4.6. The problem in (4.10) is a multi-dimension generalized assignment problem, which is

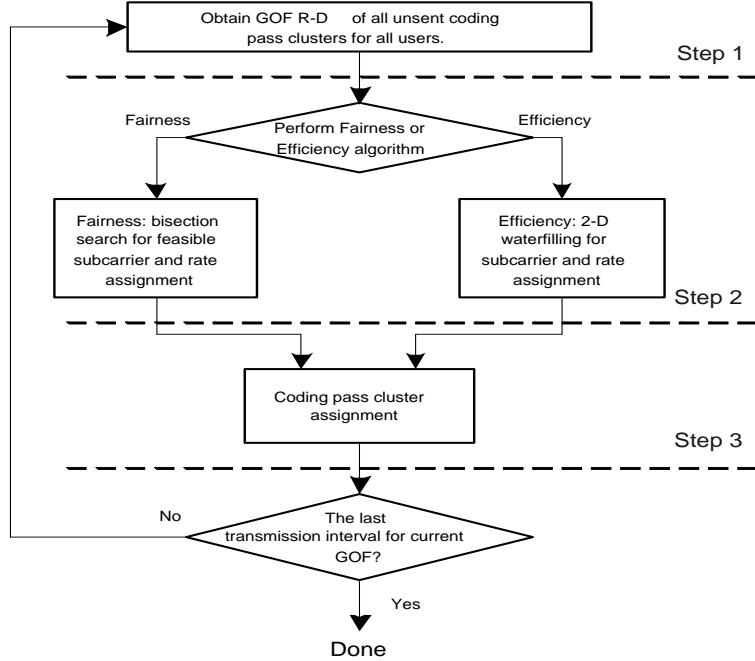


Figure 4.3: Flowchart of the proposed algorithm

an *NP* hard problem [60]. In a real-time system, a fast approximation algorithm with good performance is needed and will be designed next.

4.3.2 Proposed Algorithm for Fairness

We propose a three-stage fast algorithm to solve the optimization problem (4.10). As illustrated by the flowchart in Figure 4.3, at the first stage, we obtain continuous GOF R-D functions that provide a distortion-to-rate mapping to facilitate the resource allocation. At the second stage, we determine the subcarrier assignment matrix \mathbf{A} and rate assignment matrix \mathbf{R} to find the largest distortion reduction that the OFDMA system can support. This goal can be achieved through a bisection search on the R-D functions obtained at Stage 1. At the third stage, the coding pass cluster assignment $\Delta \mathbf{T}^l$ is decided subject to the allocated subcarrier

Table 4.2: GOF R-D used in each transmission interval

a) Sort all $\{\lambda_{t,b,k}\}$ for all $t \geq T_k^{b,l-1}$ in a decreasing order for user k
b) For each (t, b) for user k , we have indices $I_k(t, b) \in \{1, 2, \dots, M_k^l\}$ and $I_k^{-1}(m) \in \{(t, b)\}$ s.t. $\lambda_{i,b,k} > \lambda_{j,c,k}$ for $I_k(i, b) < I_k(j, c)$
c) Set $D_k^l[0] = D_k^{l-1}$ and $R_k^l[0] = 0$ For $m = 1, \dots, M_k^l$ $D_k^l[m] = D_k^l[m-1] - \Delta d_{I_k^{-1}(m),k} $ $R_k^l[m] = R_k^l[m-1] + \Delta r_{I_k^{-1}(m),k}$
d) Construct the continuous GOF R-D function $D_k^l(\gamma_k^l) = \frac{D_k^l[m+1] - D_k^l[m]}{R_k^l[m+1] - R_k^l[m]} (\gamma_k^l - R_k^l[m]) + D_k^l[m],$ for $R_k^l[m] \leq \gamma_k^l \leq R_k^l[m+1]$ and $m = 0, \dots, M_k^l - 1.$

and rate assignment at Stage 2. We explain the details of each stage below:

Stage 1

At this stage, a continuous GOF R-D function of the unsent coding pass clusters for each user is obtained. The goal for determining the GOF R-D function is to provide a one-to-one mapping between rate and distortion such that we can know the amount of rate increment necessary for a given amount of reduction in distortion.

Suppose there are M_k^l unsent coding pass clusters for user k at the beginning of the transmission interval l . Define the distortion-to-length slope for a coding pass cluster with the rate increment $\Delta r_{t,b,k}$ and the distortion reduction $\Delta d_{t,b,k}$ as

$$\lambda_{t,b,k} \triangleq |\Delta d_{t,b,k}| / \Delta r_{t,b,k}. \quad (4.11)$$

The distortion-to-length slope represents how much distortion a coding pass cluster can reduce by given one unit of rate. We can sort all distortion-to-length slopes of all unsent coding pass clusters ($\lambda_{t,b,k}$, where $T_k^{b,l-1} \leq t$) in a decreasing

order and obtain the corresponding mapping indices $I_k(t, b)$ and inverse indices $I_k^{-1}(m) \in \{(t, b)\}$. For example, if the sorting result is $\lambda_{t_1, b_1, k} > \lambda_{t_2, b_2, k} > \dots$, we assign $I_k(t_1, b_1) = 1$, $I_k(t_2, b_2) = 2$, and $I_k^{-1}(1) = (t_1, b_1)$, $I_k^{-1}(2) = (t_2, b_2)$. Then, a decreasing discrete R-D function $(R_k^l[m], D_k^l[m])$ for quality layer l can be obtained according to this sorting result, as shown in Table 4.2 (c). To facilitate the distortion-to-rate searching, we relax the constraints on integer value of rate and integer number of coding pass clusters to allow them to be real numbers; and construct a continuous R-D function through linear interpolation of the discrete R-D function as follows:

$$D_k^l(\gamma_k^l) = \frac{D_k^l[m+1] - D_k^l[m]}{R_k^l[m+1] - R_k^l[m]}(\gamma_k^l - R_k^l[m]) + D_k^l[m], \quad (4.12)$$

$$\text{for } R_k^l[m] \leq \gamma_k^l \leq R_k^l[m+1] \text{ and } m = 0, \dots, M_k^l - 1,$$

where γ_k^l is the required bit rate. We can calculate the least required rate, γ_k^l , to achieve the targeted distortion, D , by finding the inverse function of $D_k^l(\cdot)$. We summarize the algorithm used in this stage in Table 4.2. The complexity of this stage for each user is $O(M_k^l \log(M_k^l))$ due to sorting.

Stage 2

At this stage, the goal is to minimize the maximal distortion supported by the OFDMA subsystem through a bi-section search procedure. By checking the continuous GOF R-D functions obtained in (4.12), the resource allocator can calculate the minimum transmission rates $\{\gamma_k^l\}$ necessary to achieve the same targeted distortion among all users. Then the resource allocator checks whether these requested rates can be supported by the OFDMA subsystem under current channel conditions. If the requested rates are feasible, the resource allocator tries to further decrease the targeted distortion by increasing the requested rates. Otherwise, the

Table 4.3: OFDMA resource allocation and feasibility check

a) <i>Initialization:</i> Get $\Delta R_k^l, \forall k$ and set $\mathbf{A} = 0, \mathbf{R} = 0$, and $\mathbf{P} = 0$.
b) <i>Minimal Rate Assignment:</i> While $\sum_{n=0}^{N-1} r_{kn} a_{kn} \geq \Delta R_k^l, \forall k$ not satisfied <ol style="list-style-type: none"> 1) Find $\hat{k}, \hat{n} = \arg \max_{k,n} [\mathbf{G}]_{kn}$. 2) Assign subcarrier \hat{n} to user \hat{k}. Set $G_{k\hat{n}} = 0, \forall k$. Waterfill all subcarriers of user \hat{k} to minimize power with rate $\Delta R_{\hat{k}}^l$. 3) If the requested rate of user \hat{k} is achieved, set $G_{\hat{k}\hat{n}} = 0, \forall n$. 4) If no subcarriers left and not all requested rates are satisfied, report infeasibility and exit.
c) <i>Power Reduction:</i> While there are subcarriers left <ol style="list-style-type: none"> 1) Assign user k who has the highest average power per subcarrier with a remaining subcarrier having the largest G_{kn} for user k. 2) Minimize the transmission power among the subcarrier set assigned to this user. 3) Calculate the overall power. If not greater than P_{max}, calculate \mathbf{A}, \mathbf{R}, and \mathbf{P}. Exit and report feasibility. Calculate the overall power. If greater than P_{max} , report infeasibility.

resource allocator increases the targeted distortion to reduce the requested rates and checks the feasibility again. A bi-section search algorithm is deployed to find the minimal distortion D that the OFDMA subsystem can support.

The feasibility of the requested rates depends on two factors. First, the OFDMA subsystem should be able to transmit the requested rates $\{\gamma_k^l\}$ for all users. Second, the overall transmission power, P_{sum} , cannot exceed P_{max} . We develop a fast suboptimal algorithm shown in Table 4.3 to allocate the bits and power to satisfy the rate constraint first and then the power constraint. There are three steps in the proposed algorithm for feasibility checking: initialization, minimal rate assignment, and power reduction. First, the subcarrier assignment matrix \mathbf{A} , the rate assignment matrix \mathbf{R} , and the power assignment matrix \mathbf{P} are initialized to zeros.

Next, the system tries to satisfy the requested rates. In each round, we allocate an unassigned subcarrier to a user. If $G_{\hat{k}\hat{n}}$ has the maximal value in current \mathbf{G} , subcarrier \hat{n} is assigned to user \hat{k} and we update $G_{k\hat{n}} = 0, \forall k$ to prevent this subcarrier from being assigned again. We then determine the modulation schemes and the coding rates for all subcarriers currently allocated to user \hat{k} such that the requested data rate can be accommodated and the required transmission power is minimized in the meantime. This can be implemented by the well-known water-filling algorithm with Table 4.1 and (4.9). If the requested rate of user \hat{k} can be allocated, user \hat{k} is removed from future assignment list in this step by assigning $G_{\hat{k}n} = 0, \forall n$. This step is repeated until all users' requested rates are satisfied. If all subcarriers are already assigned and not all requested rates can be allowed, infeasibility is reported and the resource allocator has to reduce the requested rates. In the third step, we try to reduce the overall transmission power P_{sum} to be below P_{max} by assigning the remaining subcarriers. In each round, we select a user who has the highest average power per subcarrier and assign him/her with one of the remaining subcarriers in which this user has the largest channel gain. Then we minimize the transmission power among the subcarrier set assigned to this user. The overall transmission power is calculated and if it is greater than P_{max} , the power reduction procedure is repeated again. Otherwise, we calculate \mathbf{A} , \mathbf{R} , and \mathbf{P} for OFDMA subsystem, report feasibility and exit. An infeasibility is reported if there is no subcarrier left and $P_{sum} > P_{max}$. Since the required power for each user in each subcarrier can be pre-calculated, the complexity of checking feasibility in each iteration is $O(N)$. The overall number of iterations, which is typically fewer than 20 in our experiment, is bounded by the bi-section search.

Table 4.4: Coding pass cluster assignment

<p>a) Find $\hat{m}_k = \arg \max\{R_k^l[m] \leq AR_k^l, \forall m\}$. Allocate coding pass clusters with indices $I_k(t, b) \leq \hat{m}_k$.</p>
<p>b) Calculate unused bandwidth $UR_k^l = AR_k^l - R_k^l[\hat{m}_k]$. While $UR_k^l \geq 0$</p> <ol style="list-style-type: none"> 1) Search the coding pass clusters set, S, whose element satisfies $\Delta r_{t,b,k} \leq UR_k^l$ and $t = T_k^{b,l}$ in all subbands. 2) If set S is empty, leave the loop. 3) Select the coding pass cluster, \hat{b}, with largest $\lambda_{t,\hat{b},k}$ in set S. 4) Update $T_k^{\hat{b},l} = T_k^{\hat{b},l} + 1$ and $UR_k^l = UR_k^l - \Delta r_{t,\hat{b},k}$.

Stage 3

At this stage, we perform the coding pass cluster assignment for each user individually. Denote the assigned rate from Stage 2 for the k^{th} user as $AR_k^l = \sum_{n=0}^{N-1} a_{kn} r_{kn}$. Due to the discrete rate provided by the OFDMA subsystem ($r_{kn} \in \nu, \forall k, n$), the assigned transmission rate AR_k^l is generally larger than the requested rate, i.e., $AR_k^l \geq \gamma_k^l$. Therefore, we have rate budget AR_k^l to allocate the coding pass clusters. We formulate the problem as minimizing the distortion subject to the rate constraint by allocating the coding pass cluster:

$$\min_{\Delta \mathbf{T}_k^l} D_k^l \text{ subject to } \Delta R_k^l \leq AR_k^l. \quad (4.13)$$

Since for each unsent coding pass cluster we need to decide whether we select it or not, the problem (4.13) is a binary knapsack problem [60], which is *NP* hard. To ensure the real-time performance, we apply a two-step greedy algorithm here. First, among all values of $R_k^l[m]$ that are not larger than AR_k^l , we find the largest one, $R_k^l[\hat{m}_k]$. We will include all coding pass clusters whose indices $I_k(t, b)$ are not larger than \hat{m}_k in the current quality layer. Notice that the sorting order $\{I_k(t, b)\}$ has ensured the decoding dependency of coding pass clusters in each subband.

This is because if user k receives the t^{th} coding pass cluster in subband b , user k must receive coding pass cluster 0 to $t-1$ since $\lambda_{t',b,k} > \lambda_{t,b,k}, \forall t' < t$ or $I_k(t', b) < I_k(t, b), \forall t' < t$ due to the convexity of R-D in subband b . Second, a round of refinement is performed to utilize the unused bandwidth, $UR_k^l = AR_k^l - R_k^l[\hat{m}_k]$. We search all unselected coding pass clusters that follow the currently selected truncation points and pick those with rates not larger than the unused bandwidth. The coding pass cluster with the largest distortion-to-length slope is selected for transmission during current transmission interval. The system updates the unused bandwidth and unselected coding pass clusters; and then repeats the above search procedure until there is no coding pass cluster with size not larger than the unused bandwidth. Since the first step directly uses the result in (4.12), the refinement step consumes most computation power in the whole coding pass cluster assignment and the complexity to search a feasible coding pass cluster is $O(B)$. We recap the algorithm used in this stage in Table 4.4.

4.4 Optimization in Resource Allocator: Focusing on Efficiency

In this section, we study the efficiency problem in which the overall distortion of all users is minimized in a transmission interval. We first formulate the efficiency problem as an optimization problem. Then, similar to the fairness case, we also propose a three-stage algorithm to determine the subcarrier assignment, rate assignment, and coding pass cluster assignment to achieve the optimization goal.

4.4.1 Formulation of the Efficiency Problem

We formulate the efficiency problem as to minimize the overall (weighted) end-to-end distortion among all users subject to constraints on subcarrier assignment, subcarrier rate, user rate, and power:

$$\min_{\mathbf{A}, \mathbf{R}, \Delta \mathbf{T}^l} \sum_{k=0}^{K-1} w_k \cdot f(D_k^l) \quad (4.14)$$

subject to

$$\left\{ \begin{array}{l} \text{Subcarrier Assignment: } \sum_{k=0}^{K-1} a_{kn} \leq 1, a_{kn} \in \{0, 1\}, \forall n; \\ \text{Subcarrier Rate: } r_{kn} \in \nu, \forall k, n; \\ \text{User Rate: } 0 \leq \Delta R_k^l \leq \sum_{n=0}^{N-1} a_{kn} r_{kn}, \forall k; \\ \text{Power: } \sum_{k=0}^{K-1} \sum_{n=0}^{N-1} a_{kn} P_{kn} \leq P_{max}. \end{array} \right.$$

The constraints are similar to the fairness case presented in Chapter 4.3.1. Similar to the fairness case, the delay constraint is implicitly imposed in the problem (4.14) so as the transmission delay is restricted within a transmission interval.

4.4.2 Proposed Algorithm for Efficiency

To solve this minimization problem, we propose a three-stage algorithm shown in Figure 4.3. The first stage is to obtain the continuous R-D functions of all unsent coding pass clusters for the current GOF. The second stage is to perform subcarrier assignment and rate assignment through a 2-D waterfilling procedure; and the third stage is the coding pass cluster assignment. The first and third stage are the same as what have been discussed in Chapter 4.3.2. Here, we focus on the second stage and consider the case of $w_k = 1, \forall k$ and $f(D_k^l) = D_k^l$.

Having the continuous R-D functions, the problem (4.14) can be simplified as follows:

$$\min_{\mathbf{A}, \mathbf{R}} \sum_{k=0}^{K-1} D_k^l(\gamma_k^l) \quad (4.15)$$

$$\text{subject to } \begin{cases} \text{Subcarrier Assignment: } \sum_{k=0}^{K-1} a_{kn} \leq 1, a_{kn} \in \{0, 1\}, \forall n; \\ \text{Subcarrier Rate: } r_{kn} \in \nu, \forall k, n; \\ \text{Power: } \sum_{k=0}^{K-1} \sum_{n=0}^{N-1} a_{kn} P_{kn} \leq P_{max}; \end{cases}$$

where $\gamma_k^l = \sum_{n=0}^{N-1} a_{kn} r_{kn}$. To solve this problem, a two-step suboptimal algorithm is proposed by first determining the subcarrier assignment matrix \mathbf{A} and then deciding the rate assignment matrix \mathbf{R} .

Subcarrier Assignment

In this step, we relax the power constraint by assuming the maximal transmission power is unlimited and thus each subcarrier can be loaded with maximum rate ν_Q to fully utilize the available bandwidth. Then, the problem (4.15) has only the subcarrier assignment constraint and the goal is to find the subcarrier-to-user assignment that can reduce most distortion by using the least amount of power. This problem can be solved by an iterative greedy algorithm. In each iteration, we evaluate which user can achieve the most distortion reduction by using the least power if we assign an unused subcarrier to him/her. There are two factors affecting this evaluation, as reflected by ψ_{kn} and defined below:

$$\psi_{kn} \triangleq \left(\frac{D_k^l(\gamma_k^l) - D_k^l(\gamma_k^l + \nu_Q)}{\nu_Q} \right) \left(\frac{\nu_Q}{\frac{\rho_Q \sigma^2}{G_{kn}}} \right). \quad (4.16)$$

γ_k^l is the accumulative allocated rate for user k in the current iteration. The first term of (4.16) evaluates the gradient of reduced video distortion with respect to the allocated rate, i.e., how much distortion we can reduce by assigning a unit of rate for user k . The second term of (4.16) evaluates the gradient of the allocated rate to the required transmission power (calculated using (4.9)), i.e., how many bits this system can transmit at $\text{BER} \leq 10^{-6}$ per unit of power. If both factors of

user k at subcarrier n are large, it implies that assigning subcarrier n to user k can use the same amount of power to reduce more distortion. Since the second term is only a function of channel gain, we can further simplify (4.16) and have a matrix Ψ as $[\Psi]_{kn} = \psi_{kn}$:

$$\psi_{kn} = (D_k^l(\gamma_k^l) - D_k^l(\gamma_k^l + \nu_Q))G_{kn}. \quad (4.17)$$

Once we obtain Ψ , we can find its entry (\hat{k}, \hat{n}) with maximal value and assign subcarrier \hat{n} to user \hat{k} . To prevent this subcarrier from being assigned again, we set $G_{k\hat{n}} = 0, \forall k$. Then, we update the current allocated rate of user \hat{k} by setting $\gamma_{\hat{k}}^l = \gamma_{\hat{k}}^l + \nu_Q$ and subcarrier assignment matrix by setting $a_{\hat{k}\hat{n}} = 1$ and $a_{k'\hat{n}} = 0$ for $k' \neq \hat{k}$. This procedure is repeated until all subcarriers are assigned. The complexity of this step is $O(N)$.

Rate Assignment

Based on the subcarrier assignment in the previous step, we determine how much rate should be assigned to each subcarrier. To facilitate our discussion, let $\theta_{kn} \in \{0, 1, \dots, Q\}$ be the subcarrier usage index, which indicates the selected row in Table 4.1 for user k at subcarrier n . For example, $\theta_{kn} = q$ represents that user k has loaded $r_{kn} = \nu_q$ bits in subcarrier n and the required SNR to achieve $\text{BER} \leq 10^{-6}$ is ρ_q . Further, we define a set of incremental rate $\Delta\nu = \{\Delta\nu_1, \Delta\nu_2, \dots, \Delta\nu_Q\}$ and a set of incremental power $\Delta\rho = \{\Delta\rho_1, \Delta\rho_2, \dots, \Delta\rho_Q\}$, where $\Delta\nu_q = \nu_q - \nu_{q-1}$ and $\Delta\rho_q = \rho_q - \rho_{q-1}$ for $q = 1, \dots, Q$, respectively. We solve this rate assignment problem using a 2-D discrete waterfilling algorithm. At the beginning, we set all required rate $\{\gamma_k^l\}$ and all subcarrier usage indices $\{\theta_{kn}\}$ to zeros. In each iteration, similar to Step 1, we select the subcarrier setting that can achieve the most distortion reduction by using the least power when we evaluate the results of

filling each subcarrier an incremental rate $\Delta\nu_{\theta_{kn}+1}$. This procedure is repeated until all subcarriers are fully loaded or the overall required power reaches the maximal available amount. The evaluation of distortion-to-power ratio for all subcarriers and users can be quantified as a matrix Φ with $[\Phi]_{kn} = \phi_{kn}$:

$$\phi_{kn} \triangleq \left(\frac{D_k^l(\gamma_k^l) - D_k^l(\gamma_k^l + \Delta\nu_{\theta_{kn}+1})}{\Delta\nu_{\theta_{kn}+1}} \right) \left(\frac{\Delta\nu_{\theta_{kn}+1}}{\Delta\rho_{\theta_{kn}+1}\sigma^2/G_{kn}} \right). \quad (4.18)$$

The first term of (4.18) represents how much distortion user k can reduce with an extra unit of rate and the second term of (4.18) represents how many bits to transmit for user k at subcarrier n with a unit of power. The overall ϕ_{kn} represents how much distortion user k will reduce at subcarrier n with one extra unit of power.

After obtaining Φ , we select the entry (\hat{k}, \hat{n}) with largest value. If subcarrier \hat{n} does not belong to user \hat{k} , i.e., $a_{\hat{k}\hat{n}} = 0$, we set $\phi_{\hat{k}\hat{n}} = 0$ and search the highest value in Φ again. If so, we update user \hat{k} 's subcarrier usage index

$$\theta_{\hat{k}\hat{n}} = \theta_{\hat{k}\hat{n}} + 1, \quad (4.19)$$

the overall transmission rate of user \hat{k}

$$\gamma_{\hat{k}}^l = \gamma_{\hat{k}}^l + \Delta\nu_{\theta_{\hat{k}\hat{n}}}, \quad (4.20)$$

and the overall transmission power

$$P_{sum} = P_{sum} + \frac{\Delta\rho_{\theta_{\hat{k}\hat{n}}}\sigma^2}{G_{\hat{k}\hat{n}}}. \quad (4.21)$$

If subcarrier \hat{n} is overloaded, (i.e., $r_{\hat{k}\hat{n}} > \nu_Q$), or the overall required power exceeds the P_{max} , we need to pick other entry by setting $\phi_{\hat{k}\hat{n}} = 0$ and search the highest value in Φ until a valid one is found. The search algorithm terminates if no more valid assignment is found. The whole algorithm is presented in Table 4.5. Since the accumulative rate $\gamma_{\hat{k}}^l$ is updated for the selected user \hat{k} only, the complexity

Table 4.5: Proposed algorithm to minimize overall distortion

<p>a) <i>Initialization</i>: $P_{sum} = 0$, $Exit = False$.</p>
<p>b) <i>Subcarrier Assignment</i>:</p> <p>Set $\gamma_k^l = 0$, $\forall k$.</p> <p>While not all subcarriers are assigned</p> <p> Calculate Ψ using (4.17).</p> <p> Find $\hat{k}, \hat{n} = \arg \max_{k,n} [\Psi]_{kn}$.</p> <p> Set $a_{\hat{k}\hat{n}} = 1$ and $a_{k',\hat{n}} = 0$ for $k' \neq \hat{k}$. Set $G_{k,\hat{n}} = 0 \forall k$.</p> <p> Update user's rate $\gamma_{\hat{k}}^l = \gamma_{\hat{k}}^l + \nu_Q$.</p>
<p>c) <i>Rate Assignment</i>:</p> <p>Set $\gamma_k = 0$, $\forall k$; and set $Exit$ as <i>False</i>.</p> <p>While $Exit == False$</p> <p> Calculate Φ using (4.18).</p> <p> Set $Found$ as <i>False</i>.</p> <p> While $Found == False$</p> <p> Find $\hat{k}, \hat{n} = \arg \max_{k,n} [\Phi]_{kn}$.</p> <p> If $\phi_{\hat{k}\hat{n}} == 0$, set $Exit$ as <i>True</i> and $Found$ as <i>True</i>.</p> <p> Else If $a_{\hat{k}\hat{n}} == 0$, set $\phi_{\hat{k}\hat{n}} == 0$.</p> <p> Else</p> <p> If $r_{\hat{k}\hat{n}} + \Delta\nu\theta_{\hat{k}\hat{n}} + 1 \leq \nu_Q$ and $P_{sum} + \frac{\Delta\rho_{\theta_{\hat{k}\hat{n}}+1}\sigma^2}{G_{\hat{k}\hat{n}}} \leq P_{max}$</p> <p> Set $Found$ as <i>True</i>,</p> <p> Update system using (4.19)(4.20)(4.21).</p> <p> Else, set $\phi_{\hat{k}\hat{n}} = 0$.</p>

of updating Φ in each iteration is $O(n_{\hat{k}})$, where $n_{\hat{k}}$ is the number of subcarriers assigned to user \hat{k} . The maximal number of iteration in the rate assignment is NQ ; and the actual number of iteration depends on the transmission power level and channel condition.

4.5 Network-Level Service Objectives

In this section, we discuss the extended functionalities based on the proposed algorithms in Chapter 4.3 and 4.4. We first address how to achieve a desired tradeoff between system fairness and efficiency. Then we investigate how to incorporate unequal error protection in the proposed framework to increase system's efficiency.

4.5.1 Tradeoff between Fairness and Efficiency

We have proposed two solutions to ensure fairness and improve efficiency in each transmission interval, respectively. If we apply fairness algorithm in all transmission intervals (L intervals), the received video qualities for all users will be similar to each other. However, the users whose video programs require more rates to achieve the same video quality or who are in bad channel conditions will become a bottleneck in the whole system and degrade the overall video qualities. If we apply efficiency algorithm in all transmission intervals, the system will achieve the highest overall video qualities. Nevertheless, the users in good channel conditions with low video content complexity will be assigned more system resources. Consequently, some users will have unnecessarily good video qualities while others will have very bad qualities. In other words, a system achieving more efficiency will suffer from more unfairness. We are interested in how to design a system with

partial fairness and partial efficiency. To achieve this tradeoff, for each GOF, we propose to apply fairness algorithm in the first several transmission intervals (x intervals) to ensure the baseline fairness, and then apply the efficiency algorithm in the rest transmission intervals ($y = L - x$ intervals) to improve the overall video qualities. We denote this strategy as F_xE_y algorithm. Note that F_LE_0 algorithm is the pure fairness algorithm and F_0E_L algorithm is the pure efficiency algorithm.

4.5.2 Unequal Error Protection

It has been shown that the unequal error protection (UEP) can improve the expected video qualities [2, 61, 83]. Relaxing the requirement from lower targeted BER to higher targeted BER but sending the same bit rate, the required power can be reduced. In other words, if the overall transmission power is fixed, the overall bit rate using higher targeted BER can be higher than the one with lower targeted BER. It is potential to improve the overall expected video qualities. The UEP takes the advantage of different priorities within a video bit stream by using different targeted BER. For the video bit stream with higher priority, the UEP adopts stronger error protection (lower targeted BER) to increase the probability of successful transmission. For the video bit stream with lower priority, the UEP applies weaker error protection (higher targeted BER) to utilize a larger effective bandwidth for statistical performance gain.

Because the EWW bit stream exhibits a strong decoding dependency, all received coding pass clusters in a subband should be adjacent to each other and also a truncated version of the original bit stream starting from MSB. Assuming all bits in the quality layer $0 \sim l - 1$ are received correctly, given a targeted BER_i ,

the expected distortion of the quality layer l can be represented as:

$$E_i[D_k^l] = D_k^{l-1} - \sum_{b=0}^{B-1} \sum_{t=T_k^{b,l-1}}^{T_k^{b,l}-1} p_{t,b,k}^i \Delta d_{t,b,k}. \quad (4.22)$$

Here $p_{t,b,k}^i$ is the probability that the receiver can correctly receive all coding pass clusters from $T_k^{b,l-1}$ to t in subband b and can be expressed as follows:

$$p_{t,b,k}^i = (1 - \text{BER}_i)^{\Delta R_{t,b,k}}. \quad (4.23)$$

Here $\Delta R_{t,b,k}$ is the cumulative number of bits from coding pass cluster $T_k^{b,l-1}$ to t in subband b and can be expressed as

$$\Delta R_{t,b,k} = \sum_{t'=T_k^{b,l-1}}^t \Delta r_{t',b,k}. \quad (4.24)$$

Quality layer l has higher priority than quality layer k if $l < k$ since both layers may have coding passes in the same subband so that coding passes in quality layer k have decoding dependency on the ones in quality layer l due to (4.4). We incorporate the unequal error protection strategy in the proposed framework by considering the priorities of quality layers in different transmission intervals. In the first $L - 1$ transmission intervals, we solve the original problem (4.15) using the proposed algorithm shown in Table 4.5 with the strongest error protection. At the last transmission interval, we solve the problem (4.15) but replacing D_k^l with $E_i[D_k^l]$ using several different BER_i as shown in Table 4.1 ($\text{BER}_0 = 10^{-6}$ and $\text{BER}_1 = 10^{-5}$ in our case). We pick the BER setting that achieves the lowest overall expected distortion.

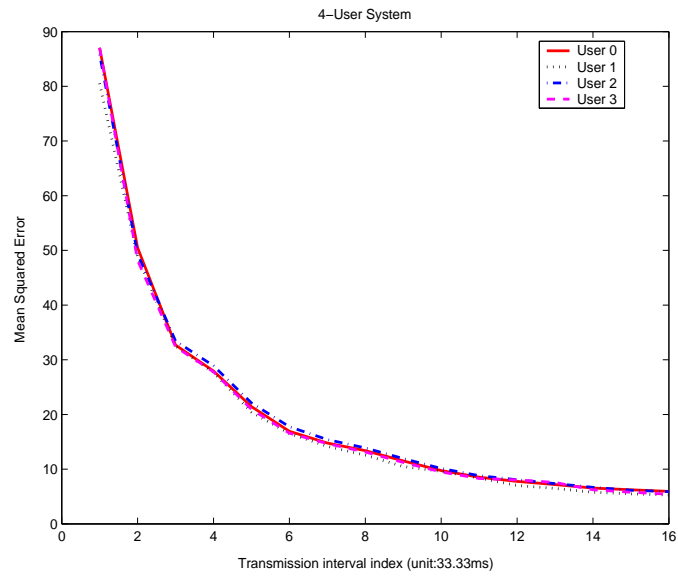
4.6 Simulation Results

The simulations are set up as follows. The OFDMA system has 32 subcarriers over a total 1.6MHz bandwidth. The delay spread in root mean square is 3×10^{-7} s.

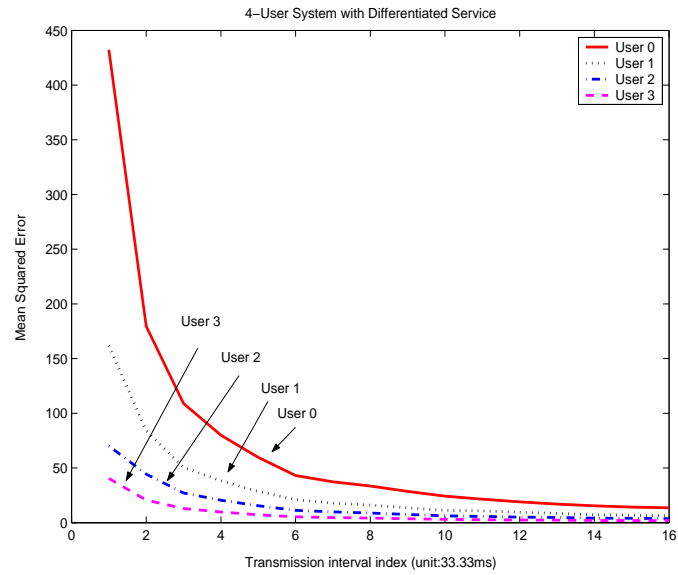
An additional $5\mu s$ guard interval is used to avoid inter-symbol interference due to channel delay spread. This results in a total block length as $25\mu s$ and a block rate as 40K per second. The Doppler frequency is 10Hz and the transmission interval is $33.33ms$. The mobile is uniformly distributed within the cell with radius of 50m and the minimal distance from mobile to the base station is 10m. The noise power is 5×10^{-9} Watts, and the maximal transmission power is 0.1 Watts. The propagation loss factor is 3 [75]. The video refreshing rate is 30 frames per second with CIF resolution (352x288). The GOF size is 16 frames and each GOF is encoded by the codec [35] using Daubechies 9/7 bi-orthogonal filter with 4-level temporal decomposition and 3/2/1 spatial decomposition in low/mid/high temporal subbands, respectively.

4.6.1 Performance of the Fairness Algorithm

We first demonstrate how the proposed algorithm $F_{16}E_0$ achieves pure fairness among all users when all users request uniform quality. We consider a four-user system, where each user receives 10 GOFs from one of the four video sequences, *Foreman*, *Hall Monitor*, *Mother and daughter*, and *Silent*, respectively. Figure 4.4(a) shows the received video quality of the first GOF in terms of mean-squared error in every transmission interval. As we can see, all four users have similar video quality in each transmission interval and the received video quality is improved by receiving more quality layers till the last transmission interval. We also show the corresponding subcarrier assignment of the first GOF in each transmission interval in Figure 4.5(a). As the source coding rate of each user is allocated in different time and frequency slots according to the channel conditions and source characteristics, the diversities of frequency, time, and multiuser are jointly exploited. Figure 4.6(a)

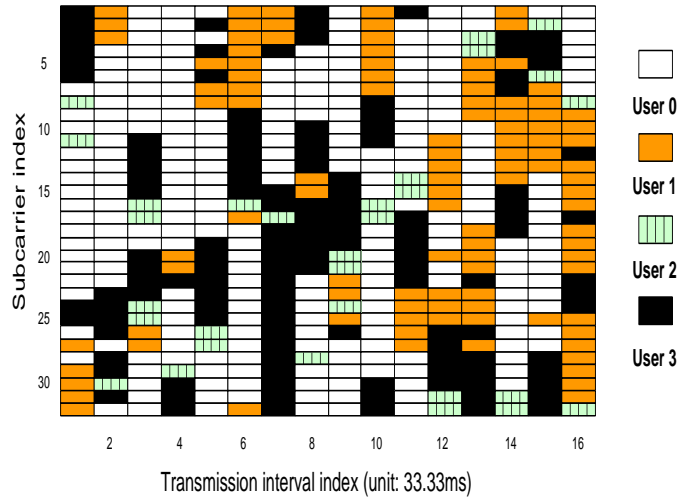


(a) $F_{16}E_0$ system

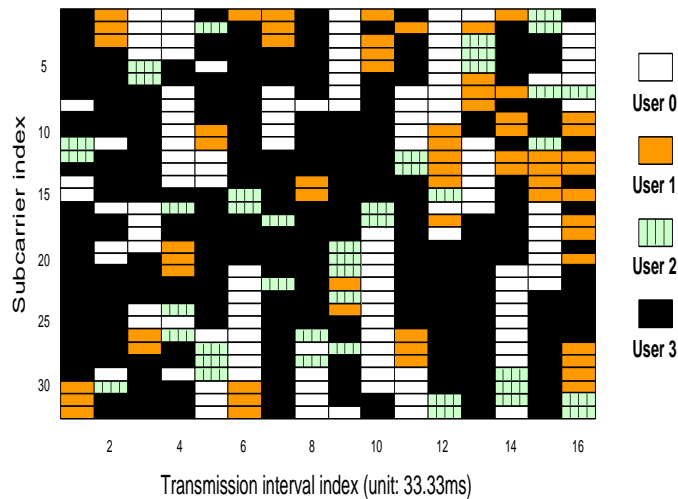


(b) $F_{16}E_0$ system with differentiated service

Figure 4.4: Comparison for the $F_{16}E_0$ system providing uniform quality and differentiated service.

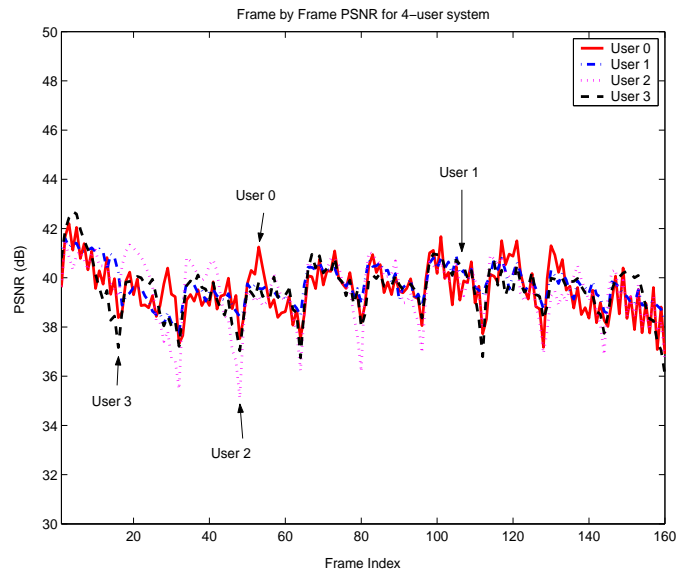


(a) $F_{16}E_0$ system

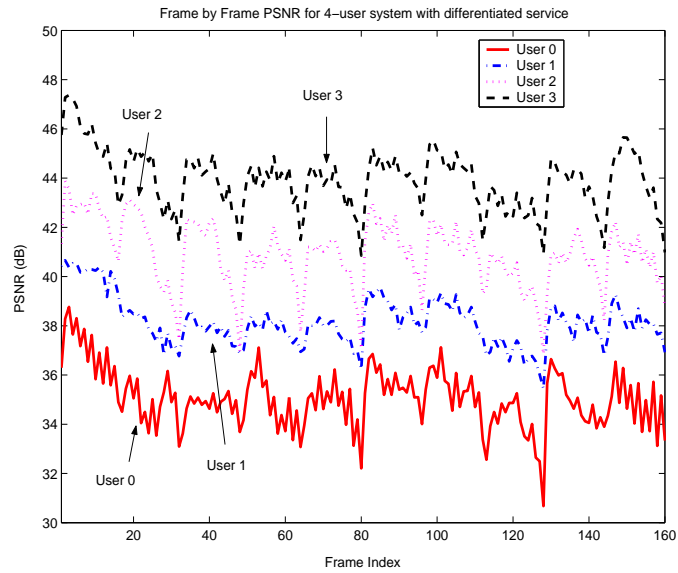


(b) $F_{16}E_0$ system with differentiated service

Figure 4.5: Subcarrier assignment for the $F_{16}E_0$ system in each transmission interval (a) Uniform quality. The system assigns more subcarriers to user 0 at most intervals due to the required rate of video sequence 0 to achieve the same quality is higher than other sequences. (b) Differentiated service. The system assigns more subcarriers to user 3 due to the highest requested quality.



(a) $F_{16}E_0$ system



(b) $F_{16}E_0$ system with differentiated service

Figure 4.6: Frame-by-frame PSNR for the $F_{16}E_0$ system with uniform quality and with differentiated service.

shows the frame-by-frame PSNR along 10 GOFs for all users. The average PSNR along the received 160 frames for each user is 39.52, 39.71, 39.46, 39.54dB, respectively. The deviation of users' received quality is small and within 0.25dB. Thus, the pure fairness algorithm, $F_{16}E_0$, can provide similar visual qualities among all users during the whole transmission time.

As we have mentioned that the proposed framework can provide differentiated service by appropriately setting the quality weighting factor $\{w_k\}$ in (4.10). We repeat the above experiment with a new set $\{w_k\}$ as $w_0 = 0.25, w_1 = 0.5, w_2 = 1$, and $w_3 = 2$. The PSNR difference between user i and $i + 1$ is expected to be 3dB. Figure 4.4(b) shows the mean-squared error of the first GOF received by each user in every transmission interval. As we can see, the video qualities received by all users maintain the desired quality gap in every transmission interval. The differentiated service is achieved when we receive all quality layers. Figure 4.5(b) shows the corresponding subcarrier assignment of the first GOF in each transmission interval. Compared to Figure 4.5(a), User 3 occupies more subcarriers in the system with differentiated service than in the system with uniform quality. Figure 4.6(b) shows the frame-by-frame PSNR along 10 GOFs for all users. As we have expected, User 3 has the highest received video quality and User 0 has the lowest PSNR. The average PSNR received by each user is 35.06, 38.16, 40.91, 43.92dB, respectively. The PSNR differences between user i and $i + 1$ for $i = 0, 1, 2$ are 3.10, 2.75, 3.01dB, respectively, which is close to the design goal of 3dB differentiated service.

4.6.2 Performance of the $F_x E_y$ Algorithm Family

Next, we evaluate the proposed algorithm $F_x E_y$ with different values of x . Here we also compare the proposed algorithm with the TDM algorithm². Instead of allowing subcarriers in a transmission interval to be allocated among multiple users, the TDM algorithm assigns all subcarriers in one transmission interval to only one user whose current end-to-end distortion is the largest. Thus, the multiuser and frequency diversity is not explored in this TDM algorithm.

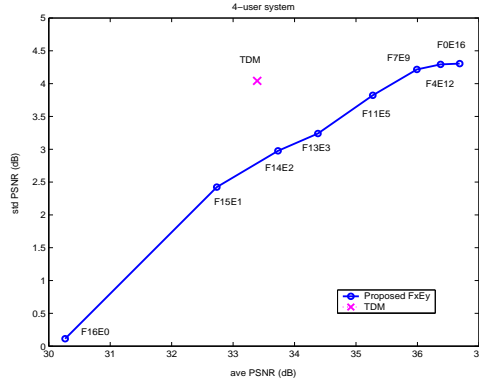
We concatenate 15 classic CIF video sequences to form one testing video sequence of 4064 frames. The 15 sequences are 288-frame *Akiyo*, 144-frame *Bus*, 288-frame *Coastguard*, 288-frame *Container*, 240-frame *Flower*, 288-frame *Foreman*, 288-frame *Hall Monitor*, 288-frame *Highway*, 288-frame *Mobile*, 288-frame *Mother and daughter*, 288-frame *MPEG4 news*, 288-frame *Paris*, 288-frame *Silent*, 256-frame *Tempete*, and 256-frame *Waterfall*. The video for the k^{th} user is 160 frames long and from frame $256 \times k + 1$ to frame $256 \times k + 160$ of the testing sequence.

Two performance criteria are used to measure the proposed algorithm and TDM algorithm. We first calculate the average received video quality of all 160 frames for each user and denote it as $PSNR_k$ for the k^{th} user. To measure the efficiency, we average the $PSNR_k$ for all users, i.e.

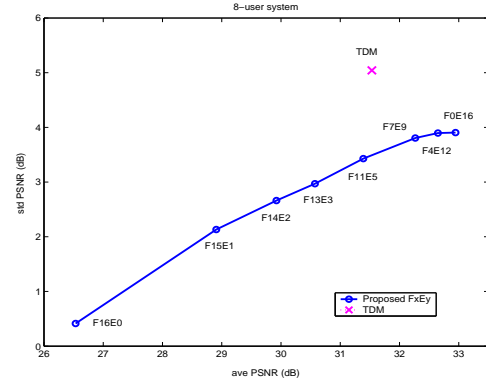
$$\text{avePSNR} = \frac{1}{K} \sum_{k=0}^{K-1} PSNR_k. \quad (4.25)$$

The higher avePSNR is, the higher system efficiency of overall video quality we have. To measure the fairness, we take the standard deviation for each user's

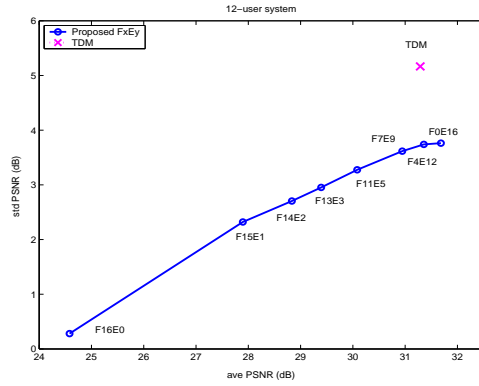
²As discussed in Chapter 2.2, the current IEEE 802.11 medium access control (MAC) protocol supports two kinds of access methods: distributed coordination function (DCF) and point coordination function (PCF). In both mechanisms, only one user occupies all the bandwidth at each time, which is similar to TDM technology.



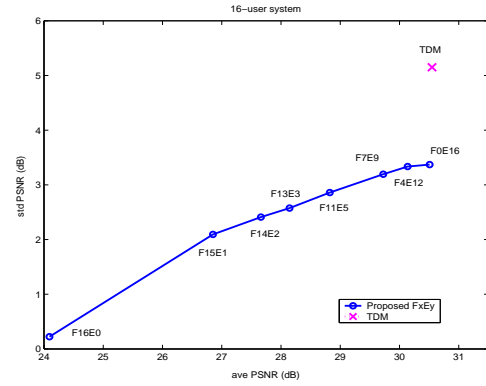
(a) 4-user system



(b) 8-user system



(c) 12-user system



(d) 16-user system

Figure 4.7: avePSNR and stdPSNR results of the $F_x E_y$ algorithm family and TDM algorithm.

average received video quality, i.e.

$$\text{stdPSNR} = \sqrt{\frac{1}{K} \sum_{k=0}^{K-1} (\text{PSNR}_k - \text{avePSNR})^2}. \quad (4.26)$$

The lower stdPSNR is, the fairer quality each user receives. If stdPSNR is high, it implies some users receive video programs with high quality and the other users receive video programs with poor quality.

Figure 4.7 shows the fairness and efficiency results for the proposed algorithms and TDM algorithm. We first compare the performances for eight settings of the

F_xE_y algorithm family, including F_0E_{16} , F_4E_{12} , F_7E_9 , $F_{11}E_5$, $F_{13}E_3$, $F_{14}E_2$, $F_{15}E_1$, and $F_{16}E_0$. We see that the pure fairness algorithm $F_{16}E_0$ achieves the lowest PSNR deviation among all algorithms but has the lowest average PSNR; and the pure efficiency algorithm F_0E_{16} achieves the highest average PSNR but has the highest PSNR deviation. The tradeoff between avePSNR and stdPSNR can be adjusted by selecting different number of transmission intervals for fairness algorithm. As revealed from Figure 4.7, the F_xE_y algorithm has higher average received video quality but higher quality deviation than the $F_{x-1}E_{y+1}$ algorithm. The second comparison included in Figure 4.7 is between the F_xE_y algorithm family and the TDM algorithm. As shown in Figure 4.7, for achieving the same avePSNR, the proposed F_xE_y algorithm family has about 1~1.8 dB lower deviation in PSNR than the TDM algorithm. In other words, the proposed algorithm provides fairer quality than the TDM algorithm. This is because the proposed scheme employs additional diversity in frequency and multiuser.

To evaluate the received video quality along the time axis, we show the frame-by-frame PSNR using TDM algorithm and the F_xE_y algorithm family for each user in a four-user system in Figure 4.8. We choose three algorithms from F_xE_y algorithm family, namely, the pure efficiency algorithm (F_0E_{16}), the pure fairness algorithm ($F_{16}E_0$), and one example of the partial fairness-efficiency algorithms ($F_{14}E_2$). As we can see, the F_0E_{16} algorithm has higher PSNR than $F_{14}E_2$, $F_{16}E_0$ for all users except User 1. This is due to two factors: one is that the video content of User 1 requires higher rate to achieve the same video quality than the other users and the other reason is that the channel condition for User 1 is the worst among all users. Therefore, the F_0E_{16} algorithm assigns more rates to the other users than User 1 to achieve higher average received video quality of all users.

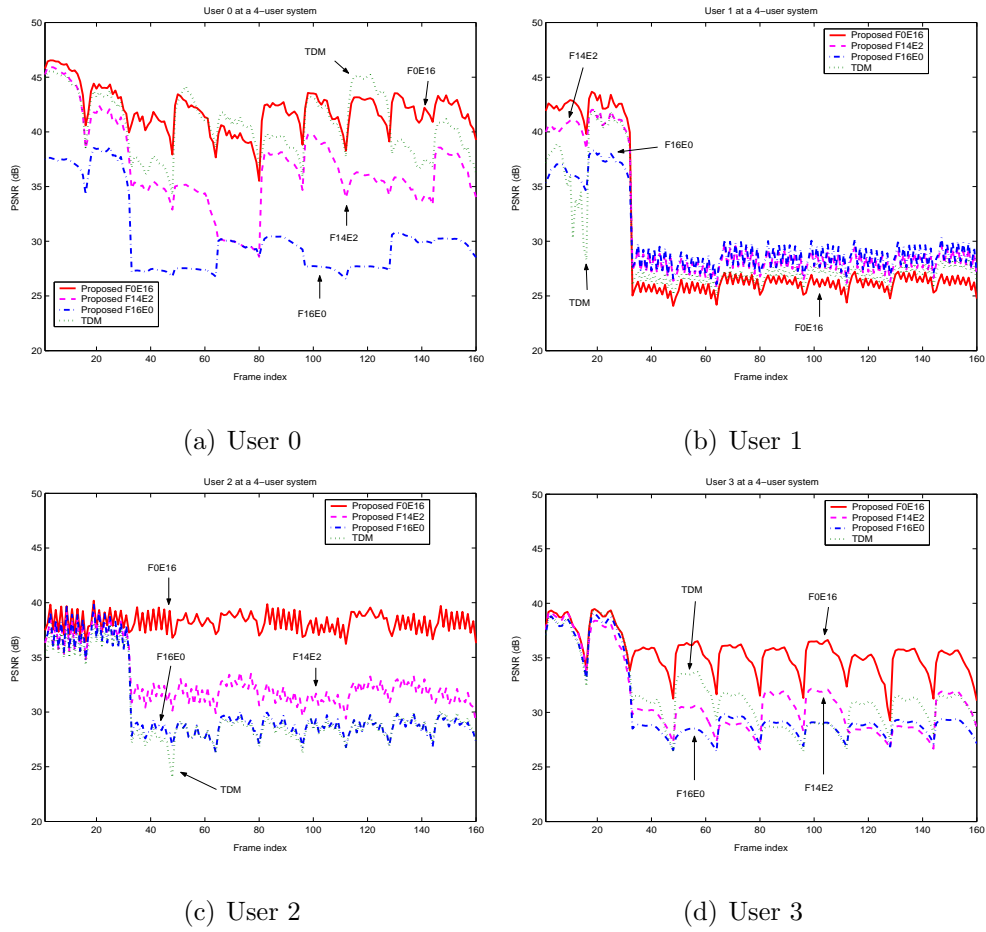


Figure 4.8: Frame-by-frame PSNR for different algorithms of a 4-user system.

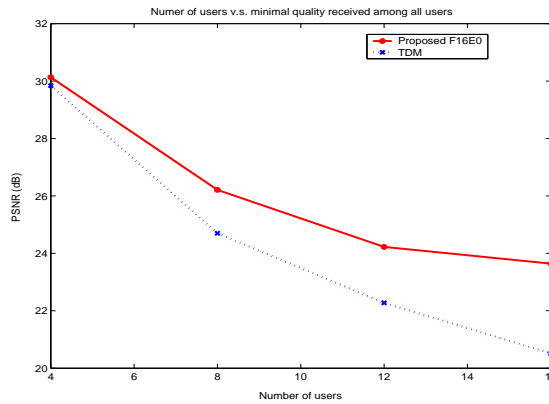


Figure 4.9: Performance comparison for the worst quality received among all users using the proposed algorithm and TDM algorithm.

Figure 4.9 shows the average value of the worst received PSNR among all users from 10 different terminals' locations for different number of users in the system. We can see that the proposed algorithm, $F_{16}E_0$, can improve the minimal PSNR better than that of the TDM algorithm. There is about 0.5~3dB gain for different number of users. The performance gap increases when the number of users increases owing to the multiuser diversity.

In Table 4.6, we show the performance gain that the unequal error protection scheme outperforms the equal error protection (EEP) for different numbers of users in the system using F_0E_{16} algorithm. For the UEP strategy, the targeted BER of the first 15 transmission intervals is set to 10^{-6} . The targeted BER of the last transmission interval is chosen from $\{10^{-5}, 10^{-6}\}$, depending on which BER setting achieving better expected distortion using (4.22). For the EEP, the targeted BER for all transmission intervals is 10^{-6} . The video content and channel setting are the same as before. For each setting, we run the simulation 10,000 times. As revealed in Table 4.6, the UEP can improve the expected average PSNR per user only about 0.05~0.13dB.

This small improvement using the UEP is due to several reasons. First, although a system with higher targeted BER has potential to attain higher bit rate throughput, the distortion introduced by the channel becomes significant. Thus, the increased effective bandwidth is limited, which limits the reduction of video distortion. Second, the EWV codec has a high compression ratio at very low bit rate but its R-D curve becomes flatter at higher bit rate due to the distortion-to-length slope sorting. So, for a system that has already received a large amount of video data at the last transmission interval, the improved distortion due to the increased effective bandwidth using the UEP is limited. Third, the joint multiple video cod-

Table 4.6: Unequal error protection versus using equal error protection

Number of users	4	8	12	16
average PSNR gain per user (dB)	0.13	0.09	0.07	0.05

ing has explored the video content complexities for all videos and the system has assigned more system resources to users who are in good channel conditions with simple video content complexity to achieve the highest system efficiency. Thus, the extra bit rate budget benefited from the UEP will be distributed to users who are in bad channel conditions with complex video complexity, whose overall distortion can only be improved by a limited amount. Further, the selection of targeted BER is based on the expected distortion calculated from (4.22). If the targeted BER is selected as 10^{-6} , the UEP is equivalent to the EEP and no performance gain can be obtained. We also observe that the more users the system has, the smaller performance improvement we have. It is because the increased bandwidth due to higher targeted BER is roughly a constant and is shared by all users. When the number of user increases, the increased bandwidth assigned to each user will reduce and the quality improvement will reduce.

4.7 Video over Interference-limited Networks

In the previous sections, we have considered wireless networks where allocating these resources to one user would affect other users due to the limited amount of resources. In some other wireless networks, such as CDMA, the performance of each individual user is further affected by the interference of simultaneous usage from other users. In this section, we extend the principle of resource allocation

discussed in the previous sections to the scenario that sending multiple real-time encoded video programs to multiple mobile users over downlink Multicode CDMA systems [31, 90].

4.7.1 System Description

To facilitate the rate adaptation, we adopt MPEG-4 FGS, which has been discussed in Chapter 3.2. The video codec has a *rate constraint* that the transmitted rate for each video frame should be between the base-layer rate and the maximal available FGS rate plus the base-layer rate.

Multicode CDMA (MC-CDMA) system [36, 37] provides a digital bandwidth-on-demand platform by allocating multiple codes according to users' rate requests. MC-CDMA system has a *code constraint* that a code can be assigned to at most one user. We need to determine the *code assignment*, namely, which code should be assigned to which user. The system has a *power constraint* that the overall transmission power for all C codes should not be larger than the maximal transmission power, P_{max} , which is to limit the co-interference between cells and to operate within the working range of communication circuits.

To protect bitstreams from bit error during transmission, we use rate-compatible punctured convolutional code (RCPC) [27], which provides a wide range of channel coding rates within $[T_{min}, T_{max}]$. The goal of channel coding is to provide sufficiently low BER on the bitstream level such that the end-to-end video quality is controllable. For MPEG-4, the degradation of video quality can be kept negligible if we enable the error resilient features and error concealment mechanism as well as keep the BER below a threshold set around 10^{-6} [90]. To achieve the BER requirement, the received SINR should not be less than a targeted SINR that can be

approximated by an exponential function of the channel coding rate of RCPC [90].

Although a CDMA code with higher channel coding rate can carry more source bits, the required power to meet the BER requirement is higher. The received SINR for each code is subject to interference from other codes, because of non-orthogonality among codes caused by multipath fading [6]. Since the overall power is limited, we need to determine the *channel coding rate assignment* of each code to achieve the optimal video quality subject to the power constraint and interference.

Overall, the key issue is how to jointly perform rate adaptation, code allocation, and power control to achieve the required perceptual qualities of received video. We consider system efficiency during each video frame refreshing interval by determining the code assignment and channel coding rate assignment, subject to the constraints on CDMA codes, power, and video rate. This problem is a mixed integer programming problem, which is *NP* hard. In searching for an effectively real-time solution, we have found an important heuristic: Since distortions can be reduced by using extra either power or codes, the code and power resource should be used in a balanced way to avoid exhausting one resource first while having the other resource left, leading to low system performances.

4.7.2 Distortion Management Algorithm

To manage the distortion, a balanced code and power usage algorithm (BCP) is developed. An example shown in Figure 4.10 illustrates how the algorithm works. We first allocate the resources for delivering the base-layer data to provide the baseline video quality for each user, as shown in Position A. Then, we allocate resources for FGS layer by keeping a guideline to the ratio of the current power to the number of assigned CDMA code. We assign one code at a time to a user and

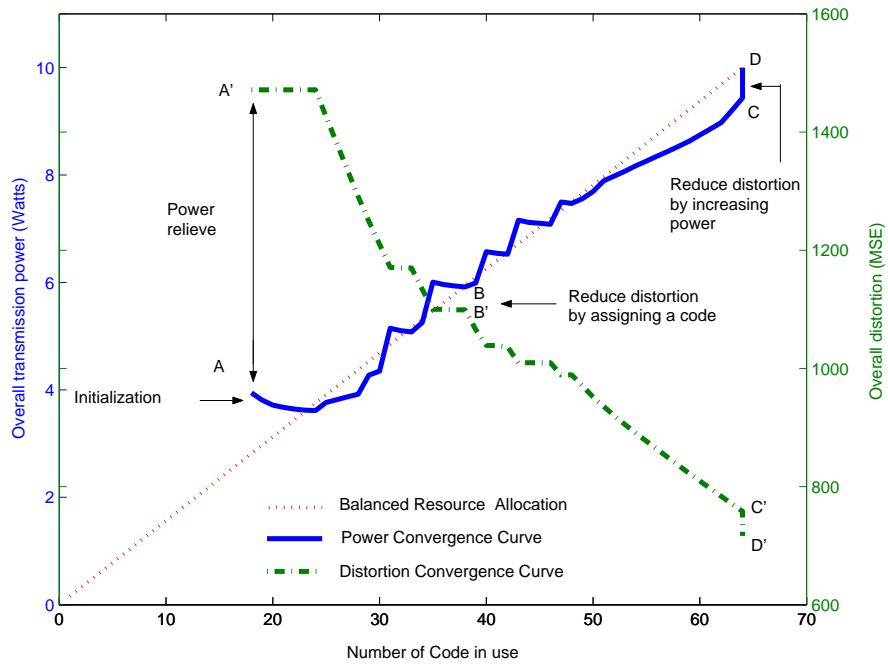


Figure 4.10: Power (solid line) and distortion (dot-dashed line) convergence track vs. the number of assigned codes. The maximal power is 10W and the total number of codes is 64. Position A, B, C, and D are examples indicating the overall power when different numbers of codes are assigned and Position A', B', C', and D' are the corresponding overall distortion.

perform different algorithms according to the current power-to-code usage ratio. If the current ratio is larger than the ratio of the maximal power over the total number of CDMA codes (as shown by the dotted line), the system consumes higher than average power per code. Position A is one of the examples. In this case, a new code is assigned to a user by keeping user's source coding rate unchanged but redistributing the source rates among user's already assigned codes plus the new code. Subsequently, the channel coding rates for those codes, the required SINR, and the overall power are all reduced. If the current power-to-code ratio is smaller than the ratio of the maximal transmission power over the total number of CDMA codes, the system consumes moderate power per code. One of such example is Position B. In this case, we will assign a new code with maximal channel coding rate to carry video source bits such that the overall distortion is reduced. By doing so, the total power consumption would increase. After all codes are assigned, we perform a round of refinement to further reduce distortion by using the remaining power quota. Position C to Position D is an example for quality refinement. At the end, all available power and code resources are fully utilized.

Owing to the mixed integer programming nature of the problem, it is difficult to evaluate how close to the optimal solutions the proposed algorithm performs. However, we have found it is possible to derive optimal solutions for code-limited case and performance bounds for power-limited case with linear complexity.

The code-limited case refers to the situation when the transmission power can be viewed as unbounded, while there are only a limited number of pseudo-random codes. This happens when all users are close to the base station, so that the necessary transmission power is much less than the maximal transmission power. We also assume the number of users is large enough or the requested rates for video

transmission are large enough, so that all C codes are used. In this case, in order to have the highest distortion reduction, each code should carry as much information as possible. So all channel coding rate of each code should be equal to the maximal channel coding rate, T_{max} . The optimal solution with linear computation complexity can be obtained as follows: we first divide each video stream into several segments with equal length T_{max} ; then we calculate the distortion reduction of each segment if the corresponding mobile user receives it; we sort all distortion reduction for all segments from all bitstreams in a decreasing order and choose the first C segments with the largest distortion reduction; finally we assign codes to the corresponding users for transmitting the selected bit stream segments.

The power-limited case refers to the situation that all available power is used and there might still be some codes left. This happens when all users are far away from the base station. We also assume the number of users is small or the requested rates are not large, so that it is not limited by power and code simultaneously. Under this condition, the channel coding rate T_{min} for all codes will have the minimal overall transmission power. So if the power is limited, by using the minimal channel coding rate, we can have the highest source rates and corresponding minimal distortions. We further relax the code constraint, namely, a code is shared by multiple users; and assume the orthogonality among all codes holds. We can derive a performance bound with linear computation complexity as the following procedures: we first divide each video stream into several segments with equal length T_{min} ; then we calculate the distortion reduction of each segment and the power increment if we select this segment and assign it a code; we evaluate the ratio between the distortion reduction and power increment for each segment; we sort all ratios of all segments from all bitstreams in a decreasing order and

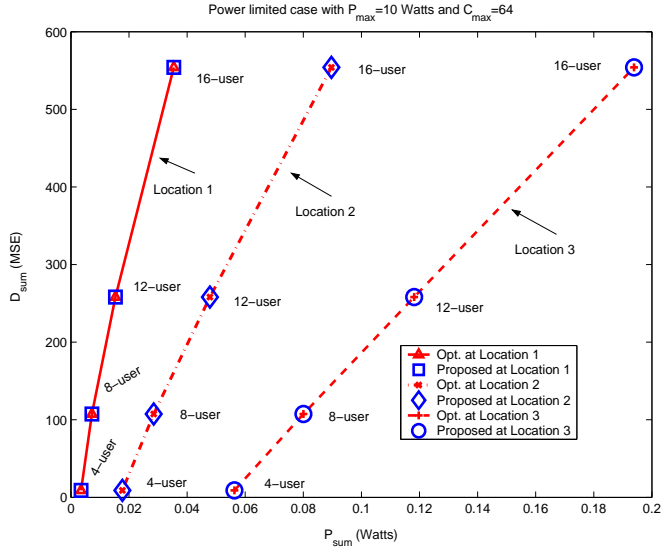


Figure 4.11: Code limited case: optimal solutions

choose the first C segments with the largest ratio; finally we assign codes to the corresponding users for transmitting the selected bit stream segments.

4.7.3 Simulation Results

We concatenate 15 classic QCIF (176×144) video sequences to form a basic testing sequence of 2775 frames [90]. The content program for each user is 100 frames and starts from a randomly selected frame of the testing sequence. The video refreshing rate is 15 frames per second (fps). To study the proposed algorithm performance under code-limited (power-unlimited) case we compared its total distortion with that of the optimal solution of the code-limited case for different number of users, N , and different locations. We set all mobile users at locations 1, 2, and 3, at distances near 100m, 150m, and 200m, respectively. From the results shown in Figure 4.11, we can see that the proposed scheme always achieves the optimal solution.

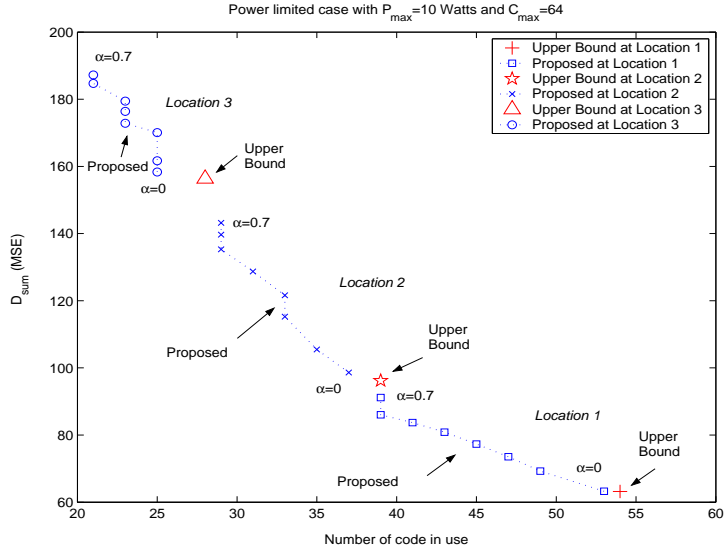


Figure 4.12: Power limited case: close to performance upper bounds

To study the proposed algorithm performance under power-limited (code unlimited) case, we considered four users which are located at different locations. All mobile users in location 1,2 and 3, are near 1100 m, 1200 m, and 1300 m, respectively. Figure 4.12 shows the results comparing the proposed algorithm with orthogonality factor between 0 and 0.7 and the performance upper bound for the power-limited case. As we can see, the performance of the proposed algorithm with small orthogonality factor is close to the performance upper bound that assumes the orthogonality factor to be zero. The loss is because two reasons. First, the performance upper bound is obtained by allowing a non-integer number of codes, so the bound has a better performance than the optimal solution. Second, the proposed algorithm might reach local minima instead of the global minima. The above two simulation results demonstrate the effectiveness of the proposed algorithm in both special cases.

For the case where both power and code are constrained, we compare the

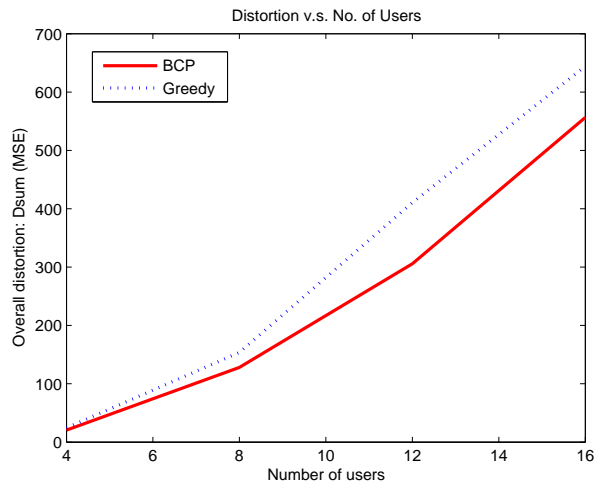


Figure 4.13: Performance comparison of BCP and greedy algorithms

proposed algorithm with a modified greedy algorithm [30]. For each iteration, the greedy algorithm tries to assign a candidate code with maximal channel coding rate to every user, calculates the distortion reduction, and assigns a new code to the user with the largest value. The location for each user is uniformly distributed within the cell with radius from 20m to 1000m. Figure 4.13 shows the comparison result for different number of users vs. the total distortion, D_{sum} . The simulation results demonstrate that BCP algorithm outperforms the greedy algorithm 14% ~ 26%. The main reason for this gain is that the greedy algorithm ignores the balance between power and code usages and thus depletes one resource while wasting other resources.

4.8 Chapter Summary

In this chapter, we have constructed a framework sending multiple scalable video programs over downlink wireless networks. We first study video over OFDMA

system. By leveraging the frequency, time, and multiuser diversity of the OFDMA system and the scalability of the 3-D embedded wavelet video codec, the proposed framework can allocate system resources to each video stream to achieve the desired video quality. Two service objectives are addressed: fairness and efficiency. For fairness problem, we formulate the system to achieve fair quality among all users as a min-max problem. For efficiency problem, we formulate the system to attain the lowest overall video distortion as a minimization problem. To satisfy the real-time requirement, two fast algorithms are proposed to solve the above two problems.

The simulation results have demonstrated that the proposed fairness algorithm outperforms TDM algorithm by about 0.5~3dB for the worst received video quality criterion. The proposed $F_x E_y$ algorithm family can achieve a desired tradeoff between fairness among users and overall system efficiency. At the same average video quality among all users, the proposed algorithm has about 1~1.8dB lower PSNR deviation among all users than the TDM algorithm. So, the proposed algorithm can provide better system efficiency and stricter fairness. In addition, the proposed fairness algorithm can allow differentiated service by appropriately setting values for quality weighting factors. We also extend the proposed framework to incorporate unequal error protection. In summary, the proposed framework is a promising solution for broadband multiuser video communication.

We have also applied the resource allocation strategy to MC-CDMA networks and demonstrated the effectiveness of the proposed principle. We have further considered two special cases which have optimal solutions and performance bounds with linear complexity and demonstrated the performance by utilizing the proposed principle of resource allocations can achieve or close to the optimal solutions.

Chapter 5

End-to-end Video Conferencing

5.1 Introduction

In Chapter 4, we have discussed how to transmit multiple video programs over downlink wireless networks with a single cell. Video-on-demand service is one of the applications. There are many other applications which may transmit video bit-streams in both uplinks and downlinks within a single cell and/or multiple cells. One of the promising services is interactive video conferencing, whereby a pair of mobile users at different locations can exchange video streams with each other in real time. Besides the real time requirement, a wireless system providing interactive video conferencing faces more challenges than the typical video-on-demand service. For instance, in each conversation session, there are two video streams being exchanged between a conversation pair; each video stream is transmitted through at least two paths, namely, an uplink to an access point and a downlink from the access point. The transmitted packets of each video stream experience different channel conditions in both links. Because the radio bandwidth resources are limited for different users' transmissions over uplink and downlink and the

channel conditions change over time, dynamically allocating the limited network resources to all users can significantly improve the end-to-end quality. Moreover, various video programs exhibit different content complexities and require different amount of bandwidth to achieve similar video quality. To provide satisfactory video quality to all users, a multiuser wireless video conferencing system should integrate cross-layer design methodology and dynamic multiuser resource allocation. In this chapter, we address the above issues and propose an interactive video conferencing framework to support multiple conversation pairs over WLANs [91, 94].

For a wireless system with limited bandwidth resources, it is critical to determine the amount of bandwidth allocated to uplink and downlink to achieve high spectrum utilization and system service objectives. A static strategy is to allocate equal bandwidth to both links and perform optimal uplink resource allocation and optimal downlink resource allocation individually. As this simple strategy of allocating equal bandwidth to both links is inefficient due to uneven load, several works using unequal bandwidth assignment have been proposed. A scheme was proposed in [43] to address the unbalanced capacity and asymmetric channel bandwidth usage problem. Several call admission control schemes were presented in [41, 42, 122] to explore the asymmetric traffic load in both links. A scheduler to control generic data traffic in both uplink and downlink simultaneously for IEEE 802.11a networks was proposed in [112]. Bandwidth resource allocation for transmitting video over WLAN in real time is more challenging than for transmitting generic data since compressed video bitstreams exhibit different characteristics from generic data as discussed in Chapter 1.2. This motivates us to investigate dynamical bandwidth allocation for both links of all video streams.

In general, the channel conditions along the whole end-to-end transmission

path of a video stream are heterogeneous. To achieve the same bit error rate, the required level of error protection in each path may not be the same. The FEC transcoding strategy, which applies optimal level of FEC in each intermediate path, can provide higher effective end-to-end throughput than the traditionally end-to-end fixed FEC strategy. Furthermore, adopting FEC transcoding in each intermediate transmission node can recover certain amount of corrupted packets transmitted through the preceding paths, thus preventing from further quality degradation accumulatively in following the transmission paths. With fixed allocated bandwidth and prior knowledge of the channel condition for each path, systems can be formulated as to maximize the overall throughput by determining which intermediate nodes should perform FEC transcoding for the unicast scenario [81] and for the multicast scenario [73].

In this chapter, we propose a framework which explores the diversity of video content and the heterogeneity of uplink and downlink channel conditions experienced by different users [92]. With vertical integration of different communication layers, a cross-layer unequal error protection mechanism is proposed for graceful quality degradations. In the proposed framework, the bandwidth of uplink and downlink is dynamically allocated according to the needs. The system performance can be further improved by jointly choosing the optimal channel coding rate and the bandwidth in both uplink and downlink, and performing FEC transcoding in the server located at the access point.

5.2 System Description

In this section, we present an overview of our proposed video conferencing system, as shown in Figure 5.1. We first analyze the throughput and error protection

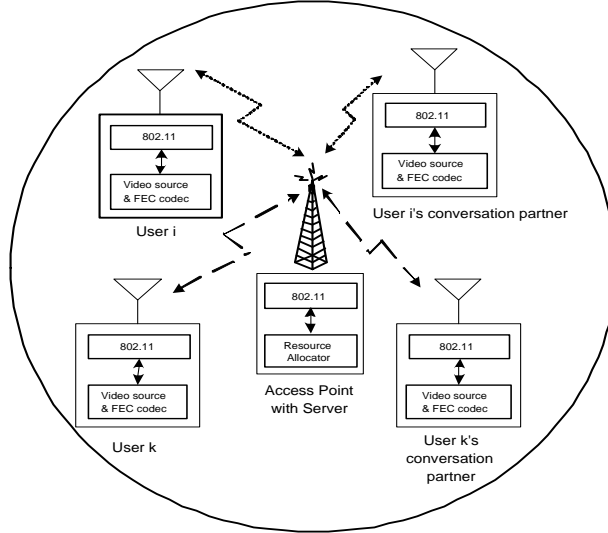


Figure 5.1: System block diagram for single-cell case.

provided by the communication subsystem. We then discuss the cross-layer error protection schemes and the required signalling in the proposed distortion control framework.

5.2.1 IEEE 802.11a

We use the IEEE 802.11a Physical layer as an example to present the proposed framework. Other wireless LAN standards can be incorporated in a similar way. The IEEE 802.11a Physical (PHY) layer provides eight PHY modes with different modulation schemes and different convolutional coding rates, and can offer various data rates. The configurations of these eight PHY modes are listed in Table 5.1.

A major task of the proposed system is to select the optimal PHY modes for each uplink and downlink of each user. Let $P_{max}^{(U)}$ and $P_{max}^{(D)}$ be the maximal available transmission power for uplink and downlink, respectively; and $G_i^{(U)}$ and $G_i^{(D)}$ the uplink and downlink channel gain from user i to his/her conversation partner at

Table 5.1: Physical layer mode for 802.11a

Mode	Modulation	Channel Coding	Data Rate
1	BPSK	1/2	6 Mbps
2	BPSK	3/4	9 Mbps
3	QPSK	1/2	12 Mbps
4	QPSK	3/4	18 Mbps
5	16-QAM	1/2	24 Mbps
6	16-QAM	3/4	36 Mbps
7	64-QAM	2/3	48 Mbps
8	64-QAM	3/4	54 Mbps

the current time slot. Without loss of generality, we assume the same thermal noise level, σ^2 , for all users. Thus, the maximal SNR for uplink and downlink are

$$\Gamma_i^{(U)} = \frac{P_{max}^{(U)} G_i^{(U)}}{\sigma^2} \text{ and } \Gamma_i^{(D)} = \frac{P_{max}^{(D)} G_i^{(D)}}{\sigma^2}, \text{ respectively.} \quad (5.1)$$

The BERs of BPSK, QPSK, 16-QAM, and 64-QAM modulation are given by the following equations as functions of the received symbol SNR denoted by Γ [32]:

$$P_b^{\text{BPSK}}(\Gamma) = 0.5 \left(1 - \sqrt{\frac{\Gamma}{1+\Gamma}} \right), \quad (5.2)$$

$$P_b^{\text{QPSK}}(\Gamma) = 0.5 \left(1 - \sqrt{\frac{\Gamma}{2+\Gamma}} \right), \quad (5.3)$$

$$P_b^{\text{16QAM}}(\Gamma) = 0.5 \left[\left(1 - \sqrt{\frac{\Gamma}{10+\Gamma}} \right) + \left(1 - \sqrt{\frac{9\Gamma}{10+9\Gamma}} \right) \right], \quad (5.4)$$

and

$$P_b^{\text{64QAM}}(\Gamma) = \frac{1}{24} \left(14 - 7\sqrt{\frac{\Gamma}{42+\Gamma}} - 6\sqrt{\frac{9\Gamma}{42+9\Gamma}} + \sqrt{\frac{25\Gamma}{42+25\Gamma}} - 2\sqrt{\frac{81\Gamma}{42+81\Gamma}} - \sqrt{\frac{121\Gamma}{42+121\Gamma}} + \sqrt{\frac{169\Gamma}{42+169\Gamma}} \right). \quad (5.5)$$

With convolutional code, the union bound for BER [70] can be expressed as

$$P_c(\Gamma) \leq \sum_{d=d_{free}}^{\infty} a_d P_d(\Gamma), \quad (5.6)$$

where d_{free} is the free distance of the convolutional code, a_d is the total number of error events of weight d , and $P_d(\Gamma)$ is the probability that an incorrect path at distance d from the correct path is chosen by the Viterbi decoder. When the hard decision is applied, $P_d(\Gamma)$ can be given by:

$$P_d(\Gamma) = \begin{cases} \sum_{k=(d+1)/2}^d \binom{d}{k} (P_b)^k (1 - P_b)^{d-k}, & \text{when } d \text{ is odd;} \\ \frac{1}{2} \binom{d}{d/2} (P_b)^{d/2} (1 - P_b)^{d/2} + \sum_{k=d/2+1}^d \binom{d}{k} (P_b)^k (1 - P_b)^{d-k}, & \text{when } d \text{ is even,} \end{cases} \quad (5.7)$$

where P_b is the uncoded BER depending on the modulations from (5.2) to (5.5).

If user i selects the uplink and downlink PHY mode as m_i and n_i , respectively, the BER for uplink and downlink can be approximated as a function of PHY mode and SNR level:

$$\text{BER}_{i,m_i}^{(U)} = P_{m_i}(\Gamma_i^{(U)}) \text{ and } \text{BER}_{i,n_i}^{(D)} = P_{n_i}(\Gamma_i^{(D)}), \text{ respectively,} \quad (5.8)$$

where the function $P_{m_i}(\cdot)$ and $P_{n_i}(\cdot)$ are the union bound of BER using channel coding as defined in (5.6). Since different PHY modes use different modulation schemes and channel coding rates, their coded BER performances are different. At the same SNR, systems with higher PHY mode index can provide higher throughput at a cost of higher BER than ones with lower PHY mode index.

The probability that a packet is received successfully for uplink and downlink can be calculated as

$$p_{i,m_i}^{(U)} = (1 - \text{BER}_{i,m_i}^{(U)})^L \text{ and } p_{i,n_i}^{(D)} = (1 - \text{BER}_{i,n_i}^{(D)})^L, \text{ respectively,} \quad (5.9)$$

where L is the number of bits in a packet. With a fixed packet size, $p_{i,m_i}^{(U)}$ and $p_{i,n_i}^{(D)}$ are functions of the channel gains and PHY modes.

As discussed in Chapter 2.2, the IEEE 802.11 MAC protocol supports two kinds of access methods, namely, the distributed coordination function (DCF) and the

point coordination function (PCF). The DCF is an access mechanism using carrier sense multiple access with collision avoidance (CSMA/CA). In contrast, the PCF is based on polling controlled by a point coordinator. In both mechanisms, only one user occupies all the bandwidth at each time slot. The proportion of time a user can occupy the bandwidth can be controlled by either PCF or enhanced DCF [67, 115]. In this work, we study how to determine the proportion of time allocated to each user to optimize users' video quality.

5.2.2 Application Layer FEC

In the IEEE 802.11 MAC protocol, a packet sent from uplink will be dropped if errors are detected and will not be forwarded to the next path or the upper communication layer. A packet loss in the base layer will cause error propagation for the video data that are predictively encoded using that frame as reference. In addition, FGS layer bitstream has strong decoding dependency owing to the intra-bitplane variable length entropy coding and the inter-bitplane DCT coefficient synchronization. The loss of a FGS layer packet containing significant bitplanes will make the following successfully received FGS layer packets containing lower bitplanes useless. Since we can know which packet arrives successfully at the application layer by checking the transmission index in the packet header, this wireless channel can be modelled as a packet erasure channel [52, 58, 105]. Applying FEC in the application layer across packets, such as systematic Reed-Solomon (RS) codes, has been shown as an effective way to alleviate the problem caused by packet loss [105]. An $RS(K_i, k)$ encoder will generate $K_i - k$ parity symbols for k source symbols, and a corresponding RS decoder can recover the original source symbols if it receives at least k out of K_i symbols successfully when the locations of the erased

symbols are known. We can apply RS coding across source packets to generate parity packets for recovering erasure packet loss.

Since MPEG-4 FGS codec is a two-layer scheme, we adopt different strategies for each layer, as shown in Figure 5.2. For the non-scalable base layer, we apply a strong equal error protection strategy across packets to provide the baseline video quality. To remove the strong decoding dependency of the FGS layer bit-stream and to have graceful quality fluctuation, we adopt the multiple-description forward error correction framework (MD-FEC) [71]. MD-FEC converts a prioritized bit stream into non-prioritized and packetized bit streams. Each packetized bit stream represents one description which can be independently decoded to represent the content in a coarse quality, and the final reconstructed video quality depends primarily on how many packets the receiver receives successfully, instead of depending on which packets are corrupted. The more descriptions a receiver receives successfully, the better reconstructed quality the decoder can get. The basic mechanism of MD-FEC works as follows: Let s be the number of symbols carried in a packet and K_i be the total number of packets. A segment is defined as the symbols located at the same position over the K_i packets. The FGS bit stream is converted to these K_i packets segment by segment, and an RS coding across packet is applied within each segment to provide error protection. An RS code with higher level of error protection is applied to the segment with higher priority. Figure 5.2 shows the overall error protection strategy. If the receiver receives ρ packets successfully out of K_i packets, then the segments encoded with $RS(K_i, k)$ codes for $k \leq \rho$ can be correctly decoded. The optimal configuration of RS code in each segment can be formulated as a constrained optimization problem and solved through the Lagrangian method [71]. There have been several works proposed to

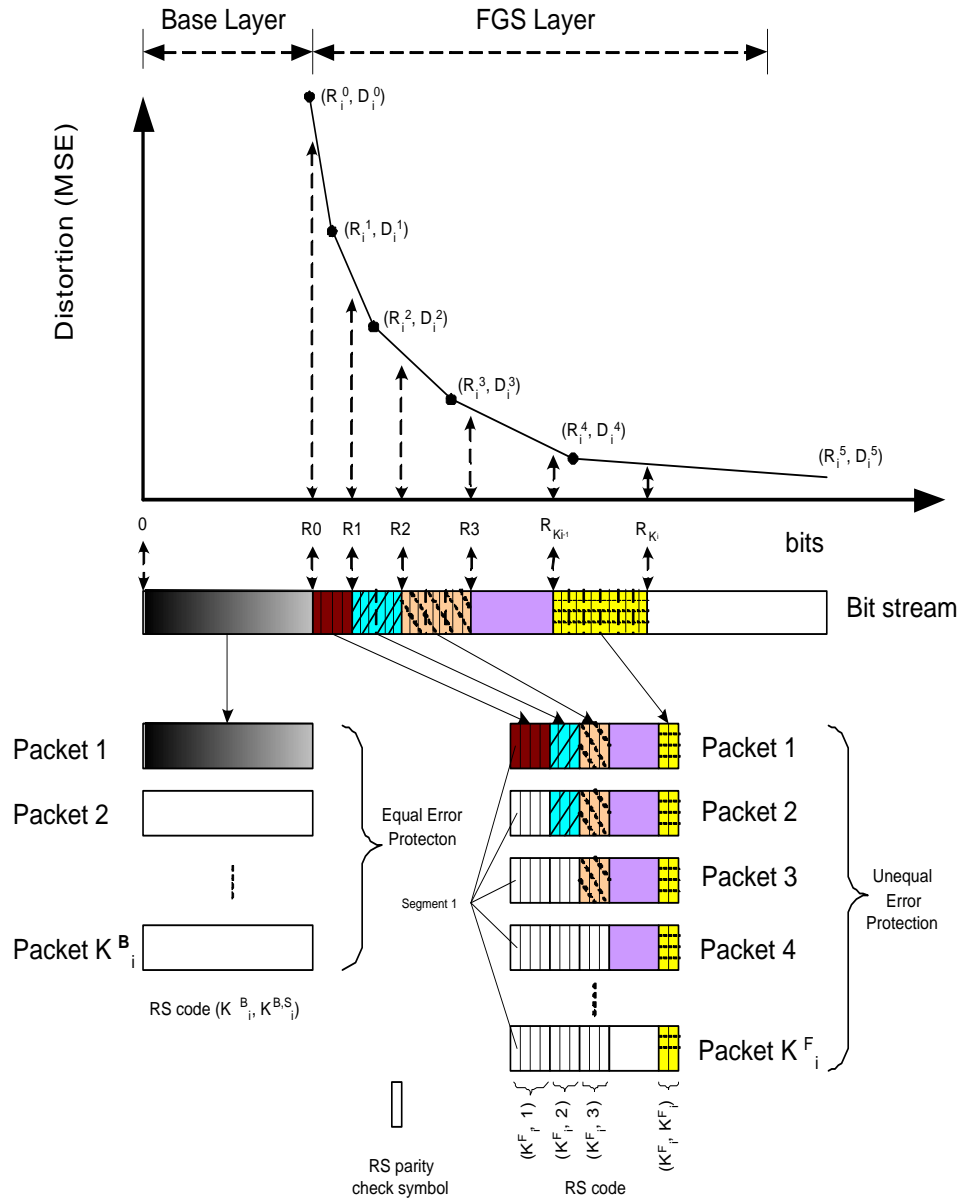


Figure 5.2: Error protection scheme for application layer FEC

reduce the computational complexity of MD-FEC. We adopt the fast local search method [87] in this work.

To decode the coded video packets using the MD-FEC framework, the RS decoder located at each client terminal needs to know the configuration of RS code used in each segment. The RS configuration is generated through an optimization according to the side information, namely, the R-D of video source, packet loss rate due to the selected PHY modes, and the allocated numbers of transmitted packets. With the side information, the RS configuration can be produced at the both client terminals within a conversation pair. In the next subsection, we will discuss how the server located at the access point coordinates the transmission of those side information and video streams.

5.2.3 Video over WLAN

Figure 5.3 illustrates a flowchart of the proposed system, where user i and j form a conversation pair. Let the video refreshing rate be F frames per second. We divide the time line into F slots per second, and perform distortion management to allocate system resources to every stream within one frame refreshing interval, $T = 1/F$. The distortion management consists of two phases, namely, an initialization phase and a video packet transmission phase. The tasks of initialization phase are to gather R-D information of compressed video streams and channel information, and then to perform resource allocation. The task of video packet transmission phase is to send video packets from users to their corresponding conversation partners.

There are three steps executed in the initialization phase. In the first step, each user's video source coder encodes video in real time and analyzes the R-D of the

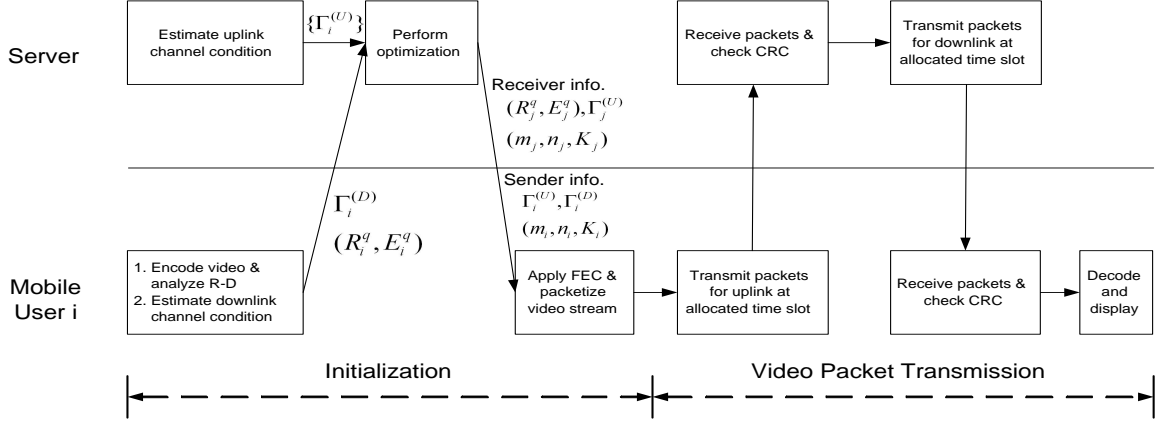


Figure 5.3: Flowchart of the proposed wireless video system. User i and user j are a conversation pair

compressed video bitstream. Meanwhile, each user's communication module estimates the downlink channel condition, and then sends the R-D models, (R_i^q, E_i^q) , along with the estimated channel conditions, $\Gamma_i^{(D)}$, to the resource allocator located at the access point. At the server side, the resource allocator estimates the channel conditions for the uplink, $\Gamma_i^{(U)}$, of all users. In the second step, the resource allocator gathers the R-D information with the channel information, and performs multi-user cross-layer optimization, which is the core of our proposed system and will be discussed in the next section. The resource allocator then informs each user of two sets of transmission configurations. One set is the transmitting configuration for encoding and sending video stream from each user to his/her conversation partner. The configuration information consists of the number of packets to be transmitted, K_i , the selected PHY modes for uplink, m_i , and downlink, n_i , and the channel condition of uplink, $\Gamma_i^{(U)}$ and downlink, $\Gamma_i^{(D)}$. The other set is the receiving configuration for receiving and decoding video stream from each user's conversation partner to himself/herself. This second set of configuration informa-

tion consists of the expected number of packets to be received, K_j , the selected PHY modes for uplink, m_j , and downlink, n_j , the uplink channel condition, $\Gamma_j^{(U)}$, and the R-D models (R_j^q, E_j^q) . In the third step, each user applies FEC and packetizes video packets according to the parameters assigned by the resource allocator. The aforementioned control signals are transmitted through control channels. We assume that the required time in this phase is negligible since the overhead rates of control signals are much smaller than the required rates for transmitting video bitstreams.

After the video is encoded, the coded video packets will be transmitted in the video packet transmission phase, which consists of two steps. In the first step, each user will transmit the FEC coded packets using the assigned PHY mode through an uplink to the access point according to the allocated time slot. In the meantime, the communication module located at the access point will check the cyclic redundancy check (CRC) of each received packet, drop the corrupted packets, and buffer the successfully received packets. In the second step, the server forwards the buffered packets to their destinations using the assigned PHY modes for the downlink path. At each mobile terminal, the communication module checks the CRC of each packet, and gathers the successfully received packets. These packets will be forwarded to the application layer for further processing so that the video frames can be reconstructed for displaying.

The critical issue in this system is how the resource allocator selects the transmission configurations for all users such that the service objective is optimized subject to the system resource constraints. We will formulate a single-cell system as an optimization problem and propose a fast algorithm in Chapter 5.3. We then extend the proposed algorithm to a multi-cell system in Chapter 5.4.

5.3 Joint Uplink and Downlink Optimization for Single-Cell Case

Based on the system described in Chapter 5.2, we first study a simple case where there is only a single cell with intra-cell calls. We begin with a discussion on the video quality model when we jointly consider the channel conditions in both uplink and downlink. The interactive video conferencing system is formulated as a min-max optimization problem, subject to the constraints of maximally allowed transmission time. We will present a fast algorithm to find the transmission configurations for both base and FGS layers.

5.3.1 Problem Formulation

Consider the system has a total of N users. In this system, for user i , we need to determine the PHY mode of the uplink, m_i , and the PHY mode of the downlink, n_i , in the physical layer, as well as the number of packets sent from user i , K_i , in the application layer. To facilitate the discussion, we use a triplet, (m_i, n_i, K_i) , to represent a transmission mode. Assuming all packets of the base layer are received successfully, the end-to-end expected distortion using a transmission mode, (m_i, n_i, K_i) , can be represented as

$$E[D_i(m_i, n_i, K_i)] = D_i^B - \sum_{k=1}^{K_i} p_i(m_i, n_i, K_i, k) \Delta D_i(K_i, k), \quad (5.10)$$

where D_i^B is the distortion after receiving all base layer packets successfully, $\Delta D_i(K_i, k)$ is the distortion reduction if user i 's conversation partner receives one more correct packet after having $k-1$ uncorrupted FGS layer packets, and $p_i(m_i, n_i, K_i, k)$ is the probability that the receiver receives at least k packets successfully when

transmitter sends K_i packets. We have

$$p_i(m_i, n_i, K_i, k) = \sum_{\alpha=k}^{K_i} p_{i,m_i}^{(U)}(K_i, \alpha) p_{i,n_i}^{(D)}(\alpha, k). \quad (5.11)$$

Here, $p_{i,m_i}^{(U)}(K_i, \alpha)$ is the probability that the server receives α packets successfully when user i sends K_i packets :

$$p_{i,m_i}^{(U)}(K_i, \alpha) \triangleq \binom{K_i}{\alpha} \left(1 - p_{i,m_i}^{(U)}\right)^{K_i - \alpha} \left(p_{i,m_i}^{(U)}\right)^{\alpha}, \quad (5.12)$$

and $p_{i,n_i}^{(D)}(\alpha, k)$ is the probability that the receiver receives at least k packets successfully when the server sends α packets

$$p_{i,n_i}^{(D)}(\alpha, k) \triangleq \sum_{\beta=k}^{\alpha} \binom{\alpha}{\beta} \left(1 - p_{i,n_i}^{(D)}\right)^{\alpha - \beta} \left(p_{i,n_i}^{(D)}\right)^{\beta}. \quad (5.13)$$

As mentioned in Chapter 5.2.3, we gather up-to-date video source R-D and channel information within each frame refreshing interval. The end-to-end expected distortion has captured the time-varying video source and channel condition.

To support interactive video conferencing, we set the maximum transmission delay as one video frame refreshing interval, i.e. T second. Thus, the encoded bitstream of each video frame should arrive at the end user within the refreshing interval of every video frame. As mentioned in Chapter 5.2.1, we consider a system where there is only one user who can send data at any moment in one cell. Let t_i be the assigned amount of time for user i to send a video frame to his/her conversation partner through uplink and then downlink. The overall transmission time of all users, $\sum_{i=1}^N t_i$, should not exceed T seconds. Note that the amount of time to transmit a fixed-length packet depends on which PHY mode we apply. Denote T_x^{max} as the required transmission time if the PHY mode x is selected to transmit a packet in a single path. Thus, if user i selects PHY mode for uplink

and downlink as m_i and n_i , respectively, and sends K_i packets from the sender to the server, the expected transmission time along user i 's uplink is

$$t_i^{(U)}(m_i, n_i, K_i) = K_i T_{m_i}^{max}. \quad (5.14)$$

The expected number of packets successfully arrived at access point is $p_{i,m_i}^{(U)} K_i$, and expected transmission time along user i 's downlink is

$$t_i^{(D)}(m_i, n_i, K_i) = K_i p_{i,m_i}^{(U)} T_{n_i}^{max}. \quad (5.15)$$

The overall expected transmission time from user i through the server to his/her conversation partner is:

$$t_i(m_i, n_i, K_i) = t_i^{(U)}(m_i, n_i, K_i) + t_i^{(D)}(m_i, n_i, K_i). \quad (5.16)$$

We formulate the overall distortion management problem in the video conferencing system as an optimization problem that searches for each user's transmission mode to minimize the maximum of all users' expected distortion, subject to the maximal available transmission time. That is,

$$\min_i \max_{\{m_i, n_i, K_i\}} w_i \cdot f(E[D_i(m_i, n_i, K_i)]) \quad (5.17)$$

$$\text{subject to } \sum_{i=1}^N t_i(m_i, n_i, K_i) \leq T,$$

where w_i is the quality weighting factor and $f(\cdot)$ the perceptual distortion function. Because of the integer valued parameters in transmission mode, the problem (5.17) is NP hard. The complexity of finding the optimal transmission modes for all N users through full search is $O(\kappa^N)$, where κ is the number of all feasible transmission modes. To meet the real-time requirement of the proposed system, we propose a fast algorithm in the next subsection to find a near-optimal

solution to problem (5.17). As a proof of concept, we consider the case of providing uniform mean squared distortion among all users, i.e., $w_i = 1, \forall i$ and $f(E[D_i(m_i, n_i, K_i)]) = E[D_i(m_i, n_i, K_i)]$.

5.3.2 Proposed Algorithm

Because the base and FGS layer have different properties and importance, we propose two different resource allocation strategies for each layer. The goal of resource allocation in the base layer is to provide a strong error protection and to reduce the overall transmission time used in the base layer so that the remaining transmission time can be used for sending the FGS layer. For the FGS layer, the resource allocation strategy is to prune out inefficient transmission modes and to find the optimal solutions that gives the lowest maximal distortion among all users.

1) *Base Layer*: The base layer data at rate R_i^0 of user i requires $K_i^{B,S} = \lceil R_i^0/L \rceil$ source packets. The remaining rates of the last source packet, $K_i^{B,S}L - R_i^0$, is filled with the first part of the FGS layer bit stream. We need to determine the uplink and downlink PHY mode (m_i, n_i) and the number of parity packets, $K_i^{B,P}$, such that the resulted transmission time for the base layer is the shortest and the end-to-end BER is kept lower than a threshold. In this chapter, we set the threshold $\text{BER}^B = 10^{-6}$ as suggested in [26].

The BER requirement can be attained in three steps: we first examine the smallest number of required parity packets for each (m_i, n_i) to achieve $p_i(m_i, n_i, K_i^{B,S} + K_i^{B,P}, K_i^{B,S}) \geq (1 - \text{BER}^B)$ using (5.11); then calculate the corresponding transmission time $t_i^B(m_i, n_i)$ using (5.16); and finally find the setting with the shortest transmission time

$$(\hat{m}_i, \hat{n}_i) = \arg \min_{\{m_i, n_i\}} t_i^B(m_i, n_i). \quad (5.18)$$

Denote t_i^B as the transmission time using mode (\hat{m}_i, \hat{n}_i) . Thus, the overall transmission time for all users is $T^B = \sum_{i=1}^N t_i^B$, and the remaining transmission time for FGS layer is $T^F = T - T^B$. An outage is reported if T^B exceeds T , which suggests that there are too many users in the system and there are not even enough resources to support base layer.

2) *FGS Layer*: To reduce the high dimensionality of the search space, we propose a two-step algorithm by first obtaining a one-to-one mapping function between transmission time and expected distortion (T-D) for each user and then using a bisection search in all T-D functions to obtain the solutions. The T-D function can be obtained by first finding a set of efficient transmission modes. A transmission mode (m_i, n_i, K_i^F) is *efficient* if $E[D_i(m'_i, n'_i, K_i^{F'})] < E[D_i(m_i, n_i, K_i^F)]$ for all other modes $(m'_i, n'_i, K_i^{F'})$ with $t_i(m'_i, n'_i, K_i^{F'}) > t_i(m_i, n_i, K_i^F)$. We can collect all efficient transmission modes $\{(m_i, n_i, K_i^F)\}$ as set \mathcal{S}_i and the corresponding transmission time $\{t_i(m_i, n_i, K_i^F)\}$ as set \mathcal{T}_i iteratively. The search algorithm starts from receiving only base layer packets as the first efficient transmission mode. By given an efficient mode with distortion D_s and transmission time T_s , the search algorithm finds the next nearest efficient mode by first pruning out all modes with distortion no less than D_s and then searching the mode with the smallest increased transmission time from current mode. Let $\{t_{i,k}\}$ be the transmission time sorted in an increasing order in \mathcal{T}_i and the corresponding expected distortion for each transmission time $t_{i,k}$ can be obtained. Bring all $\{t_{i,k}\}$ and the corresponding expected distortion together, we have a time-distortion function $E[\tilde{D}_i[t_{i,k}]]$ for user i . The algorithm to obtain a T-D function is summarized in Table 5.2. The complexity of obtaining a T-D function for the worst case is $O(\kappa)$.

Figure 5.4 shows an example how to obtain the T-D function for a user by

Table 5.2: Proposed algorithm to obtain transmission time to expected distortion function

<p>a) Initialization:</p> <ol style="list-style-type: none"> 1) <i>Feasible set:</i> $S_i = \{(m_i, n_i, K_i), \forall m_i, n_i, K_i\}$ 2) <i>Thresholds:</i> $D_s = D_i^B$ and $T_s = 0$, 3) <i>T-D function list:</i> $k = 0, t_{i,k} = T_s, E[\tilde{D}_i[t_{i,k}]] = D_s, .$
<p>b) Obtain T-D function:</p> <p><i>While</i> $S_i > 0$</p> <ol style="list-style-type: none"> 1) Find the next efficient mode. <i>For each</i> $(m_i, n_i, K_i) \in S_i$ $NT(m_i, n_i, K_i) \triangleq t_i(m_i, n_i, K_i) - T_s.$ $(\hat{m}_i, \hat{n}_i, \hat{K}_i) = \arg \min_{\{m_i, n_i, K_i\}} \{NT(m_i, n_i, K_i)\}.$ 2) Add $(\hat{m}_i, \hat{n}_i, \hat{K}_i)$ to the T-D function list. $t_{i,k} = t_i(\hat{m}_i, \hat{n}_i, \hat{K}_i),$ $E[\tilde{D}_i[t_{i,k}]] = E \left[D_i(\hat{m}_i, \hat{n}_i, \hat{K}_i) \right],$ $k = k + 1.$ 3) Update thresholds T_s and D_s. $T_s = t_i(\hat{m}_i, \hat{n}_i, \hat{K}_i),$ $D_s = E \left[D_i(\hat{m}_i, \hat{n}_i, \hat{K}_i) \right].$ 4) Remove modes whose distortion $\geq D_s$ from feasible set. <i>For each</i> $(m_i, n_i, K_i) \in S_i$, <i>If</i> $E[D_i(m_i, n_i, K_i)] \geq D_s$ $S = S \setminus (m_i, n_i, K_i).$ <p><i>End</i></p>

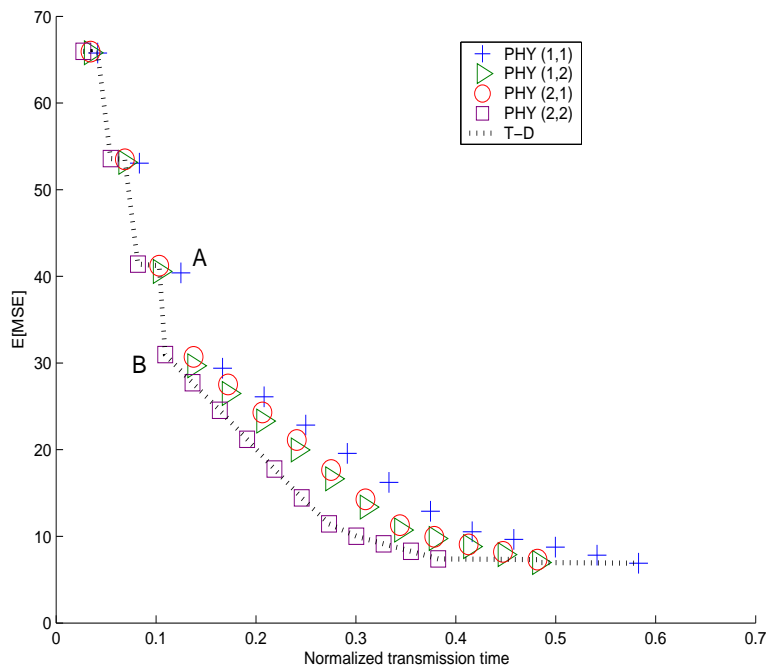


Figure 5.4: The Time-Distortion function

considering only PHY mode index 1 and 2. Let $\text{PHY}(a,b)$ represent the two selected PHY modes for uplink, a , and for downlink, b . For each $\text{PHY}(a,b)$, we can obtain a curve for the expected transmission time and the expected distortion by using different numbers of packets. Since users choose two PHY modes for uplink and two PHY modes for downlink for a packet, there are four different curves shown in Figure 5.4. As we can see, Point A is not an *efficient* transmission mode because we can find other transmission modes with smaller distortion and shorter transmission time (such as Point B). On the other hand, Point B is an *efficient* transmission mode. After finding all efficient transmission modes, we can collect them as a T-D function, as shown a dotted line in Figure 5.4. In general, the resulting T-D function contains points from different $\text{PHY}(a,b)$ modes.

After obtaining all T-D functions for all users, the problem (5.17) can be re-

formulated as

$$\begin{aligned} & \min_i \max_{\{t_{i,k}\}} E \left[\tilde{D}_i[t_{i,k}] \right] & (5.19) \\ & \text{subject to } \sum_{i=1}^N t_{i,k} \leq T^F. \end{aligned}$$

Based on the definition of efficient transmission mode, all T-D functions are monotonically decreasing. We solve the problem (5.19) using bi-section search. The search algorithm calculates the total required time to achieve a targeted distortion, and then increases the targeted distortion at the next iteration if the total required time is higher than the time constraint, T^F , and vice versa. The overall number of iterations is determined by the computation precision used in bi-section search. If T-D functions are continuous and monotonically decreasing, the solution provided by bi-section search is optimal. However, due to the discrete nature of T-D function as shown in Figure 5.4, the problem (5.19) is *NP* hard [60] and the solution provided by bi-section search is suboptimal. After determining $t_{i,k}$ for all users, we can obtain corresponding transmission mode of each user (m_i, n_i, K_i^F) from \mathcal{S}_i .

5.4 Joint Uplink and Downlink Optimization for Multi-Cell Case

In this section, we consider a video conferencing system supporting multiple cells. We first present the proposed system framework and discuss different types of conversation calls hold within multiple cells. We formulate this multi-cell system as an optimization problem to minimize the maximal distortion among all users, and extend the proposed single-cell algorithm to the multi-cell system.

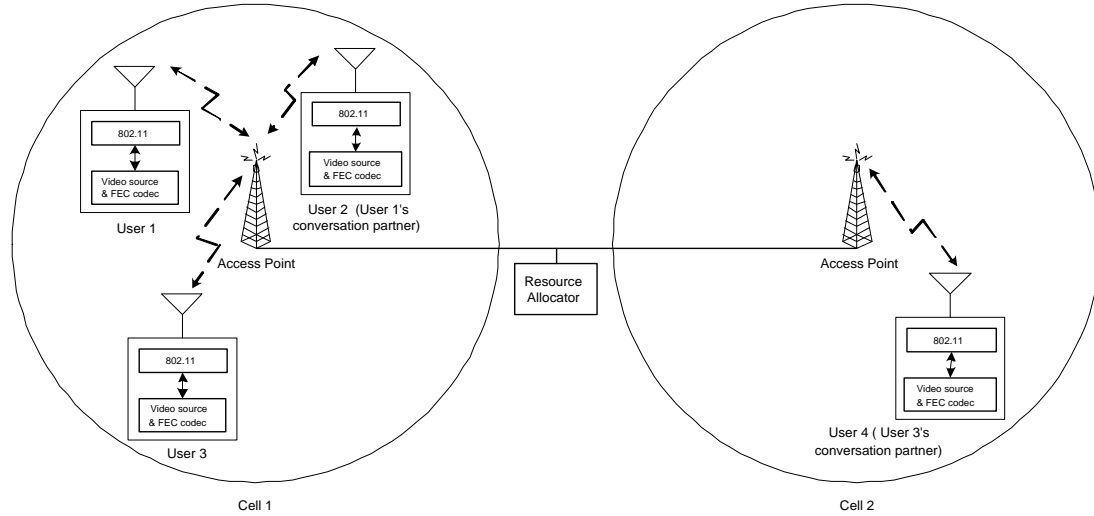


Figure 5.5: System block diagram for multi-cell case

5.4.1 System Framework

Figure 5.5 shows the proposed framework for multiple cells. Without loss of generality, here we use a system with two cells as an example to illustrate. Without loss of generality, we use a system with two cells as an example. For simplicity, we assume that these two cells are far away, i.e. two sites of a company, and won't interfere to each other. Both cells are connected by a wired channel which is reliable without any packet loss and whose bandwidth is large enough to transmit all packets. We also assume the coherent time of the channel condition is much larger than the propagation delays induced by the wired link. A user can have either an intra-cell conversation with a user within the same cell (e.g. the conversation between user 1 and 2 in Figure 5.5), or an inter-cell conversation with a user located in another cell (e.g. conversation between user 3 and 4). Similar to the distortion management used in the single-cell case, the resource allocator needs to first gather R-D information of all video streams and channel information of all

links, and then performs distortion control. Note that in this multi-cell system, there is only one transmitter which can send data at any time instance in each cell, and thus there are two transmitters which can transmit video packets to the corresponding receivers simultaneously in this two-cell system. The major tasks of the resource allocator are how to jointly consider the traffic load in both cells and how to allocate system resources to each user in each link such that the maximal distortion among all users is minimized.

5.4.2 Problem Formulation

Suppose there are C cells in the proposed conferencing system. Let $S_c^{(U)}$ and $S_c^{(D)}$ be the set of users who have requested uplink channel and downlink channel to send video streams in the c^{th} cell, respectively. As an example shown in Figure 5.5, $S_1^{(U)} = \{1, 2, 3\}$, $S_1^{(D)} = \{1, 2, 4\}$, $S_2^{(U)} = \{4\}$, and $S_2^{(D)} = \{3\}$. We can formulate this video conferencing system as an optimization problem that chooses each user's transmission mode to minimize the maximum of all users' expected distortion, subject to the maximal available transmission time constraint in each cell:

$$\min_i \max_{\{m_i, n_i, K_i\}} E [D_i(m_i, n_i, K_i)] \quad (5.20)$$

$$\text{subject to } \sum_{i \in S_c^{(U)}} t_i^{(U)}(m_i, n_i, K_i) + \sum_{i \in S_c^{(D)}} t_i^{(D)}(m_i, n_i, K_i) \leq T_c, \text{ for } c = 1, 2, \dots, C.$$

Unlike the single-cell system containing only intra-cell calls, a multi-cell system needs to consider the inter-cell conversation pairs whose packets are transmitted from cells to cells. The traffic load in different cells may be different and adjusting traffic load in one cell will affect other cells' load through the inter-cell calls. We should jointly allocate time slots in both cells for the inter-cell calls and evaluate the time constraints in both cells. In fact, the problem (5.20) is a generalized

assignment problem, which is *NP* hard [60]. To meet the real-time requirement, we propose a fast and suboptimal algorithm by extending the single-cell algorithm.

5.4.3 Proposed Algorithm

Similar to the single-cell case, we adopt a two-stage strategy to allocate system resources for the base layer first and then for the FGS layer.

1) *Base Layer* : In parallel to the single-cell case, we calculate the required number of packets, $K_i^{B,S}$, to carry all base layer's bitstream. We then find the optimal transmission mode $(\hat{m}_i, \hat{n}_i, K_i^{B,S} + K_i^{B,P})$ that has the shortest overall transmission time in both cells with end-to-end BER lower than the BER threshold, BER^B . Once the transmission modes are determined, the overall allocated transmission time for base layer in each cell can be determined as:

$$T_c^B = \sum_{i \in S_c^{(U)}} t_i^{(U)}(\hat{m}_i, \hat{n}_i, K_i^{B,S} + K_i^{B,P}) + \sum_{i \in S_c^{(D)}} t_i^{(D)}(\hat{m}_i, \hat{n}_i, K_i^{B,S} + K_i^{B,P}), \forall c. \quad (5.21)$$

Subsequently, we can calculate the rest transmission time, $T_c^F = T - T_c^B$, to transmit FGS layer's data in each cell.

2) *FGS Layer*: We first obtain the T-D functions, $E[\tilde{D}_i[t_{i,k}]]$, for all users using Table 5.2. For each valid $t_{i,k}$, we can know its corresponding transmission time in the uplink path alone, $t_{i,k}^{(U)}$, and in the downlink path alone, $t_{i,k}^{(D)}$. We reformulate the problem (5.20) as

$$\begin{aligned} & \min_i \max_{\{t_{i,k}\}} E \left[\tilde{D}_i[t_{i,k}] \right] \\ & \text{subject to } \sum_{i \in S_c^{(U)}} t_{i,k}^{(U)} + \sum_{i \in S_c^{(D)}} t_{i,k}^{(D)} \leq T_c^F, \text{ for } c = 1, 2, \dots, C. \end{aligned} \quad (5.22)$$

To solve this problem, we propose a fast algorithm performing multiple round of bi-section search on all T-D functions, as shown in Figure 5.6. For a targeted distortion, the search algorithm calculates the total required transmission time including

all uplinks and downlinks in each cell. If there is at least one cell whose overall required time is higher than the corresponding time constraint, T_c^F , the algorithm increases the targeted distortion to reduce the required amount of transmission time in the next iteration, and vice versa. Because the numbers of intra-cell calls and inter-cell calls are different in each cell, the available FGS transmission time in each cell is different. The allocated transmission time in some cells will reach the limit of time constraints first, and some cells may still have unallocated transmission time left. Thus, performing only one round of bi-section search to maintain strict fairness among all users may waste system resources in some cells. To efficiently utilize the remaining system resources, we allow further rounds of bi-section search to reduce users' distortion. A cell is defined as *inactive* if there is no more transmission time left for FGS layer. A user is *inactive* if either uplink or downlink of the corresponding video streaming path is in an inactive cell. Once a round of bi-section search is finished, the proposed multi-cell algorithm will remove the inactive cells and inactive users from the further assignment list. Then, another round of bi-section search is performed on the T-D functions of all active users subject to the set of time constraints in the active cells. The whole algorithm terminates when there are no more active users in this system.

5.5 Simulation Results

In this section, we evaluate the performance of our proposed scheme and compare it with a traditional sequential optimization scheme. This traditional scheme assigns equal bandwidth to each uplink and downlink and allocates system resources to each link independently. More specifically, the resource allocator first allocates the optimal configuration based on only the uplink channel information and the mobile

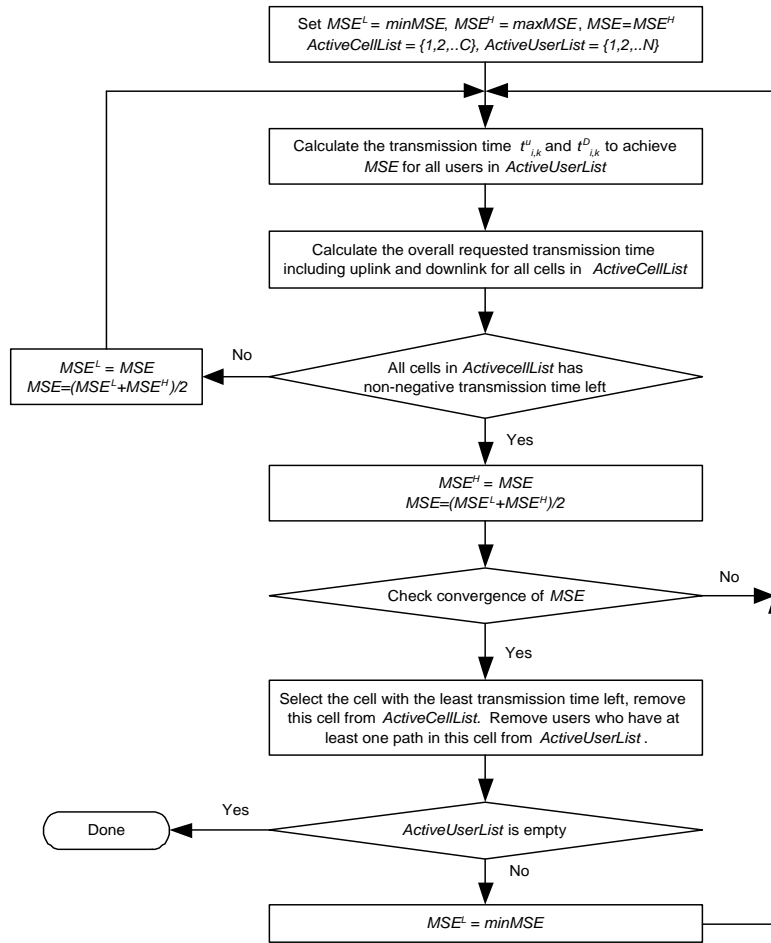


Figure 5.6: Proposed algorithm for multi-cell case.

users transmit packets to access point during the first half of available transmission time. Then, based on the packets received successfully by the access point, the resource allocator optimizes the downlink configuration and the server transmits packets to each mobile user during the second half of available transmission time. We first describe the simulation setup and the performance criteria used to examine both schemes, and then present simulation results for both schemes within a single cell and multiple cells, respectively.

5.5.1 Simulation Setup

The simulations are set up as follows. The noise power is 10^{-10} Watts and the maximal transmission power for both mobile user and server is 40 mW. The path loss factor is 2.5. Packet length L is set to 512 bytes. The video format is QCIF (176×144) with refreshing rate as 30 frames per second and thus $T = 33.33$ ms. The base layer is generated by MPEG-4 encoder with a fixed quantization step of 30 and the GOP pattern is 29 P frames after one I frame. All frames of FGS layer have up to six bit planes. We concatenate 15 QCIF video sequences to form one testing video sequence of 5760 frames. The 15 sequences are 300-frame *Akiyo*, 360-frame *carphone*, 480-frame *Claire*, 300-frame *coastguard*, 300-frame *container*, 390-frame *foreman*, 870-frame *grandmother*, 330-frame *hall objects*, 150-frame *Miss American*, 960-frame *mother and daughter*, 300-frame *MPEG4 news*, 420-frame *salesman*, 300-frame *silent*, 150-frame *Suzie*, and 150-frame *Trevor*.

A simulation profile for an N-user system is defined as follows: the video content program for each user is 90-frame long, the first video frame starts from a randomly selected frame of the concatenated video, and the location for each user is randomly selected between 20 m to 100 m. For each simulation profile, we repeat

the simulations 100 times and take the average.

5.5.2 Performance Criteria

Four performance criteria are used to evaluate the proposed scheme and the traditional scheme. Let $\text{PSNR}_{i,n}$ denote the PSNR of the received video frame n for user i . Since the service objective in the problem (5.17) is to minimize the maximal distortion, our first performance metric is the worst received video quality among all users. We measure the minimal PSNR among all users at frame n as $\text{minPSNR}_n = \min_i \{\text{PSNR}_{i,n}\}$ and take the average of the minimal PSNRs' over M video frames :

$$\text{minPSNR} = \frac{1}{M} \sum_{n=1}^M \text{minPSNR}_n. \quad (5.23)$$

The second metric is the average video quality received by all users, averaged over M frames :

$$\text{avePSNR} = \frac{1}{M} \sum_{n=1}^M \text{PSNR}_n, \quad (5.24)$$

where PSNR_n is the average received video quality of all users' n^{th} video frame. The higher avePSNR is, the higher system efficiency in terms of overall video quality we have.

The third metric measures the fairness through examining the received video qualities for users who subscribe the same video quality level. To quantify the fairness, we calculate the standard deviation for all users' n^{th} video frame and take the average along the whole M -frame video, i.e.,

$$\text{stdPSNR} = \frac{1}{M} \sum_{n=1}^M \left\{ \frac{1}{N} \sum_{i=1}^N (\text{PSNR}_{i,n} - \text{PSNR}_n)^2 \right\}^{\frac{1}{2}}. \quad (5.25)$$

The lower stdPSNR is, the fairer video quality each user receives.

The forth metric concerns the quality fluctuation. Because significant quality differences between consecutive frames can bring irritating flickering and other artifacts to viewers even when the average video quality is acceptable. To quantify the fluctuation of quality between nearby frames, we use the mean absolute difference of consecutive frames' PSNR, madPSNR , to measure the perceptual fluctuation and take the average over N users:

$$\text{madPSNR} = \frac{1}{N} \sum_{i=1}^N \left\{ \frac{1}{M-1} \sum_{n=2}^M |\text{PSNR}_{i,n} - \text{PSNR}_{i,n-1}| \right\}. \quad (5.26)$$

5.5.3 Single-Cell Case

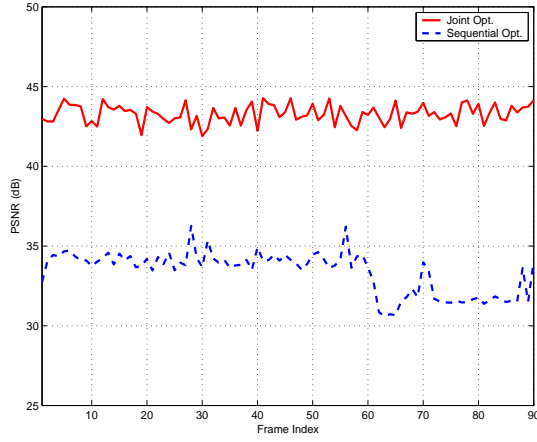
We first use a four-user system to illustrate the proposed scheme to achieve fair video quality. User 1,2 and User 3,4 are teamed up to form two conversation pairs. The locations of User 1 to 4 are 91m, 67m, 71m, and 20m away from the access point, respectively. For the video content, User 1 to 4 send one frame of video sequence, *Akiyo*, *carphone*, *Claire*, and *foreman* to their corresponding conversation partner, respectively. The selected transmission modes for the FGS layer using the proposed algorithm are summarized in Table 5.3. As we can see, User 1 to 4 selects uplink PHY modes as 4, 6, 4, and 8, respectively; and downlink PHY modes as 5, 4, 7, and 5, respectively. As expected, a link with longer transmission distance or worse channel condition requires a higher level of error protection (i.e., smaller PHY mode) to protect video packets. We then compare the required number of packets for each user. The required number of packets for User 2 and 4 are 32 and 24, respectively, which are higher than the 17 and 8 packets for User 1 and 3, respectively. This is because User 2's sequence, *carphone*, and User 4's sequence, *foreman*, have higher content complexity than the other two sequences and require more packets to achieve similar video quality. The overall transmission time for

Table 5.3: Selected transmission modes for FGS layer

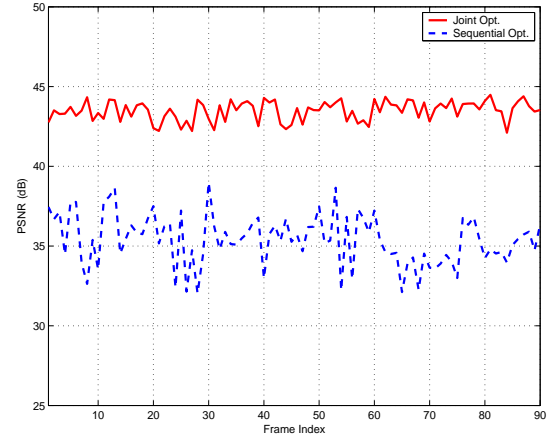
	user 1	user 2	user 3	user 4
Sent video sequence	<i>Akiyo</i>	<i>carphone</i>	<i>Claire</i>	<i>foreman</i>
Uplink distance (m)	91	67	71	20
Downlink distance (m)	67	91	20	71
Uplink PHY mode, m_i	4	6	4	8
Downlink PHY mode, n_i	5	4	7	5
Number of packet, K_i	17	32	8	24
Transmission time, t_i (ms)	6.4	8.9	2.5	5.9
Received PSNR (dB)	42.97	42.75	42.90	42.49

each video stream depends on the number of packets along with the selected PHY modes; and is calculated using (5.14), (5.15), and (5.16). Finally, we evaluate the final reconstructed video quality. As shown in Table 5.3, the video quality sent from User 1 to 4 are 42.97, 42.75, 42.90, and 42.49 dB, respectively, maintaining a good amount of fairness.

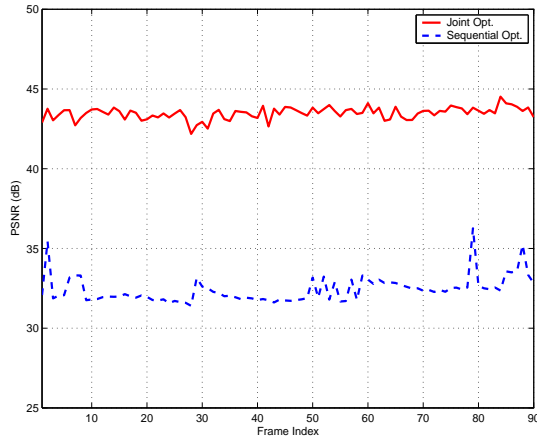
We compare the proposed scheme with the sequential optimization scheme by keeping the same simulation setting as mentioned above, except that each user sends a 90-frame video sequence to his/her conversation partner. We repeat the experiments 100 times to calculate the average PSNR for each frame. Figure 5.7 shows the frame-by-frame PSNR. As shown, the proposed scheme can provide higher minimal and average PSNR, more uniform video quality among all users, and lower quality fluctuation along each received video sequence than the sequential optimization scheme. The performance gain is due to dynamical bandwidth allocation by the proposed scheme to users in uplink and downlink transmission paths. Note that the sequential optimization scheme allocates fixed $T/2$ seconds for overall uplinks and another $T/2$ seconds for overall downlinks. Because the



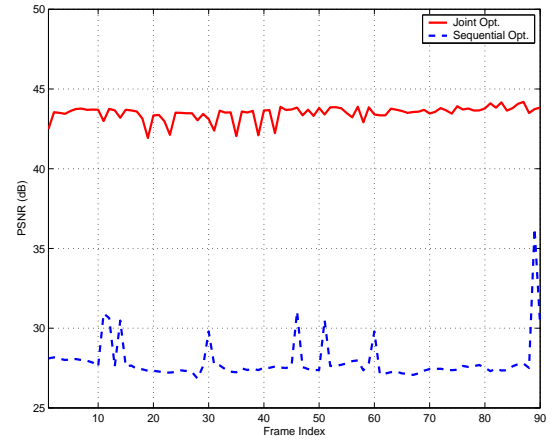
(a) User 1



(b) User 2



(c) User 3

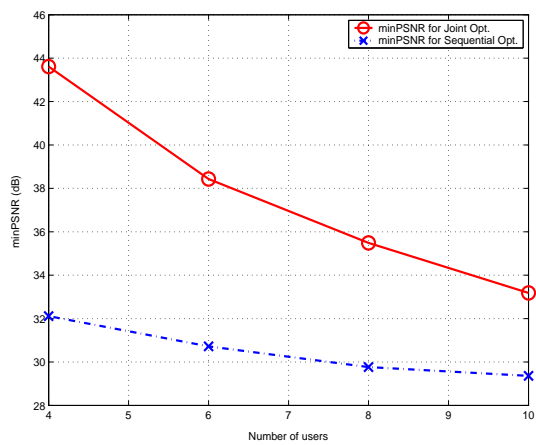


(d) User 4

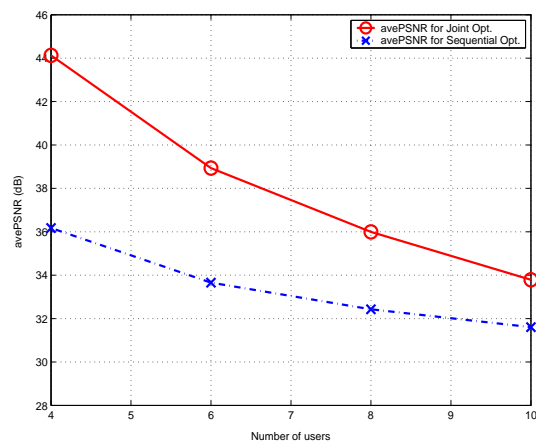
Figure 5.7: Frame-by-frame PSNR for User 1 to User 4.

asymmetric channel conditions along uplink and downlink for each video stream and the time heterogeneity of video content, the sequential optimization scheme lacks the freedom to dynamically adjust the time budget for uplink and downlink to attain better video quality.

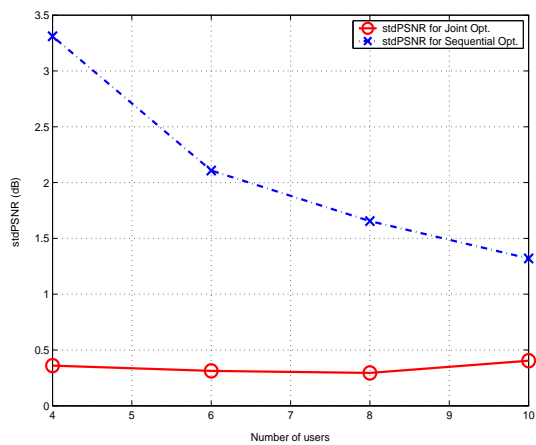
We evaluate the performance of both schemes with different number of users within a single cell and show the results in Figure 5.8. We average the results from 100 simulation profiles as described in Chapter 5.5.1, and calculate the minPSNR,



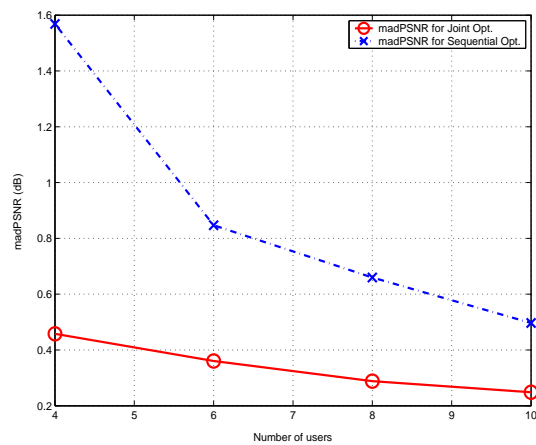
(a) minPSNR



(b) avePSNR



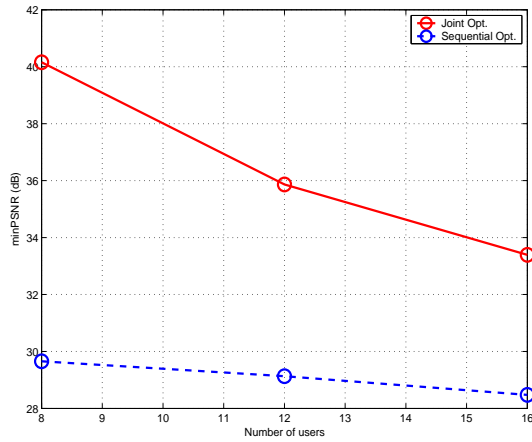
(c) stdPSNR



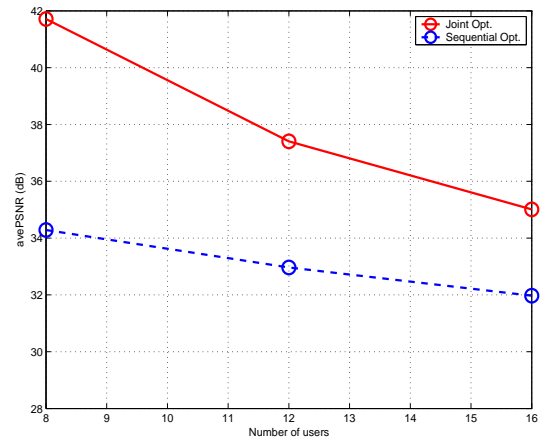
(d) madPSNR

Figure 5.8: PSNR performance results for different number of users for single-cell case.

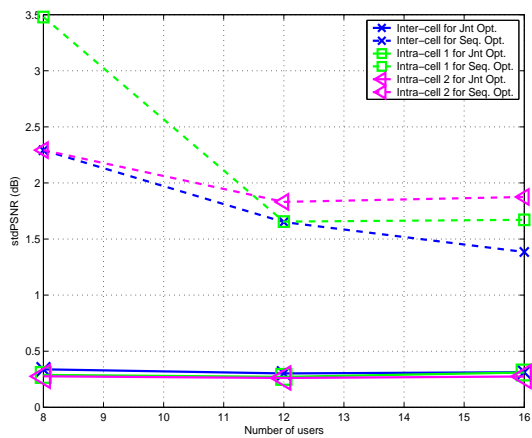
avePSNR, stdPSNR, and madPSNR as defined in Chapter 5.5.2. We can see in Figure 5.8(a) that, for the minPSNR criterion, the proposed joint optimization scheme outperforms the sequential optimization scheme 3.82 ~ 11.50 dB. In other words, the worst received quality among all users in the proposed scheme has a substantial improvement over the one in the sequential optimization scheme. Comparing the avePSNR as shown in Figure 5.8(b) and the stdPSNR as shown in Figure 5.8(c), the proposed scheme has higher overall quality by 2.18 ~ 7.95 dB and lower quality deviation by 0.92 ~ 2.95 dB among all users than the sequential optimization scheme. The proposed algorithm can provide not only higher overall users' video quality but also more uniform video quality among all users. In general, a system with more users can leverage the diversity of video content complexity to provide more consistent video qualities to all users. However, we observe that the stdPSNR for the proposed system with ten users is slightly higher than the one with eight users. This is because the available FGS transmission time for the system with ten users is close to 0. In most cases, the system can allocate transmission time for the base layer only. Consequently, there are less transmission time budget left for FGS bitstreams to compensate the quality deviation among users contributed to the base layer, which results in higher stdPSNR. Figure 5.8(d) shows the quality fluctuation along each received video sequence for both schemes. The proposed scheme can achieve 0.25 ~ 1.11 dB lower than the sequential optimization scheme. By exploring multiuser diversity, the more users the proposed system has, the lower quality fluctuation each user experiences.



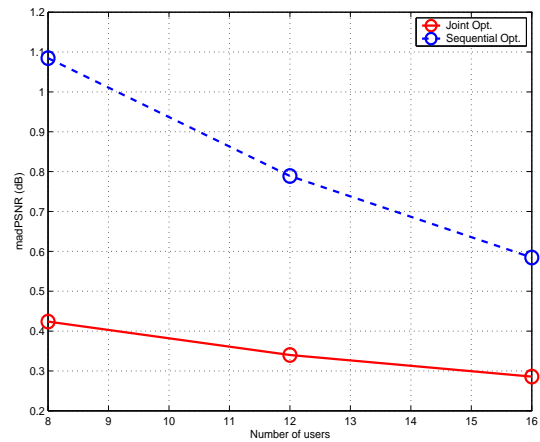
(a) minPSNR



(b) avePSNR



(c) stdPSNR



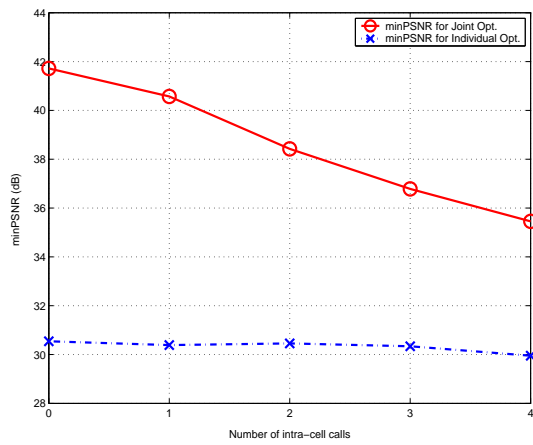
(d) madPSNR

Figure 5.9: PSNR results for different number of users for two-cell case. In (c), \times , \square , and \triangleleft represent the results for inter-cell, intra-cell 1, and intra-cell 2 call, respectively. Solid line and dotted line represent the results for the proposed algorithm and the traditional sequential algorithm, respectively.

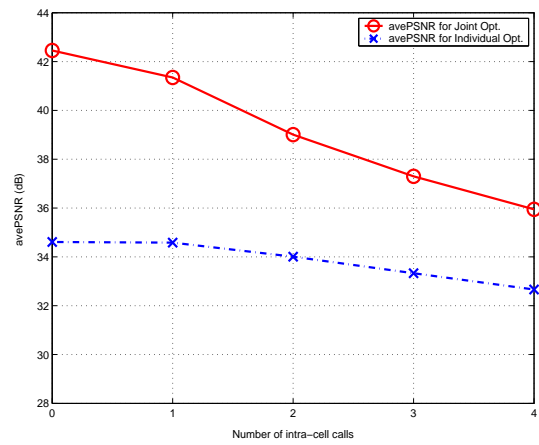
5.5.4 Multiple-Cell Case

For the multi-cell case, without loss of generality, we simulate a two-cell system in which there are 8, 12, and 16 users. For each simulation profile, each user is randomly located in either cell, the distance from each user to his/her cell's access point is randomly selected between 20 m to 100m, and each user's first video frame is also randomly picked from the testing video sequence. We repeat the simulation using 100 different profiles and average the results to evaluate the performance. Figure 5.9(a) and 5.9(b) show the minPSNR and avePSNR using both schemes for different number of users in this system, respectively. The proposed joint uplink and downlink optimization scheme outperforms the sequential uplink and downlink optimization scheme by 4.92 ~ 10.50 dB for the minimal PSNR and by 3.04 ~ 7.43 dB for the average PSNR. Since there are three different types of video conferencing calls in this system, namely, inter-cell call between cell one and two, intra-cell call within cell one, and intra-cell call within cell two. we shall compare the stdPSNR for each type of call separately. As revealed by Figure 5.9(c), the proposed algorithm can provide lower quality deviation for all three types of calls. Figure 5.9(d) shows the quality fluctuation along each received video sequence for both schemes, suggesting that the proposed scheme provides lower quality fluctuation than the sequential optimization scheme. In summary, the proposed scheme can provide higher minPSNR, higher avePSNR, lower stdPSNR, and lower madPSNR, which again demonstrates the superiority of joint uplink and downlink optimization.

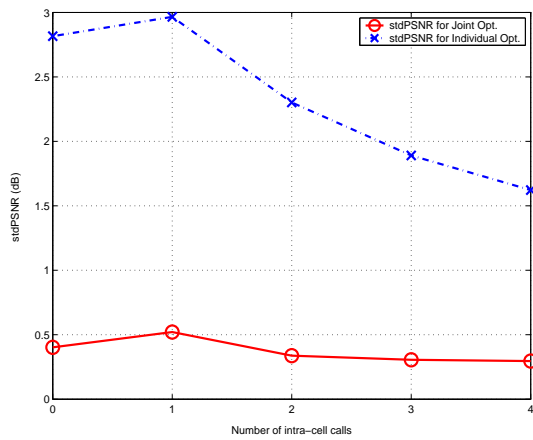
To study the bottleneck effect caused by different traffic loads over different cells, we conduct another simulation in which there are 8 users and there are only two types of calls, namely, inter-cell call between cell one/two and intra-cell call



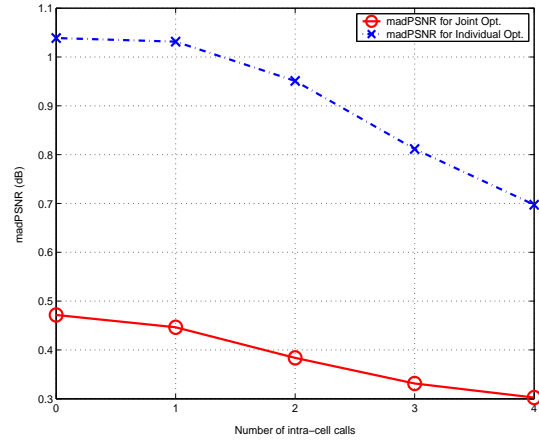
(a) minPSNR



(b) avePSNR



(c) stdPSNR



(d) madPSNR

Figure 5.10: PSNR results for different number of intra-cell calls for two-cell case with 8-users.

within cell one. The PSNR performances with various number of intra-cell calls in cell one are shown in Figure 5.10. If the system has more intra-cell calls within cell one, there are more users requesting bandwidth to deliver video streams such that cell one becomes the system performance bottleneck. Consequently, the allocated bandwidth for each user is reduced, and the received video quality decreases.

Figure 5.10(c) shows that the stdPSNR of a system with only one intra-cell call is slightly higher than the one without any intra-cell calls. This is because 7% of simulation profiles have all users in cell two being far away from the access point. These users adopt higher level of error protection to transmit video streams and thus require longer transmission time along the corresponding uplinks and downlinks in cell two than in cell one. Therefore, the available transmission time in cell two will saturate earlier than cell one. To utilize unassigned transmission time in cell one, our algorithm performs another round of bi-section search in cell one. It results in two different levels of video quality in the overall system and the quality deviation among all users increases.

5.6 Chapter Summary

In summary, we have constructed a joint network-aware and source-aware video conferencing framework for multiple conversation pairs within IEEE 802.11 networks. The proposed framework dynamically performs multi-dimensional resource allocation by jointly exploring the cross-layer error protection, multi-user diversity, and the heterogeneous channel conditions in all paths. We formulate the system as a min-max optimization problem to provide satisfactory video quality for all users. A fast algorithm that converts system resources into time-distortion functions is proposed to obtain the transmission configuration for each user in both single-cell

and multi-cell scenario.

We compare the proposed scheme with a traditional scheme that performs sequential optimization for uplink and downlink. Our experiments demonstrated that the proposed scheme for a single cell scenario can obtain a 2.18 ~ 7.95 dB gain for the average received PSNR of all users and a 3.82 ~ 11.50 dB gain for the minimal received PSNR among all users. For a two-cell case, the proposed scheme can achieve a 4.92 ~ 10.50 dB gain for the worst received quality among all users and a 3.04 ~ 7.43 dB gain for the average video quality. In addition, the proposed scheme can provide more uniform video quality among all users and lower quality fluctuation along each received video sequence.

Chapter 6

PDMA-Based Error Protection

6.1 Introduction

In this chapter, we discuss the scenario that multiple video streams are merged at aggregation points and the merged stream is transmitted between multiple aggregation points till final destinations [66]. Multi-point video conferencing [12] and digital video surveillance are such examples. There may exist multiple paths with different channel conditions between two aggregation points. The main challenge is how to jointly apply stream aggregation, rate control, and error protection such that the transmitted streams can arrive in time with the highest overall received quality.

When there are multiple transmission paths between a sender and a receiver, path diversity can be explored: at a given time, the available bandwidth in each path may not be the same, and the probability for all paths having severe packet loss simultaneously is very low [3, 4, 24, 59]. If the channel conditions can be fed back from the receiver to the sender, the streaming video systems can be formulated as optimization problems to achieve the optimal expected video quality

by determining the stream partitioning, such as through multiple-layer coding [5] or multiple-description coding [110]; and the corresponding path assignment, namely, which partition is transmitted through which path [118, 129]. An R-D optimized transmission policy was proposed [10] to schedule packet transmission time to meet playback deadline. Video transmission system can further deploy FEC in each transmission path to alleviate the bit error and packet loss problem [47, 119]. By jointly selecting the channel coding rate and source coding rate in each individual transmission path, the overall end-to-end video quality reconstructed from multiple partitions can be improved.

Encoding video sources into multiple layers or multiple descriptions has penalty on compression efficiency: given the same bit rate budget, the perceptual quality reconstructed from multiple layers or descriptions is lower than single layer or single description [65, 110]. To take the advantage of higher coding efficiency achieved by adopting only one single video stream, we can apply joint FEC across multiple transmission paths to protect a single stream instead of individual FEC protection for each substream in each path. The basic idea of a cross-path FEC scheme is to jointly consider heterogeneous channel conditions in all paths and determine only one set of the optimal FEC configuration to generate parity check packets. Then, we will split FEC coded packets into groups according to the bandwidth in each path and transmit each group along the corresponding path.

Several works on cross-path error protection have been proposed by either assuming homogeneous channel conditions among all paths [120] or considering heterogeneous channel conditions among all paths [57, 63]. Cross-path error protection has the potential to have lower packet loss rate after FEC decoding compared to individual-path error protection in each path when both schemes have the same

number of source and parity check packets summed over all paths. This is because cross-path error protection aggregates packets in different paths together to form a stronger FEC code thus providing higher level of error protection. In this chapter, we analyze cross-path FEC in the proposed scheme and demonstrate that the reconstructed video protected by cross-path FEC can have better perceptual quality than the one by individual-path FEC.

As with every resource allocation mechanism, we concern how to allocate bandwidth to each stream. Traditional bandwidth allocation scheme allocates a set of complete packets to each video stream, and multiple video streams are transmitted through Time Division Multiplexing Access (TDMA). Motivated by the better performance provided by cross-path error protection owing to spreading data to more packets, we propose a novel multi-stream cross-path error protection for real-time compressed video over multiple paths based on Packet Division Multiplexing Access (PDMA) [89]. By allowing each packet in all paths carry partial data from different streams, or equivalently, spreading one video bitstream to all packets in all paths, PDMA-based scheme can have a higher level of error protection than TDMA-based scheme when both schemes have the same number of source and parity check bits summed over all paths. When both schemes achieve the same level of error protection given the same bit rate budget, the PDMA-based scheme requires fewer parity check bits and thus allows for more source bits to carry video data to improve video quality than TDMA-based scheme. In this chapter, we will show the performance gain achieved by PDMA-base scheme compared to TDMA-based scheme by first examining effective throughput for generic data and then evaluating the reconstructed video quality for compressed video streams.

6.2 Multi-Source Multi-Path Transmission

In this section, we first describe the scenario where multiple streams are transmitted through multiple paths. We then discuss how to apply cross-path error protection when we transmit a stream through multiple paths. Two different bandwidth allocation schemes, namely, TDMA and PDMA, with cross-path error protection are presented and the corresponding optimal expected capacities after FEC decoding are analyzed at the end of this section.

6.2.1 System Overview

We consider a packet network with erasure channels where all packets have equal size of L bytes. Suppose there exist M independent paths between a sender node and a receiver node, as shown in Figure 6.1. Path m can carry N_m packets per unit time and the packet loss rate is q_m . The total bandwidth from sender node to receiver node is $N = \sum_{m=1}^M N_m$ packets per unit time. There are U video bitstreams to be merged and transmitted from sender node to receiver node. Let n_{um} be the number of packets carrying data of Stream u along Path m per unit time and n_{um} should not be larger than N_m . To facilitate our discussion, we use $\mathbf{n}_u \triangleq [n_{u1}, n_{u2}, \dots, n_{uM}]$ to denote the packet assignment for Stream u in each path. The total number of packets in all M paths to carry data from Stream u is $\sum_{m=1}^M n_{um} = W_u$ packets per unit time.

As discussed in Chapter 5.2.2, an $RS(n, k)$ encoder generates $n - k$ parity symbols for k source symbols. The corresponding RS decoder can recover the original k source symbols if it receives at least k out of n symbols successfully when the locations of the erased symbols are known. For erasure channels, encoding symbols located at the same position across the collected packets with RS code

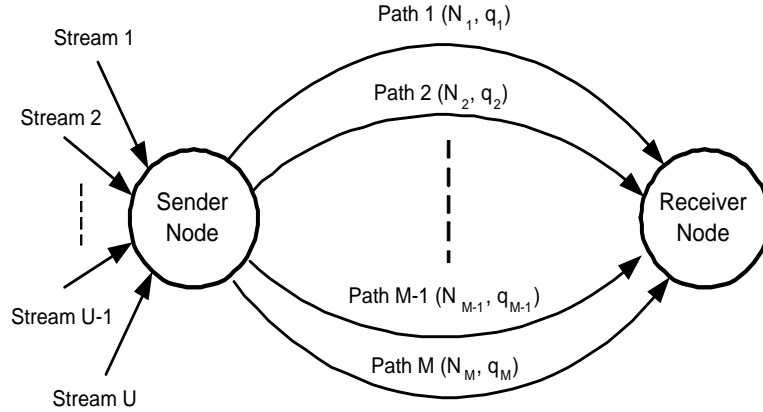


Figure 6.1: Multiple sources over a multiple-path system

can alleviate packet loss problem [105]. The setting of RS code to achieve optimal video quality can be obtained by jointly considering video source R-D and channel conditions. As different paths exhibit heterogeneous channel conditions, we will jointly consider channel conditions of all paths and apply FEC across packets in all paths.

If the receiver receives λ_{um} packets successfully when the sender transmits n_{um} packets for Stream u along Path m , the probability that receiver receives exactly λ packets from all M paths by applying Stream u 's packet assignment \mathbf{n}_u is

$$p(\mathbf{n}_u, \lambda) = \sum_{\sum_{m=1}^M \lambda_{um} = \lambda; \lambda_{um} \leq n_{um}, \forall m} \left(\prod_m \binom{n_{um}}{\lambda_{um}} p_m^{\lambda_{um}} q_m^{n_{um} - \lambda_{um}} \right) \quad (6.1)$$

$$= \sum_{\sum_{m=1}^M \lambda_{um} = \lambda} \left(\prod_m \binom{n_{um}}{\lambda_{um}} p_m^{\lambda_{um}} q_m^{n_{um} - \lambda_{um}} \right), \quad (6.2)$$

where $p_m = 1 - q_m$.

Applying de Moivre-Laplace Theorem [68], we approximate a binomial distri-

bution by a normal distribution:

$$\binom{n_{um}}{\lambda_{um}} p_m^{\lambda_{um}} q_m^{n_{um}-\lambda_{um}} \approx \frac{1}{\sqrt{2\pi\sigma_{um}^2(n_{um})}} e^{-\frac{(\lambda_{um}-\mu_{um}(n_{um}))^2}{2\sigma_{um}^2(n_{um})}}, \quad (6.3)$$

where $\mu_{um}(n_{um}) = n_{um}p_m$ and $\sigma_{um}^2(n_{um}) = n_{um}p_mq_m$.

The sum of M normally distributed independent variables with means $\{\mu_{um}(n_{um})\}$ and variances $\{\sigma_{um}^2(n_{um})\}$ follows a normal distribution with mean $\mu_u(\mathbf{n}_u) = \sum_{m=1}^M \mu_{um}(n_{um})$ and variance $\sigma_u^2(\mathbf{n}_u) = \sum_{m=1}^M \sigma_{um}^2(n_{um})$. Bringing in this normal sum distribution and (6.3), we can approximate (6.1) as

$$p(\mathbf{n}_u, \lambda) \approx \frac{1}{\sqrt{2\pi}\sigma_u(\mathbf{n}_u)} e^{-\frac{(\lambda-\mu_u(\mathbf{n}_u))^2}{2\sigma_u^2(\mathbf{n}_u)}}. \quad (6.4)$$

The probability that the receiver node receives at least k packets from Stream u 's packet assignment \mathbf{n}_u can be expressed as

$$\mathbf{P}(\mathbf{n}_u, k) = \sum_{\lambda=k}^{W_u} p(\mathbf{n}_u, \lambda) \quad (6.5)$$

$$\begin{aligned} &\approx \int_k^{W_u} \frac{1}{\sqrt{2\pi}\sigma_u(\mathbf{n}_u)} e^{-\frac{(t-\mu_u(\mathbf{n}_u))^2}{2\sigma_u^2(\mathbf{n}_u)}} dt \\ &\approx \frac{1}{2} \operatorname{erfc}\left(\frac{k - \mu_u(\mathbf{n}_u)}{\sqrt{2}\sigma_u(\mathbf{n}_u)}\right), \end{aligned} \quad (6.6)$$

where $\operatorname{erfc}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$.

6.2.2 Cross-Path FEC for TDMA and PDMA Scheme

We consider two different cross-path FEC schemes supporting multi-stream aggregation based on TDMA and PDMA, respectively, as illustrated in Figure 6.2. We study the optimal expected throughput achieved by each scheme when we allocate each stream an equal amount of bandwidth, including source and parity check

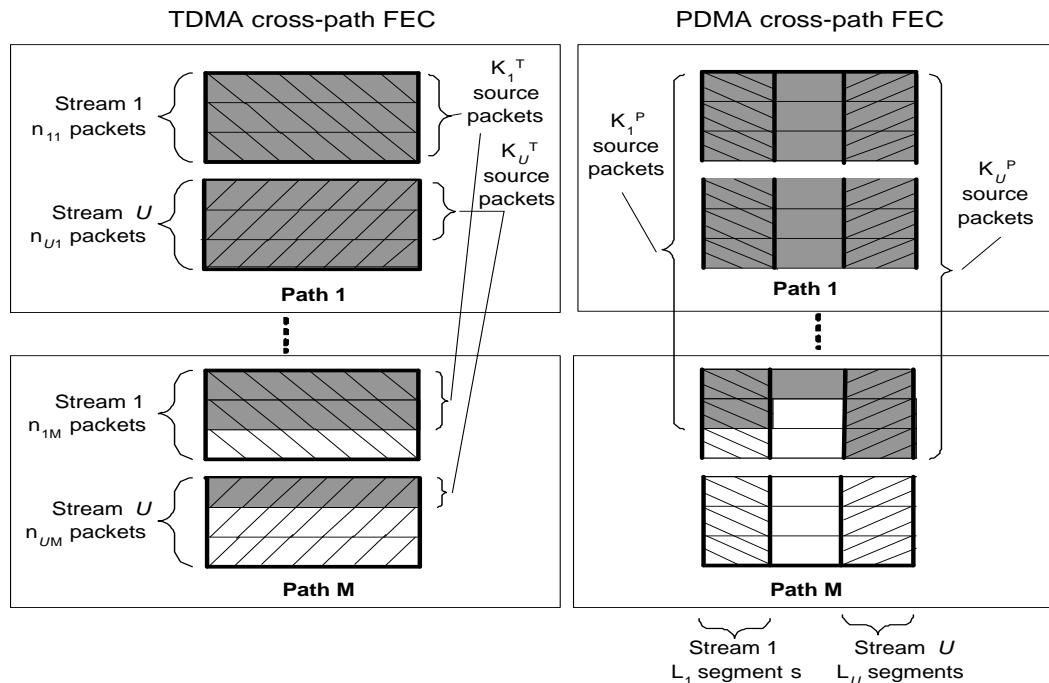


Figure 6.2: Illustration of TDMA and PDMA-based multi-stream error protection. Gray boxes represent source symbols and white ones represent parity symbols. Symbols assigned to different streams are distinguished by different texture patterns.

symbols. This case fits well to the transmission of generic data. We defer the discussion on how to achieve the optimal expected perceptual quality for transmitting video streams to Chapter 6.3.

In the conventional TDMA-based scheme, each packet can only be assigned to at most one stream. To perform cross-path error protection for multiple streams, we need to determine the number of packets assigned to each stream in different paths subject to the constraints on bandwidth and delay. We can further simplify the constraints by merging bandwidth and delay constraints together as the constraint on the maximal number of packets per unit time (according to the maximal

tolerable delay). In each path, packets belonging to different streams are transmitted in a TDMA fashion to meet the constraint of maximal packet number. As an example shown in Figure 6.2, the maximal number of packets per unit time that we can transmit in Path 1 and Path M are N_1 and N_M , respectively. We transmit $n_{11} \leq N_1$ and $n_{1M} \leq N_M$ packets for Stream 1 in Path 1 and Path M , respectively, according to the allocated time slot.

In contrast, the proposed PDMA-based scheme allows each packet to carry data from multiple streams. We apply cross-path error protection for each stream in all packets across all paths, and each stream will be transmitted in a PDMA fashion to satisfy the constraint of maximal packet number in each path. In PDMA-based scheme, we need to determine which portion of a packet is assigned to which stream. To facilitate our discussion, we denote *segment* as the smallest unit that can be assigned to a stream in a packet. We set the size of a segment as the number of bits used in an RS symbol, which is one byte in this work. Denote L_u as the number of segments assigned to Stream u within a packet. As shown in Figure 6.2, we allocate the first L_1 segments of all packets to Stream 1.

In TDMA-based scheme, a packet can be assigned to only one stream, namely, $L_u = L, \forall u$. Thus, $\sum_{u=1}^U n_{um} = N_m, \forall m$. To allocate the same amount of bandwidth to all streams, we assign equal number of packets to each stream in each path, i.e., $n_{um} = N_m/U$. For simplicity in analysis, we assume that the value of N_m is integer multiple of U . The total number of packets in each stream is $W_u = N/U, \forall u$. Let \mathbf{n}_u^T be the corresponding packet assignment and k_u^T the overall number of source packets for Stream u in all paths. The expected throughput using RS(W_u, k_u^T) code for Stream u is

$$f_T(k_u^T) = k_u^T L \cdot \mathbf{P}(\mathbf{n}_u^T, k_u^T). \quad (6.7)$$

Since we only consider throughput in this section, the packet assignment and the corresponding optimal value of k_u^T will be the same for all streams. The optimal RS code to achieve optimal expected throughput can be obtained as follows:

$$k_{opt}^T = \arg \max_{k_u^T} f_T(k_u^T). \quad (6.8)$$

In PDMA-based scheme, a packet can be assigned to multiple streams, namely, $\sum_{u=1}^U L_u = L$. Each stream can allocate data to all packets in each path, namely, $n_{um} = N_m, \forall u, m$, and $W_u = N, \forall u$. To allocate the same amount of bandwidth to all streams, we assign an equal number of segments to each stream in each packet, namely, $L_u = L/U, \forall u$. For simplicity in analysis, we assume that the value of L is an integer multiple of U . Let \mathbf{n}_u^P be the corresponding packet assignment and k_u^P be the overall number of source packets for Stream u in all paths. The expected throughput using RS(W_u, k_u^P) code for Stream u is

$$f_P(k_u^P) = k_u^P \frac{L}{U} \cdot \mathbf{P}(\mathbf{n}_u^P, k_u^P). \quad (6.9)$$

To optimize for throughput, the packet assignment and the corresponding optimal value of k_u^P will be the same for all streams. The optimal RS code to achieve optimal expected throughput can be obtained as follows:

$$k_{opt}^P = \arg \max_{k_u^P} f_P(k_u^P). \quad (6.10)$$

Note that $\mathbf{n}_u^T = \mathbf{n}_u^P/U$. As shown in Appendix A.3, with continuous relaxation of k_u^P and k_u^T and bringing (6.6) into (6.7) and (6.9), we take the derivative of both expected throughputs and can show that the difference between the two optimal solutions can be approximated as:

$$f_P(k_{opt}^P) - f_T(k_{opt}^T) \propto \sigma \sqrt{2 \ln \frac{\mu}{\sigma}} \left(1 - \sqrt{\frac{\ln U}{U}}\right), \quad (6.11)$$

under the condition that

$$\ln \frac{\mu}{\sigma} > \frac{\ln 2\pi}{2} + \frac{\ln U}{2(U-1)}, \quad (6.12)$$

where $\mu = \frac{1}{U} \sum_{m=1}^M N_m p_m$ and $\sigma^2 = \frac{1}{U} \sum_{m=1}^M N_m p_m q_m$.

When there is only one path with bandwidth N and packet loss rate q , the performance gain for PDMA-based scheme over TDMA-based scheme can be approximately proportional to

$$f_P(k_{opt}^P) - f_T(k_{opt}^T) \propto \sqrt{Npq(\ln N + \ln \frac{p}{q})} \cdot (1 - \sqrt{\frac{\ln U}{U}}). \quad (6.13)$$

Figure 6.3 demonstrates the throughput gain for transmitting two streams over single path with different packet loss rates and bandwidths. The performance gain is calculated by solving problem (6.8) and (6.10) via full search. As expected from (6.13), the effective throughput gain increases when the bandwidth (in terms of number of packets) increases and the packet loss rate is fixed. We also observe that when the total bandwidth is fixed, the gain increases when the packet loss rate increases till around 0.5, and then it decreases as the packet loss rate becomes higher. This is because the system needs a larger number of parity check packets to overcome higher packet loss rate. Thus, the allocated number of source packets reduces, and the effective throughput gain decreases.

Figure 6.4 illustrates the performance gain when there are different number of streams in a system where the allocated bandwidth of each stream is $20L$ bytes. The result is calculated by solving problem (6.8) and (6.10) via a full search. As we can see, the performance gain per stream increases when the number of streams increases with fixed packet loss rate. Since the number of packets assigned to each stream in TDMA-based scheme is fixed, the RS code configuration and the corresponding effective throughput for TDMA-based scheme is thus fixed even

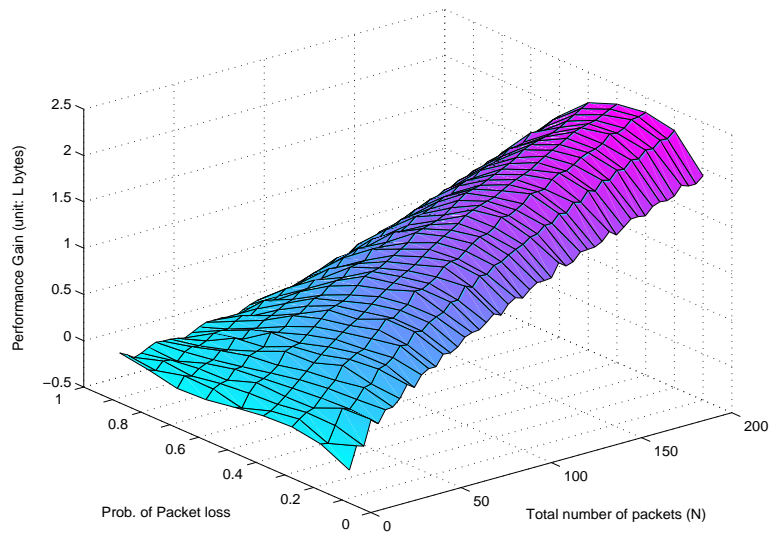


Figure 6.3: Performance gain using PDMA-based scheme over TDMA-based scheme for two-stream case

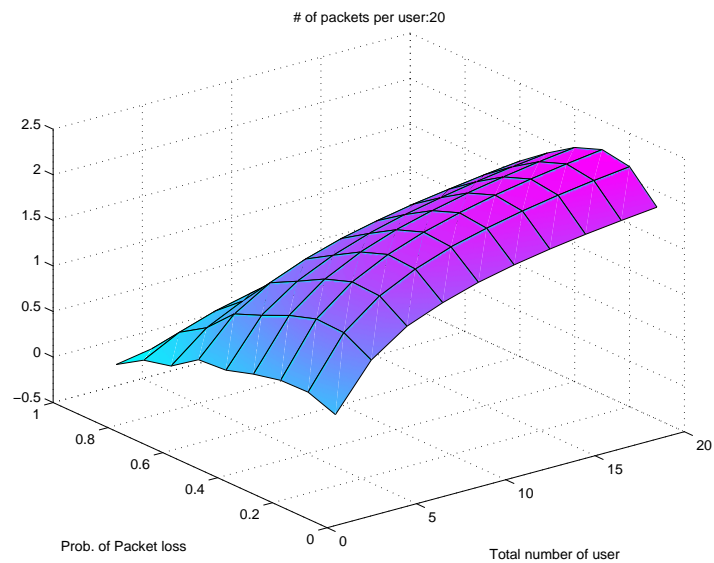


Figure 6.4: Performance gain using PDMA-based scheme over TDMA-based scheme for different number of streams

when the number of streams increases. On the other hand, the PDMA-based scheme can aggregate more packets from different streams to construct a stronger RS code when the number of streams increases. Therefore, with the increase of number of streams in a system, PDMA-based scheme offers higher resilience to packet loss and has an increased performance gain than TDMA-based scheme.

6.3 PDMA Error Protection for Video System

In this section, we first discuss how to apply cross-path PDMA-based FEC to protect multiple video streams and transmit the aggregated stream in real time. We formulate the resource allocation for this multi-stream system as an optimization problem, and propose a fast algorithm to achieve near-optimal solution. We also extend the proposed scheme to the scenario where end users experience different channel conditions in each transmission path.

6.3.1 Multi-Stream Video System

The block diagram of the proposed multi-stream video system is shown in Figure 6.5. There are three main subsystems in the sender aggregation node, namely, a video source coder, a communication module, and a resource allocator. In this work, we use 3D-SPIHT [46] codec as an example for video source coder. Other codecs with similar coding methodologies can also be incorporated. Note that 3D-SPIHT codec needs to collect a group of pictures (GOP) of a total of H frames for temporal/spatial wavelet filtering, so within a video source coder component, each stream has its own frame buffer to gather incoming raw video frames. After collecting a GOP, those raw video frames for each stream are moved to a 3D-SPIHT

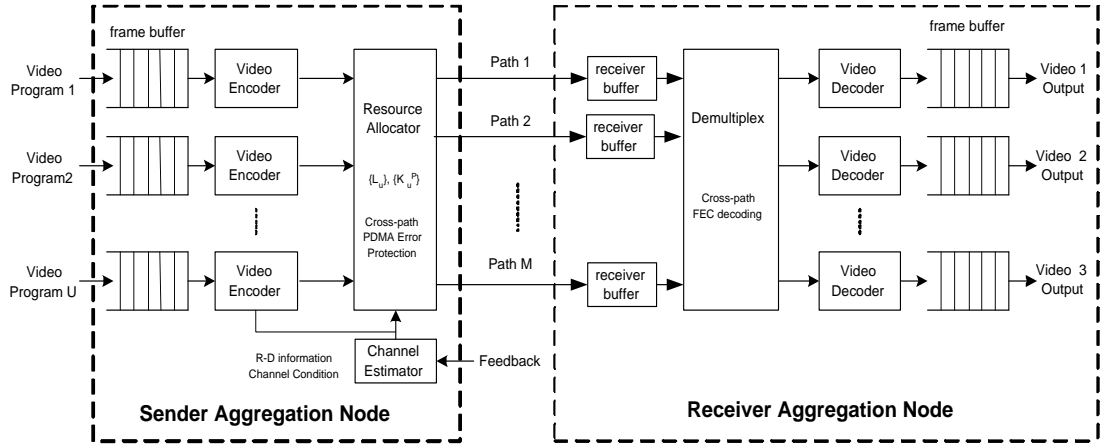


Figure 6.5: Block diagram of multi-stream video system using cross-path PDMA-based error protection

video encoder for compression as an embedded bitstream. The corresponding R-D information for each stream is also analyzed. The resource allocator will determine the truncation point with the required level of error protection for each coded video bitstream and transmit it during the next GOP transmission time of H/F second long, where F is the video frame sampling rate.

The main task of resource allocator is to determine the optimal configuration of cross-path PDMA-based error protection for each GOP. To capture the varying channel conditions and video content characteristics, the resource allocator should obtain the channel information for all M transmission paths fed back from the receiver aggregation node and the R-D information of each GOP from the video coder. With the estimated channel conditions, the resource allocator can calculate the required level of error protection in each transmission path. With the R-D information, the resource allocator can estimate the quality of the reconstructed video frames after decoding at each end decoder. By jointly considering the R-D information and the estimated channel conditions, the resource allocator performs

resource optimization to achieve the required service objectives. According to the allocated resources, the sender aggregation node will truncate video streams, allocate bitstreams in the assigned segments in each packet, and generate the required level of parity check symbols. The communication module will then transmit those coded packets through the assigned transmission paths.

On the receiver aggregation node, a communication module will buffer the received data until the end of the current GOP transmission time. Then, a demultiplexing module will demultiplex different FEC protected bitstreams from each packet according to the allocated segments and perform FEC decoding to obtain the video bitstreams. Those FEC decoded data are moved to a 3D-SPIHT video decoder for decoding and the decoded frames are sent for display.

6.3.2 Problem Formulation

Since the resource allocator applies cross-path PDMA-based error protection and performs optimization for every GOP, we will omit the notation for GOP index in each video stream. Let N_m be the number of packets we can transmit along Path m per GOP time of H/F second and N be the total number of packets summed over all paths per H/F second. We also denote $D_u(r_u)$ as Stream u 's distortion when the receiver receives r_u bytes; $D_u(0)$ as the distortion without decoding any bitstream; and r_u^{max} as the maximal number of bytes in Stream u 's embedded bitstreams. Suppose we allocate L_u bytes for Stream u and apply RS(N, k_u^P) code across packets to protect Stream u . There will be k_u^P source packets and the total bandwidth to transmit video source for Stream u is $k_u^P L_u$ bytes. The corresponding packet assignment \mathbf{n}_u^P equals to $[N_1, N_2, \dots, N_M]$. The expected distortion can be

represented as follows:

$$E^P\{D_u(k_u^P, L_u)\} = D_u(0) - \mathbf{P}(\mathbf{n}_u^P, k_u^P) \cdot (D_u(0) - D_u(k_u^P L_u)), \quad (6.14)$$

where $\mathbf{P}(\mathbf{n}_u^P, k_u^P)$ is defined in (6.5) as the probability that the receiver aggregation node receives at least k_u^P packets when the sender aggregation node sends N packets through all M paths.

The key issue is how to jointly choose the number of bytes, $\{L_u\}$, and the RS code configuration, $\{k_u^P\}$, for each stream. We formulate the problem as to minimize the overall end-to-end distortion of all streams:

$$\min_{\{k_u^P\}, \{L_u\}} \sum_{u=1}^U E^P\{D_u(k_u^P, L_u)\} \quad (6.15)$$

subject to $\left\{ \begin{array}{l} \text{Packet length: } \sum_{u=1}^U L_u \leq L; \\ \text{RS code: } 0 \leq k_u^P \leq N, \forall u; \\ \text{Stream Rate: } 0 \leq k_u^P L_u \leq r_u^{max}, \forall u. \end{array} \right.$

We solve this optimization problem by selecting the RS code, $\{k_u^P\}$, and the number of segments, $\{L_u\}$, for each stream, subject to three constraints: The first constraint limits the overall number of bytes assigned to all streams in a packet not exceed the packet length; the second one restricts the RS code selection; and the third one is the feasible rate range for each video bitstream. The problem (6.15) is an integer programming problem, which is *NP* hard. Finding the optimal solution requires full search with complexity $O(L^U)$, which is prohibited for real-time applications. In a real-time system, a fast approximation algorithm with near-optimal performance is desired.

6.3.3 Proposed Algorithm

We propose a fast, near-optimal algorithm to solve the problem (6.15). There are three main steps, namely, obtaining a segment-to-distortion (S-D) function for each stream, generating the corresponding convex approximation of S-D function, and performing water-filling on all convex S-D functions. We summarize the proposed algorithm in Table 6.1 and examine each step in more details.

STEP 1: Obtain the S-D function for each stream. Given a fixed number of segments, ℓ , for Stream u , we can calculate the optimal number of source packets, $K_u[\ell]$,

$$K_u[\ell] = \arg \min_{k_u^P} E^P \{D_u(k_u^P, \ell)\}, \quad (6.16)$$

to achieve the minimal expected distortion:

$$D_u[\ell] = E^P \{D_u(K_u[\ell], \ell)\}. \quad (6.17)$$

We can calculate the optimal number of source packets and the corresponding distortion for $\ell = 1 \sim L$, arriving at a segment-to-distortion function $D_u[\ell]$.

STEP 2: If the S-D function is convex, the optimal solution can be obtained via the well-known water-filling technique. However, S-D function is not convex in practice. Although we can obtain the optimal solution via full search, the complexity of full search is $O(L^U)$, which is not suitable for real-time application. To facilitate searching near-optimal solution in real time, we should construct a convex function that very well approximates the original S-D function. This can be achieved by finding the convex hull of the original S-D function [7]. We first search the segments that do not maintain convexity. Then, we replace the expected distortion of these segments with the values interpolated linearly from the nearby segments that satisfy convexity. This process gives a convex approximation of S-D function, $\hat{D}_u[\ell]$.

STEP 3: In this step, we perform water-filling among the convex S-D functions to find the solution. We first calculate the distortion reduction for each stream as follows:

$$\Delta\hat{D}_u[\ell] = \hat{D}_u[\ell - 1] - \hat{D}_u[\ell]. \quad (6.18)$$

Then, we sort all streams' $\Delta\hat{D}_u[\ell]$ in a decreasing order. We choose L segments with the largest distortion reduction from $\{\Delta\hat{D}_u[\ell]\}$ and calculate the total number of segments, L_u , assigned to each stream. Given L_u segments for Stream u , we can obtain the optimal RS code configuration, k_u^P , via (6.16). The complexity of Step 1 to 3 are $O(UL)$, $O(UL)$, and $O(L)$, respectively.

6.3.4 Heterogeneous Channel Conditions Along Each Path

In the proposed framework, each video stream is reconstructed by receiving packets from multiple transmission paths. In previous sections, we consider the scenario that transmission paths are the paths between two aggregation nodes. In this subsection, we generalize the proposed framework by extending the concept of transmission path to the end-to-end transmission path experienced by each individual stream between its own video encoder and decoder. For example, each stream can be further forwarded to its own destination after the receiver aggregation point has received the aggregated stream as shown in Figure 6.1. Since different streams are transmitted through different paths at the last hop, each stream experiences different end-to-end channel conditions for each end-to-end transmission path. Another example is the multi-stream transmission service over wireless networks, such as Orthogonal Frequency Division Multiplexing Access (OFDMA) wireless networks, despite of the absence of receiver aggregation point. In this scenario, a subcarrier serves as a transmission path, and each end user can access multiple subcarriers

Table 6.1: Proposed algorithm for cross-path PDMA-based scheme

<p>a) Initialization:</p> $D_u[0] = D_u(0)$
<p>b) Obtain S-D function:</p> <p><i>For each stream</i></p> <p><i>For $\ell = 1 : 1 : L$</i></p> $K_u[\ell] = \arg \min_{k_u^P} E\{D_u(k_u^P, \ell)\}$ $D_u[\ell] = E\{D_u(K_u[\ell], \ell)\}$ $\Delta D_u[\ell] = D_u[\ell - 1] - D_u[\ell]$ <p><i>End</i></p> <p><i>End</i></p>
<p>c) Maintain convexity of S-D function:</p> <p><i>For each stream u</i></p> <p>ConvexFlag = FALSE ; SlopeStart = 1 ; SlopeEnd = 1 ;</p> <p><i>While</i> ConvexFlag == FALSE</p> <p>ChangeFlag = FALSE ;</p> <p><i>For $\ell = 2 : 1 : L$</i></p> <p><i>If</i> $\Delta D_u[\ell] > \Delta D_u[\ell - 1]$</p> <p>SlopeEnd = ℓ; ChangeFlag = TRUE;</p> <p>$S = \text{mean}(\Delta D_u[\text{SlopeStart} : \text{SlopeEnd}])$;</p> <p>$\Delta D_u[\text{SlopeStart} : \text{SlopeEnd}] = S$;</p> <p><i>Else</i></p> <p>SlopeStart = ℓ, SlopeEnd = ℓ;</p> <p><i>End</i></p> <p><i>End</i></p> <p><i>If</i> ChangeFlag == TRUE</p> <p>ConvexFlag = TRUE ;</p> <p><i>End</i></p> <p><i>End</i></p> <p>$\Delta \hat{D}_u[\ell] = \Delta D_u[\ell], \forall \ell$</p> <p><i>End</i></p>
<p>d) Sort all $\Delta \hat{D}_u[\ell]$ in a decreasing order :</p> <p>$\tau(u, \ell) \in \{1, 2, \dots, UL\}$ and $\tau^{-1}(m) \in \{(u, \ell)\}$</p> <p>s.t. $\Delta \hat{D}_i[a] > \Delta \hat{D}_j[b]$ for $\tau(i, a) < \tau(j, b)$</p>
<p>e) Determine parameters :</p> <p><i>For each stream u</i></p> $L_u = \arg \max_{\ell} \tau(u, \ell) \leq L$ $k_u^P = K_u[L_u]$ <p><i>End</i></p>

and will experience heterogeneous channel conditions along each subcarrier.

We can easily extend the proposed framework and solution in previous subsection to the heterogeneous scenario. Denote the packet loss rate for Path m experienced by Stream u as q_{um} . By replacing q_m by q_{mu} in (6.1), the probability that receiver receives exactly λ packets from all M paths by applying Stream u 's packet assignment \mathbf{n}_u^P becomes

$$p_u(\mathbf{n}_u^P, \lambda) = \sum_{\sum_{m=1}^M \lambda_{um} = \lambda} \left(\prod_m \binom{n_{um}}{\lambda_{um}} p_{um}^{\lambda_{um}} q_{um}^{n_{um} - \lambda_{um}} \right), \quad (6.19)$$

where $p_{um} = 1 - q_{um}$. The probability that the end user u receives at least k out of \mathbf{n}_u^P for Stream u is

$$\mathbf{P}_u(\mathbf{n}_u^P, k) = \sum_{\lambda=k}^{W_u} p_u(\mathbf{n}_u^P, \lambda). \quad (6.20)$$

The expected distortion is

$$E^P\{D_u(k_u^P, L_u)\} = D_u(0) - \mathbf{P}_u(\mathbf{n}_u^P, k_u^P) \cdot (D_u(0) - D_u(k_u^P, L_u)). \quad (6.21)$$

The expected distortions (6.14) and (6.21) have the same expression except the probability of receiving at least k out of \mathbf{n}_u^P is different. This new system has the same form of problem formulation as (6.15), and the solutions can be solved using the proposed algorithm presented in Chapter 6.3.3.

6.4 Simulation Results

In this section, we examine the performance of the proposed cross-path PDMA-based error protection scheme. We compare the proposed fast algorithm discussed in Chapter 6.3.3 (PDMA-CP) with three other alternatives. The first one is to employ full search to find the optimal solution for the cross-path PDMA-based scheme

(PDMA-CP-opt). This approach requires a high amount of computation. The second one is the TDMA-based scheme with cross-path error protection (TDMA-CP), the details of which is reviewed in Appendix A.4. The optimal solution of TDMA-CP is found via full search. The last one is the TDMA-based scheme without cross-path error protection (TDMA-EP), where each path can be assigned to at most one video stream. When the number of video streams is equal to the number of transmission paths, we can employ Hungarian method to reduce the computation complexity to obtain the optimal solutions [60].

6.4.1 Simulation Setup

The simulations are set up as follows. We concatenate 15 classic QCIF (176x144) video sequences to form one testing video sequence of 4688 frames. The 15 sequences are 288-frame *News*, 144-frame *Suzie*, 320-frame *Hall objects*, 288-frame *Coastguard*, 368-frame *Carphone*, 400-frame *Foreman*, 288-frame *Container*, 288-frame *Akiyo*, 480-frame *Claire*, 144-frame *Miss American*, 448-frame *Salesman*, 288-frame *Silent*, 144-frame *Trevor*, 400-frame *Grandmother*, and 400-frame *Mother and Daughter*. We use 3D-SPIHT [46] codec to compress video sequence with Group of Picture (GOP) size of 16 frames. The packet length is 1024 bytes.

To simulate the burstness of the channel loss condition, we use the Gilbert-Elliott channel model [20, 22], which is a two-state Markov model at the packet level. We consider transmitting three video programs over three transmission paths. The default bandwidth for Path 1 to 3 is 61.44Kbps, 76.8Kbps, and 92.16Kbps, respectively, and the default packet loss rate for Path 1 to 3 is 0.1, 0.2, and 0.3. A simulation profile is defined as follows: the video content program for each stream is 160-frame long; the first video frame starts from a randomly

selected frame of the concatenated testing video. For each simulation profile, we repeat the simulations 10,000 times for packet loss and take the average to obtain the statistical results.

6.4.2 Performance of Homogeneous Channel Conditions

In this section, we study the performance when all streams experience homogeneous channel conditions along each path. In Figure 6.6, we show one example of the frame-by-frame PSNR for different schemes under the default bandwidth and packet loss rate for each path. The first video frame of Stream 1 to 3 starts from frame 689, 3297, and 4241 from the concatenated testing video sequence, respectively, and the first GOP of each stream starts with video sequence *Hall objects*, *Salesman*, and *Grandmother*, respectively. As we can see, the proposed cross-path PDMA-based scheme has the best PSNR for all streams along the time axis compared to other schemes. Comparing both TDMA-based schemes, TDMA-CP has better PSNR than TDMA-EP for Stream 1 and 2 and slight performance loss for Stream 3.

We list the selected parameters of TDMA-EP, TDMA-CP, and PDMA-CP scheme for the first GOP in Table 6.2, 6.3, and 6.4, respectively. Comparing TDMA-CP and TDMA-EP scheme, we see that both schemes have the same number of bytes allocated to each video sequence, namely, 2048, 2048 and 1024 bytes for Stream 1, 2, and 3, respectively. However, for Stream 1 and 2, the cross-path error protection in TDMA-CP scheme helps provide higher probability to recover video source after suffering from packet loss, as indicated by the higher value of $\mathbf{P}(\mathbf{n}_u^T, k_u^T)$ in Table 6.3 than $\mathbf{P}(N_u, k_u)$ in Table 6.2. Because cross-path error protection scheme distributes each stream to multiple paths and the probability

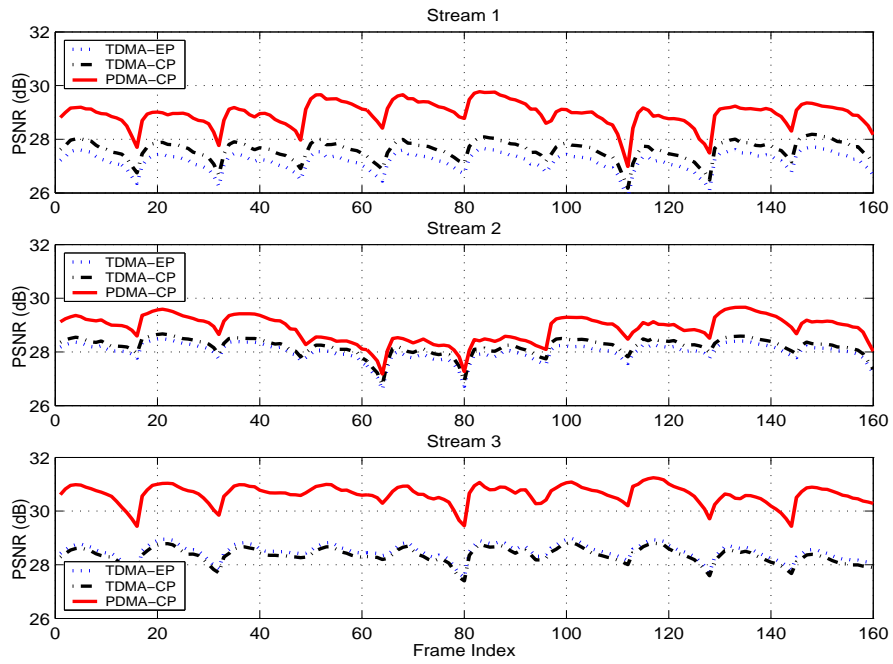


Figure 6.6: Frame-by-frame PSNR for different schemes when bandwidth of Path 1, 2, and 3 is 61.44Kbps, 76.8Kbps, and 92.16Kbps, and the packet loss rate for Path 1 to 3 is 0.1, 0.2, and 0.3.

Table 6.2: Selected parameters of TDMA-EP

Stream ID, u	1	2	3
video sequence	<i>Hall Objects</i>	<i>Salesman</i>	<i>Grandmother</i>
Selected Path	1	2	3
RS code (N_u, k_u)	(4,2)	(5,2)	(6,1)
Prob. RS decoding, $\mathbf{P}(N_u, k_u)$	0.9963	0.9933	0.9993
Number of bytes in source, $k_u L$	2048	2048	1024
Expected MSE	598	206	102

Table 6.3: Selected parameters of TDMA-CP

Stream ID, u	1	2	3
video sequence	<i>Hall Objects</i>	<i>Salesman</i>	<i>Grandmother</i>
number of packets in Path 1, n_{u1}	2	0	2
number of packets in Path 2, n_{u2}	2	3	0
number of packets in Path 3, n_{u3}	2	3	1
number of total packets, W_u	6	6	3
number of source packets, k_u^T	2	2	1
RS code (W_u, k_u^T)	(6,2)	(6,2)	(3,1)
Prob. RS decoding, $\mathbf{P}(\mathbf{n}_u^T, k_u^T)$	0.9989	0.9957	0.9970
Number of bytes in source, $k_u^T L$	2048	2048	1024
Expected MSE, $E^T\{D_u(\mathbf{n}_u^T, k_u^T)\}$	197	147	143

of simultaneous packet loss in all paths is lower, cross-path error protection has stronger error protection. Thus, TDMA-CP scheme can provide better expected video quality for Stream 1 and 2 compared to TDMA-EP scheme. To achieve the optimal overall video quality summed over all streams, TDMA-CP trades in a little weaker code than TDMA-EP for Stream 3 (thus a little performance loss) to obtain larger performance improvement in the overall video quality.

Next, we compare PDMA-CP and TDMA-CP scheme. As revealed from Table 6.4, PDMA-CP scheme allocates 2526, 2541, and 1680 bytes to Stream 1, 2, and 3,

Table 6.4: Selected parameters of PDMA-CP

Stream ID, u	1	2	3
video sequence	<i>Hall Objects</i>	<i>Salesman</i>	<i>Grandmother</i>
number of segments, L_u	421	363	240
number of source packets, k_u^P	6	7	7
RS code (N, k_u^P)	(15,6)	(15,7)	(15,7)
Prob. RS decoding, $\mathbf{P}(\mathbf{n}_u^P, k_u^P)$	0.9999	0.999	0.999
Number of bytes in source, $k_u^P L_u$	2526	2541	1680
Expected MSE, $E^P\{D_u(k_u^P, L_u)\}$	97	85	81

respectively. Compared to TDMA-CP scheme, PDMA-CP scheme provides higher bitrate for video source in all paths. In addition, PDMA-CP scheme can provide higher probability to recover video source than TDMA-CP scheme for all streams, as shown by the higher value of $\mathbf{P}(\mathbf{n}_u^P, k_u^P)$ in Table 6.4 than $\mathbf{P}(\mathbf{n}_u^T, k_u^T)$ in Table 6.3. This is because PDMA-CP scheme can allocate each stream's data to all packets in all paths so that the path diversity can be fully explored. Owing to higher source bitrates with stronger error protection, PDMA-CP scheme can achieve the best performance among all these three schemes.

To evaluate the performance for the proposed scheme under different packet loss rate, we vary the packet loss rate of Path 2 from 0.05 to 0.4. To further assess the performance under diverse video content, we run 50 simulation profiles for each packet loss rate setting and the starting frame of each stream is randomly chosen from the concatenated testing sequence for each simulation profile. Figure 6.7 shows the average received PSNR of all streams over 50 simulation profiles for different packet loss rate of Path 2. As we can see, all cross-path error protection schemes can have better performance compared to the one without cross-path protection. This demonstrates the benefit of using cross-path error protection.

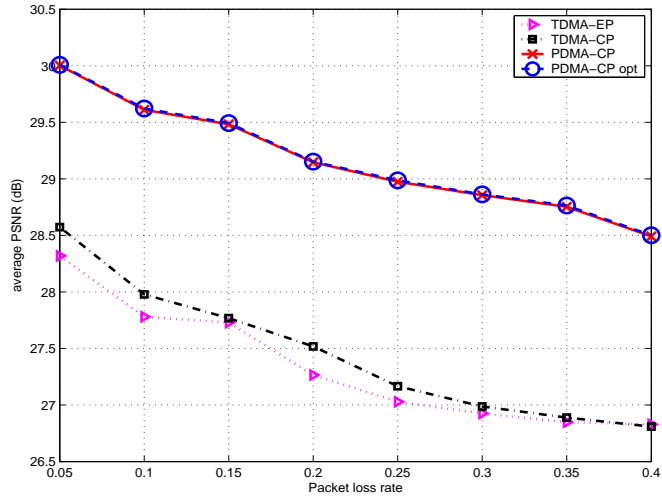


Figure 6.7: Performance comparison of the TDMA-CP, and proposed PDMA-CP, and the corresponding optimal PDMA-CP-opt with full search for different packet loss rate

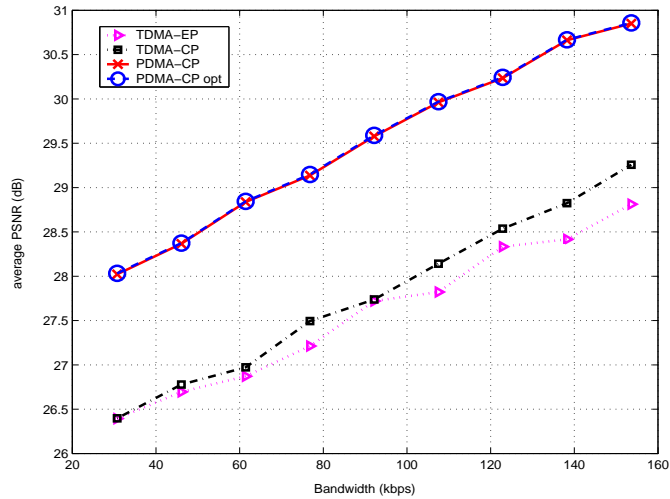


Figure 6.8: Performance comparison of the TDMA-CP, and proposed PDMA-CP, and the corresponding optimal PDMA-CP-opt with full search for different bandwidth

Between the two cross-path error protection schemes providing optimal solutions, the full search result of PDMA-based scheme, PDMA-CP-opt, outperforms optimal TDMA-based scheme, TDMA-CP, by 1.43 ~ 1.88 dB. The performance gain increases when the packet loss rate increases. This verifies the throughput gain analyzed in Chapter 6.2.2. Compared to the full search result of PDMA-based scheme, the proposed fast algorithm of PDMA-based scheme has performance loss no more than 0.01dB. Moreover, PDMA-CP requires much lower computation than PDMA-CP-opt algorithm.

We also examine the performance under different bandwidth provided by the transmission paths. Figure 6.8 shows the average received PSNR of all streams over 50 simulation profiles when we vary the bandwidth of Path 2 from 30.72Kbps to 153.6Kbps. Again, all cross-path error protection schemes can have better performance compared to TDMA-EP. Between the two cross-path error protection schemes providing optimal solutions, PDMA-CP-opt outperforms TDMA-CP by 1.59 ~ 1.87 dB. We also observe that PDMA-CP scheme has only 0.01dB performance loss compared to the optimal PDMA-based solution.

6.4.3 Performance of Heterogeneous Channel Conditions

In this subsection, we evaluate the performance of cross-path PDMA-based error protection under the scenario that end users experience heterogeneous channel conditions even along the same path, as discussed in Chapter 6.3.4. We fix the bandwidth of Path 1 and 3 as 61.44Kbps and 92.16Kbps, respectively, and vary the bandwidth of Path 2 from 30.72Kbps to 153.6Kbps. To simulate the heterogeneity of channel conditions experienced by different end users, the packet loss rate for each transmission path experienced by each end user is randomly chosen within the

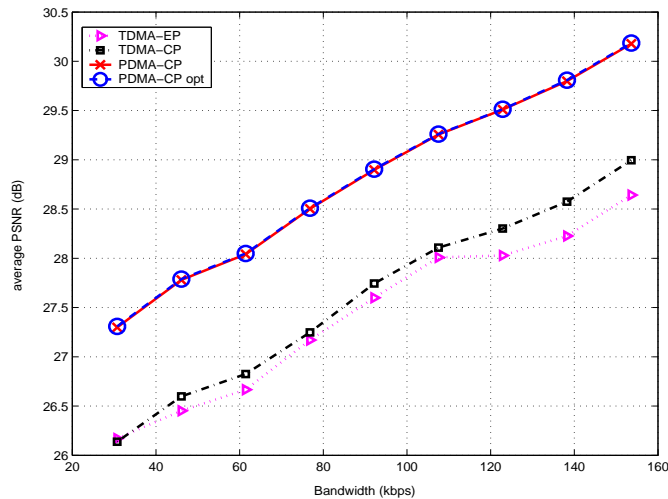


Figure 6.9: Performance comparison for heterogeneous channel conditions of the TDMA-CP, and proposed PDMA-CP, and the corresponding optimal PDMA-CP-opt with full search for different bandwidth

range of 0.1 to 0.4 in each simulation profile. To further examine the performance of transmitting different video programs, we simulate 100 simulation profiles for each bandwidth setting and the starting frame of each stream is randomly chosen from the testing sequence for each simulation profile.

Figure 6.9 shows the average received PSNR of all streams over 100 simulation profiles. Similar to the performance in Figure 6.8, cross-path error protection can provide better performance than individual-path error protection. When we compare the perceptual quality provided by the proposed PDMA-CP scheme with TDMA-CP scheme, we can have 1.15 dB to 1.26 dB performance gain. Even though different streams experience different packet loss rates, by given a bandwidth budget constraint, the proposed PDMA-CP scheme spreads data to all packets along all paths to fully explore path diversity for higher perceptual quality.

6.5 Chapter Summary

In this chapter, we have proposed a cross-path PDMA-based error protection scheme to transmit multiple streams over multiple error-prone paths which exhibit heterogeneous channel conditions. The throughput gain using PDMA-based vs. TDMA-based scheme is analyzed. We have shown that PDMA-based scheme can provide higher effective bandwidth with stronger error protection than TDMA-based scheme. The proposed multiuser video streaming system jointly explores the path diversity and video content diversity. We formulate this video system as an optimization problem to minimize the overall distortion summed over all streams, subject to the constraints on bandwidth, delay, FEC code selection, and video coding rate. To satisfy the real-time requirement, we propose a fast algorithm, providing near-optimal solution.

The simulation results demonstrate that the proposed cross-path PDMA-based scheme outperforms traditional TDMA-based scheme by 1.43~1.88 dB for the averaged PSNR of all received streams when all end receivers experience homogeneous channel conditions along each path and 1.15~1.26 dB when all end receivers experience heterogeneous channel conditions along each path. In addition, the proposed fast algorithm has only 0.01dB performance loss compared to the optimal solution obtained through full search. The proposed scheme can further incorporate unequal error protection to explore the unequal importance within each embedded bitstream. Thus, the proposed scheme is a promising framework supporting real-time transmission for multiple video streams over multiple transmission paths, such as multi-point video conferencing and digital video surveillance.

Chapter 7

Conclusions and Perspectives

In this dissertation, we have discussed the real-time multiuser video transmission over resource-limited communication networks from a resource allocation point of view. Different from the traditional approaches, we focus on two aspects of design issues: cross-layer optimization and multiuser diversity. A general resource-allocation framework is presented to address how to control the quality of multiple video bitstreams that share the limited resources in the communication networks. Based on the proposed framework, we study the available resources for allocation and the diversity of resources that can be explored for performance improvement. Realizing the inter-constrained usage of resources among users and across layers in real time, we present a general design principle to formulate systems as constrained optimization problems and derive solutions for real-time applications.

Several systems transmitting video programs over 3G, 4G, WLAN/WMAN communication channels and the corresponding algorithms for real-time applications are proposed in this dissertation. We have applied the proposed resource allocation methodology to several scenarios. We consider the scenario with error-free channel in Chapter 3 and the scenario where BER after FEC decoding is

sufficiently low such that the end-to-end video quality is controlled by the source coding rate in Chapter 4. We also examine the scenarios where end-to-end video distortion is jointly determined by source lossy compression and channel induced errors in Chapter 5 and Chapter 6. The simulation results demonstrate that the proposed multiuser cross-layer design with dynamic resource allocation can provide better video quality than the traditional schemes.

Instead of achieving the highest average perceptual quality for one user as targeted by the traditional single-user video transmission system, we discuss the importance of addressing service objectives in two levels, namely, network level and individual user level. We investigate the resulting video quality received by each end user and point out the significance of having tradeoffs among multiple service objectives. The corresponding real-time algorithms to achieve the desired tradeoffs are also proposed. The simulation results demonstrate that the tradeoffs among several service objectives obtained via the proposed framework can provide more satisfactory video quality than the system targeting at only one service objective.

We have studied the design principles of multi-user cross-layer video transmission system in this dissertation. The general principle can be applied to other scenarios and applications. Here are several interesting perspectives that can be pursued further.

1. Multi-point video conferencing Multi-point video conferencing, which involves multiple conferees and realizes a virtual conference room, is one of the promising applications. However, there are still many design challenges remaining in a multi-point video conferencing system. First, each conferee transmits his/her real-time compressed video stream through a resource-limited uplink as the first hop and receives multiple streams containing all other conferees in time through

a resource-limited downlink as the last hop. The network resources among all intermediate nodes are also resource-limited. How to allocate resources to all participants in all hops is an important issue. Second, the channel conditions in different hops are heterogeneous. The optimal error protection for different stream along the same hop may not be the same. An effective approach should explore the diversity of video streams and jointly perform error protection for those streams subject to limited network resources. Third, a multi-point video conferencing requires real-time streaming and a strict delay constraint is imposed to each stream to maintain the interactivity within a conference session. Simply applying the existing single-user streaming methods or multicasting would not be efficient.

A possible solution is to deploy video stream combiners in a distributed manner to support this multi-point video conferencing. Video stream combiners, which are located in different geographical areas and serve as portals for conferees, aggregate incoming streams supplied by local users along with another aggregated streams from nearby video stream combiners. In addition, users may have different preferences for the incoming video streams and may want to have better quality for some streams than others. The video combiners should take those preferences into account when they perform multi-stream aggregation. We can also extend this technology developed for multiple-point video conferencing system to ad-hoc networks, such as military applications, with considerations of time-varying channel conditions, radio interferences caused by all other users. For mobile ad-hoc network, we also need to consider how to maintain and update the transmission paths among all participants owing to dynamic topology.

2. Cross-Subcarrier PDMA error protection over OFDMA Networks

As demonstrated in Chapter 6, the proposed cross-path PDMA-based error protection can provide better perceptual quality than the traditional schemes even when each user experiences different channel condition along each path. We can apply the proposed cross-path PDMA-based error protection to OFDMA networks as discussed in Chapter 4, since each subcarrier can be viewed as a transmission path. In this case, it is important to study how to employ cross-subcarrier PDMA error protection in the physical layer for multiple users by jointly considering heterogeneous channel conditions of different subcarriers. We are also interested in what kind of conditions that offers performance gain compared to the traditional schemes. To improve spectrum utilization, adaptive modulation and channel coding in the physical layer can be further incorporated into the cross-subcarrier PDMA error protection. We would need to address how to apply cross-subcarrier error protection in the application layer for packets with different payload size, and how to derive fast and near-optimal solutions to solve this integer programming problem for real-time applications.

3. Distributed video telephony The current 3G wireless systems, such as CDMA2000 1xEV-DO [21], do not have fully centralized mechanism for multiuser cross-layer R-D optimization. There are limited amount of bandwidth providing side information, such as power control and R-D information, which can be exchanged between the base stations and mobile users to facilitate video quality control in the network level. A research issue is how to use such a limited amount of side information to provide satisfactory quality of service. Under this scenario, each user needs to adjust the parameters in both video source and wireless channel in a distributed manner owing to the lack of central authority. Pricing mechanism

or game theoretic approach are potential solutions.

Moreover, for terminal-to-terminal applications through current 3G networks, such as interactive video telephony, the video streams often are transmitted through hybrid networks, namely, wireless and wired networks. These two types of networks have different impacts on the received end-to-end video quality. It is desirable to have an effective way to differentiate these two different impacts and overcome the quality degradation from both networks. Potential solutions include equipping proxies at the edges of networks for transcoding and sending feedback information from the receiver to facilitate rate adjustment in the source coder.

Appendix A

Proofs and Review

A.1 The Impact of Budget Factor on Perceptual Quality

In this section, we present detailed rationale behind the bandwidth resource allocation algorithm proposed in Section 3.3.4. In particular, for a video scene consisting of similar R-D characteristics, we analyze the trend of the *aveMSE* and the *madMSE* as β changes, and derive the result of (3.14) for β_0 .

Consider an M -frame video clip with similar visual contents, whereby the rate-distortion model of each frame within the clip are similar to each other. As such, it is reasonable to assume that the feasible range of FGS data rates for all frames are within the same bit plane, and $R_j^p \approx r f_{j-1}$. The rate-distortion model can be expressed as follows for rates falling in the range of interest:

$$D_j(r_j) = k_1 - k_2 r f_j, \text{ for all } j \tag{A.1}$$

where k_1 and k_2 are constants. We denote \overline{rb} and \overline{rf} as the average rate of base layer and FGS layer within this video sequence, respectively. Here \overline{rb} is fixed due

to the use of constant quantization step for the base layer of all M frames.

Let us first consider $madMSE$, which can be represented as follows using (A.1):

$$madMSE = \frac{k_2}{M-1} \sum_{j=2}^M |\Delta r f_j|. \quad (\text{A.2})$$

where $\Delta r f_j = r f_j - r f_{j-1}$. We examine the absolute difference of FGS rates between two frames at time slot j :

$$\begin{aligned} |\Delta r f_j| &= |(1 - w_p)\beta[(C_j - C_{j-1}) \\ &+ (B_{j-2}^e - B_{j-1}^e) + (r b_{j-1} - r b_j)] + w_p \Delta r f_{j-1}|. \end{aligned} \quad (\text{A.3})$$

With a fixed w_p , a larger β would usually result in larger $|\Delta r f_j|$ hence larger $madMSE$ by (A.2).

Next, we consider $aveMSE$. With the R-D model in (A.1), we can express $aveMSE$ as follows:

$$aveMSE = k_1 - k_2 \overline{r f}. \quad (\text{A.4})$$

Summing up from $r f_1$ to $r f_M$ using (3.11) and (3.12) and taking the average, we obtain

$$\overline{r f} = w_p \overline{R^p} + (1 - w_p)\beta (B_{max}^e + \overline{C} - \overline{B}^e - \overline{r b}), \quad (\text{A.5})$$

where $\overline{R^p}$, \overline{C} , \overline{B}^e represent the average of their corresponding rates or buffer occupancy. Bringing $\overline{R^p} \approx \overline{r f}$, the condition of encoder buffer occupancy $\overline{B}^e \approx 0$, and the conservation law of data flow that the overall input flow should be equal to the overall output flow:

$$\overline{C} = \overline{r b} + \overline{r f}, \quad (\text{A.6})$$

we arrive at

$$\overline{r f} \approx \beta B_{max}^e / (1 - \beta). \quad (\text{A.7})$$

Taking (A.7) into (A.4), we can represent $aveMSE$ as a function of β :

$$aveMSE \approx k_1 - k_2\beta B_{max}^e/(1 - \beta). \quad (A.8)$$

Thus if the system has a larger β , the average distortion ($aveMSE$) will decrease.

Note that w_p does not affect $aveMSE$ as long as $0 \leq w_p < 1$.

The above analysis shows that a larger β reduces $aveMSE$ but leads to larger $madMSE$. To complete the derivation for β_0 , we observe from (A.6) that since \overline{rb} is fixed, increasing \overline{C} is equivalent to increasing \overline{rf} . But when \overline{C} approaches the channel capacity C_{max} , \overline{rf} and $aveMSE$ cannot be further improved. Thus there exists β_0 such that its corresponding $\overline{rf}(\beta_0)$ from (A.7) is equal to $C_{max} - \overline{rb}$, i.e.

$$\beta_0 B_{max}^e/(1 - \beta_0) = C_{max} - \overline{rb}. \quad (A.9)$$

Recalling the results in (A.2) and (A.8), the selection of $\beta = \beta_0$ can give an excellent trade-off between $aveMSE$ and $madMSE$. Solving (A.9) for β_0 , we arrive at the trade-off point:

$$\beta_0 = \frac{C_{max} - \overline{rb}}{B_{max}^e + C_{max} - \overline{rb}}. \quad (A.10)$$

A.2 Review of An Alternative Sliding-Window Algorithm

The sliding-window algorithm is originally proposed in [124] and only concerns the channel capacity without considering the buffer and delay constraints. To determine the bit rate for a frame, the sliding window algorithm requires the complete bit rate and R-D information of L frames ahead. The algorithm then distributes the FGS rates to the current frame by solving an optimization problem that all frames within the look-ahead window have the uniform and highest possible

quality subject to a FGS rate budget for all frames in this sliding window, denoted W_j^F . This FGS rate budget is obtained by subtracting all base layer rates within the window from an overall rate budget, W_j . The rate budget W_j is updated by removing the bandwidth used for the previous frame and adding the currently available channel transmission rate. That is,

$$W_j = W_{j-1} - r_{j-1} + C_j, \quad (\text{A.11})$$

$$W_j^F = W_j - \sum_{k=j}^{j+L-1} rb_k, \quad (\text{A.12})$$

For a fair comparison with our proposed algorithms, we modify the sliding window approach by adding the delay and encoder/decoder buffer constraints. We also make two further modifications to fit in the scenarios considered in this paper.

The first modification is on W_j^F . When the sliding window does not cross the scene boundary, it is reasonable to assume that the frames within the sliding window have similar R-D characteristics. Under this assumption, the estimated total transmission rate for these L frames in the sliding window is $L \cdot C_j^{(U)}$, where $C_j^{(U)}$ is the upper bound in (3.7). To keep the occupancy of the encoder buffer low, the data left in the encoder buffer at time slot $j - 1$ is flushed out during the next L time slots. Thus, the modified FGS rate budget for a sliding window is

$$W_j^F = L \cdot C_j^{(U)} - B_{j-1}^e - \sum_{k=j}^{j+L-1} rb_k. \quad (\text{A.13})$$

Second, we observe that when a sliding window enters a segment with simple R-D characteristics from a past segment with complex R-D characteristics, the encoder buffer may overflow and force the system to drop more FGS layer data, which leads to several severe quality fluctuations near the R-D characteristics dissimilarity boundaries (which often coincide with scene changes). To overcome this problem and allow fair comparison with our proposed schemes, we detect the

change in R-D characteristics once it appears in the sliding window and adaptively reduce the window size so that the sliding window does not cross this dissimilarity boundary. After passing the boundary, the size of the sliding window is restored to the original size.

A.3 Performance Gain for PDMA over TDMA

In this section, we show the performance gain (6.11) for PDMA-based scheme vs. TDMA-based scheme and the gain region (6.12) discussed in Chapter 6.

We first consider TDMA-based scheme. To simplify our notation, denote $\mu^T = \mu(\mathbf{n}_u^T)$ and $\sigma^T = \sigma_u(\mathbf{n}_u^T)$. Bringing in (6.6) into (6.7), the expected rate of TDMA-based scheme can be expressed as follows:

$$f_T(k^T) \approx \frac{k^T L}{2} \operatorname{erfc}\left(\frac{k^T - \mu^T}{\sqrt{2}\sigma^T}\right). \quad (\text{A.14})$$

Let $\tilde{k}^T = \frac{k^T - \mu^T}{\sqrt{2}\sigma^T}$. With continuous relaxation for k^T , the optimal k^T can be found by taking the derivative of $f_T(k^T)$ with respect to k^T :

$$\frac{df_T(k^T)}{dk^T} = \frac{L}{2} \operatorname{erfc}(\tilde{k}^T) - \frac{k^T L}{\sqrt{\pi}} (e^{-\tilde{k}^2}) \left(\frac{1}{\sqrt{2}\sigma^T}\right) = 0 \quad (\text{A.15})$$

Without loss of generality, we can transform the optimal solution k_{opt}^T into the following form:

$$k_{opt}^T = \mu^T - \phi_{opt}^T \sigma^T. \quad (\text{A.16})$$

where ϕ_{opt}^T is the value to be determined. Note that

$$\tilde{k}_{opt}^T = -\frac{\phi_{opt}^T}{\sqrt{2}}. \quad (\text{A.17})$$

Rearrange (A.15) and bring in (A.16) and (A.17), we arrive at

$$\phi_{opt}^T + \frac{\sqrt{\pi}}{\sqrt{2}} \operatorname{erfc}\left(-\frac{\phi_{opt}^T}{\sqrt{2}}\right) e^{\frac{(\phi_{opt}^T)^2}{2}} = \frac{\mu^T}{\sigma^T}. \quad (\text{A.18})$$

We assume $\phi_{opt}^T > 1$ and thus $\text{erfc}(-\frac{\phi_{opt}^T}{\sqrt{2}}) \approx 2$. Then, the second exponential term dominates the LHS of (A.18) and we can approximate (A.18) as

$$\sqrt{2\pi}e^{\frac{(\phi_{opt}^T)^2}{2}} \approx \frac{\mu^T}{\sigma^T}. \quad (\text{A.19})$$

Thus, ϕ_{opt}^T can be solved approximately by rearranging (A.19) as follows:

$$\phi_{opt}^T = \sqrt{2 \ln \frac{\mu^T}{\sigma^T} - \ln 2\pi} \quad (\text{A.20})$$

Bringing (A.16) into (A.14), the expected rate of TDMA-based scheme can be represented as:

$$f_T(k_{opt}^T) \approx L \frac{\mu^T - \phi_{opt}^T \sigma^T}{2} \text{erfc}\left(\frac{-\phi_{opt}^T}{\sqrt{2}}\right). \quad (\text{A.21})$$

Similarly, for PDMA-based scheme, denote $\mu^P = \mu(\mathbf{n}_u^P)$ and $\sigma^P = \sigma_u(\mathbf{n}_u^P)$. The expected rate is:

$$f_P(k^P) = k^P \frac{L}{2U} \cdot \text{erfc}\left(\frac{k^P - \mu^P}{\sqrt{2}\sigma^P}\right). \quad (\text{A.22})$$

By applying the same technique used in TMDA-based scheme, we transform the optimal k_{opt}^P into $\mu^P - \phi_{opt}^P$. We can approximate ϕ_{opt}^P as:

$$\phi_{opt}^P = \sqrt{2 \ln \frac{\mu^P}{\sigma^P} - \ln 2\pi} \quad (\text{A.23})$$

The expected rate of PDMA-based scheme can be represented as:

$$f_P(k_{opt}^P) \approx \frac{L}{U} \frac{\mu^P - \phi_{opt}^P \sigma^P}{2} \text{erfc}\left(\frac{-\phi_{opt}^P}{\sqrt{2}}\right). \quad (\text{A.24})$$

Note that $\mu^P = U\mu^T$ and $(\sigma^P)^2 = U(\sigma^T)^2$. Thus, $\phi_{opt}^P > \phi_{opt}^T$.

The performance gain for PDMA-based scheme over TDMA-based scheme can

be approximated as follows:

$$\begin{aligned}
f_P(k_{opt}^P) - f_T(k_{opt}^T) &\approx \frac{L}{U} \frac{\mu^P - \phi_{opt}^P \sigma^P}{2} \operatorname{erfc}\left(\frac{-\phi_{opt}^P}{\sqrt{2}}\right) - L \frac{\mu^T - \phi_{opt}^T \sigma^T}{2} \operatorname{erfc}\left(\frac{-\phi_{opt}^T}{\sqrt{2}}\right) \\
&\geq \frac{L}{U} \frac{\mu^P - \phi_{opt}^P \sigma^P}{2} \operatorname{erfc}\left(\frac{-\phi_{opt}^P}{\sqrt{2}}\right) - L \frac{\mu^T - \phi_{opt}^T \sigma^T}{2} \operatorname{erfc}\left(\frac{-\phi_{opt}^P}{\sqrt{2}}\right) \\
&= \frac{L}{2} \sigma^T \left(\phi_{opt}^T - \frac{\phi_{opt}^P}{\sqrt{U}}\right) \operatorname{erfc}\left(\frac{-\phi_{opt}^P}{\sqrt{2}}\right) \\
&\approx L \sigma^T \left(\phi_{opt}^T - \frac{\phi_{opt}^P}{\sqrt{U}}\right) \\
&\propto \sigma^T \sqrt{2 \ln \frac{\mu^T}{\sigma^T}} \left(1 - \sqrt{\frac{\ln U}{U}}\right), \tag{A.25}
\end{aligned}$$

The performance gain holds when (A.25) is greater than zero. In other words,

$$\phi_{opt}^T - \frac{\phi_{opt}^P}{\sqrt{U}} > 0 \tag{A.26}$$

Bringing in (A.20) and (A.23) to (A.26), we arrive at the region where PDMA has a performance over TDMA:

$$\ln \frac{\mu}{\sigma} > \frac{\ln 2\pi}{2} + \frac{\ln U}{2(U-1)}, \tag{A.27}$$

where $\mu = \frac{1}{U} \sum_{m=1}^M N_m p_m$ and $\sigma^2 = \frac{1}{U} \sum_{m=1}^M N_m p_m q_m$.

A.4 Problem Formulation of TDMA-CP

In this section, we briefly describe the cross-path TDMA-based error protection for multi-stream video system. Suppose we adopt \mathbf{n}_u^T packet assignment for stream u and thus each stream can distribute video bitstreams into W_u packets. Then we apply $\text{RS}(W_u, k_u^T)$ code across packets to protect stream u . There will be k_u^T source packets and the total bandwidth to transmit video source for stream u is $k_u^T L$ bytes. The expected distortion can be represented as follows:

$$E^T \{D_u(\mathbf{n}_u^T, k_u^T)\} = D_u(0) - \mathbf{P}(\mathbf{n}_u^T, k_u^T) \cdot (D_u(0) - D_u(k_u^T L)), \tag{A.28}$$

In this cross-path TDMA-based scheme, we need to jointly determine the packet assignment, $\{\mathbf{n}_u^T\}$, and the RS code configuration, $\{k_u^T\}$, for each stream. We formulate the problem as to minimize the overall end-to-end distortion of all streams:

$$\min_{\{\mathbf{n}_u^T\}, \{k_u^T\}} \sum_{u=1}^U E^T \{D_u(\mathbf{n}_u^T, k_u^T)\} \quad (\text{A.29})$$

$$\text{subject to } \begin{cases} \text{Bandwidth: } \sum_{u=1}^U n_{um} \leq N_m, \forall m; \\ \text{RS code: } 0 \leq k_u^T \leq \sum_{m=1}^M n_{um}, \forall u; \\ \text{Stream Rate: } 0 \leq k_u^T L \leq r_u^{\max}, \forall u. \end{cases}$$

There are three system constraints imposed in the formulated problem: The first constraint limits the overall number of packets assigned to all streams in each path not exceed the maximal bandwidth limitation and delay requirement; the second one restricts the RS code selection; and the third one is the feasible rate range for each video bitstream. Note that the transmission delay constraint is H/F second, which is implicitly imposed by the maximal number of packets (N_m) we can send along Path m . The problem (A.29) is also an integer programming problem, which is NP hard. To compare the performance with PDMA-based scheme, we perform full search to find the optimal solutions. As presented in Section 6.4, the optimal solution of problem (A.29) obtained via full search is worse than the near-optimal solution of problem (6.15).

BIBLIOGRAPHY

- [1] M. H. Ahmed, H. Yanikomeroglu, and S. Mahmoud. Fairness enhancement of link adaptation techniques in wireless networks. In *Proc. IEEE Vehicular Technology Conference*, volume 4, pages 1554–1557, October 2003.
- [2] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan. Priority encoding transmission. *IEEE Trans. on Information Theory*, 42(6):1737–1744, November 1996.
- [3] J. G. Apostolopoulos. Reliable video communications over lossy packet networks using multiple state encoding and path diversity. In *SPIE Conf. Visual Comm. & Image Processing*, pages 392–409, January 2001.
- [4] J.G. Apostolopoulos and S.J. Wee. Unbalanced multiple description video communication using path diversity. In *IEEE International Conference on Image Processing*, volume 31, pages 966–969, October 2001.
- [5] R. Aravind, M. R. Civanlar, and A. R. Reibman. Packet loss resilience of MPEG-2 scalable video coding algorithms. *IEEE Trans. on Circuits and Systems for Video Technology*, 6(5):426–435, October 1996.
- [6] O. Awoniyi, N.B. Mehta, and L.J. Greenstein. Characterizing the orthogonality factor in WCDMA downlinks. *IEEE Trans. on Wireless Communications*, 2(4):621–625, July 2003.
- [7] M.S. Bazaraa, H.D. Sherali, and C.M. Shetty. *Nonlinear Programming Theory and Algorithms*. Wiley, 1993.
- [8] J. Beran, R. Sherman, M.S. Taqqu, and W. Willinger. Long-range dependence in variable-bit-rate video traffic. *IEEE Trans. on Communications*, 43(2):1566–1579, Feb.-Apr. 1995.
- [9] L. Boroczky, A.Y. Nagi, and E.F. Westermann. Joint rate control with look-ahead for multi-program video coding. *IEEE Trans. on Circuits and Systems for Video Technology*, 10(7):1159–1163, October 2000.

- [10] J. Chakareski and B. Girod. Rate-distortion optimized packet scheduling and routing for media streaming with path diversity. In *Proc. Data Compression Conf.*, pages 203–212, March 2003.
- [11] N.H.L. Chan and P.T. Mathiopoulos. Efficient video transmission over correlated Nakagami fading channels for IS-95 CDMA systems. *IEEE Journal on Selected Areas in Communications*, 18(6):996–1011, June 2000.
- [12] M. Chen, G.-M. Su, and M. Wu. Robust distributed multi-point video conferencing over error-prone channels. In *IEEE International Conference on Multimedia and Expo*, 2006.
- [13] Y. Chen, J. C. Ye, C. R. Floriach, and K. Challapali. Robust video streaming over wireless LAN with efficient scalable coding and prioritized adaptive transmission. In *IEEE International Conference on Image Processing*, volume 3, pages 14–17, September 2003.
- [14] R.S. Cheng and S. Verdu. Gaussian multiaccess channels with ISI: capacity region and multiuser water-filling. *IEEE Trans. on Information Theory*, 39(3):773–785, May 1993.
- [15] P.J. Cherriman, T. Keller, and L. Hanzo. Orthogonal frequency-division multiplex transmission of H.263 encoded video over highly frequency-selective wireless networks. *IEEE Trans. on Circuits and Systems for Video Technology*, 9(5):701–712, August 1999.
- [16] S.-J. Choi and J. W. Woods. Motion-compensated 3-D subband coding of video. *IEEE Trans. on Image Processing*, 8(2):155–167, February 1999.
- [17] P. S. Chow, J. M. Cioffi, and J. A. C. Bingham. A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels. *IEEE Trans. on Communications*, 43(2/3/4):773–775, Feb./Mar./Apr. 1995.
- [18] R. Cohen and H. Radha. Streaming fine-grained scalable video over packet-based networks. In *IEEE Global Telecommunications Conference*, volume 1, pages 288–292, 2000.
- [19] W. Ding. Joint encoder and channel rate control of VBR video over ATM networks. *IEEE Trans. on Circuits and Systems for Video Technology*, 7(2):266–278, April 1997.
- [20] E. O. Elliott. Estimates of error rates for codes on burst-noise channels. *Bell System Technology Journal*, 42(9):1977–1997, September 1963.

- [21] Third Generation Partnership Project 2 (3GPP2). CDMA2000 high rate packet data air interface specification, C.S20024-A, August 2005.
- [22] E. N. Gilbert. Capacity of a burst-noise channel. *Bell System Technology Journal*, 39(9):1253–1265, September 1960.
- [23] B. Girod and N. Farber. *Wireless video in Compressed video over networks*. M-T. Sun and A.R. Reibman, Eds. Marcel Dekker, Inc, 2001.
- [24] N. Gogate, D.-M. Chung, S. S. Panwar, and Y. Wang. Supporting image and video applications in a multihop radio environment using path diversity and multiple description coding. *IEEE Trans. on Circuits and Systems for Video Technology*, 12(9):777–792, September 2005.
- [25] A. Goldsmith. *Wireless communications*. Cambridge University Press, 2005.
- [26] S. Gringeri, R. Egorov, K. Shuaib, A. Lewis, and B. Basch. Robust compression and transmission of MPEG-4 video. In *ACM International conference on multimedia*, pages 113–120, June 2000.
- [27] J. Hagenauer. Rate compatible punctured convolutional (rcpc) codes and their applications. *IEEE Trans. on Communications*, 36(4):389–399, April 1988.
- [28] Z. Han, Z. Ji, and K. J. R. Liu. Low-complexity OFDMA channel allocation with Nash bargaining solution fairness. In *IEEE Global Telecommunications Conference*, pages 3726–3731, 2004.
- [29] Z. Han, Z. Ji, and K. J. R. Liu. Power minimization for multi-cell OFDM networks using distributed non-cooperative game approach. In *IEEE Global Telecommunications Conference*, pages 3742–3747, 2004.
- [30] Z. Han, A. Kwasinski, K.J.R. Liu, and N. Farvardin. Pizza party algorithm for real time distortion management in downlink single-cell cdma systems. In *Allerton Conference*, 2003.
- [31] Zhu Han, Guan-Ming Su, Andres Kwasinski, Min Wu, and K. J. Ray Liu. Multiuser distortion management of layered video over resource limited downlink MC-CDMA. *IEEE Trans. on Wireless Communications*, 5(10), October 2006.
- [32] L. Hanzo, S.X. Ng, T. Keller, and W.T. Webb. *Single and multicarrier quadrature amplitude modulation: from basics to adaptive trellis-Coded, turbo-equalised and space-time coded OFDM, CDMA and MC-CDMA systems*. Wiley, 2004.

- [33] Z. He and S.K. Mitra. A linear source model and a unified rate control algorithm for DCT video coding. *IEEE Trans. on Circuits and Systems for Video Technology*, 12(11):970–982, November 2002.
- [34] C.-Y. Hsu, A. Ortega, and A.R. Reibman. Joint selection of source and channel rate for VBR video transmission under ATM policing constraints. *IEEE Journal on Selected Areas in Communications*, 15(6):1016–1028, August 1997.
- [35] J. Hua, Z. Xiong, and X. Wu. High-performance 3-D embedded wavelet video (EWV) coding. In *IEEE Workshop on Multimedia Signal Processing*, 2001.
- [36] C.-L. I and R.D. Gitlin. Multicode cdma wireless personal communications networks. In *IEEE International Conf. on Communications*, volume 2, pages 1060–1064, June 1995.
- [37] C.-L. I, C. A. Webb III, H. C. Huang, S. t. Brink, S. Nanda, and R. D. Gitlin. IS-95 enhancements for multimedia services. *Bell System Technology Journal*, 1(2):60–87, Autumn 1996.
- [38] ISO/IEC. JTC1/SC29/WG11 MPEG-4 video verification model version 18.0, January 2001.
- [39] ISO/IEC. JTC1/SC29/WG11 applications and requirements and for scalable video coding N6880, January 2005.
- [40] S.A. Jafar and A. Goldsmith. Adaptive multirate CDMA for uplink throughput maximization. *IEEE Trans. on Wireless Communications*, 2(2):218–228, March 2003.
- [41] W. S. Jeon and D. G. Jeong. Call admission control for mobile multimedia communications with traffic asymmetry between uplink and downlink. *IEEE Trans. on Vehicular Technology*, 50(1):59–66, January 2001.
- [42] W. S. Jeon and D. G. Jeong. Call admission control for CDMA mobile communications systems supporting multimedia services. *IEEE Trans. on Wireless Communications*, 1(4):649–659, October 2002.
- [43] D. G. Jeong and W. S. Jeon. CDMA/TDD system for wireless multimedia services with traffic unbalance between uplink and downlink. *IEEE Journal on Selected Areas in Communications*, 17(5):939–946, May 1999.
- [44] Z. Ji, Q. Zhang, W. Zhu, J. Lu, and Y.-Q. Zhang. Video broadcasting over MIMO-OFDM systems. In *IEEE International Symposium on Circuits and Systems*, volume 2, pages 844–847, 2003.

- [45] M.K. Karakayali, R. Yates, and L. Razumov. Throughput maximization on the downlink of a cdma system. In *IEEE Wireless Communications and Networking Conference*, volume 2, pages 894–901, March 2003.
- [46] B.-J. Kim, Z. Xiong, and W. A. Pearlman. Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT). *IEEE Trans. on Circuits and Systems for Video Technology*, 10(8):1374–1387, December 2000.
- [47] J. Kim, R. M. Mersereau, and Y. Altunbasak. Distributed video streaming using multiple description and unequal error protection. *IEEE Trans. on Image Processing*, 14(7):849–861, July 2005.
- [48] D. Kivanc, G. Li, and H. Liu. Computationally efficient bandwidth allocation and power control for OFDMA. *IEEE Trans. on Wireless Communications*, 2(6):1150–1158, November 2003.
- [49] C. E. Koksal, H. Kassab, and H. Balakrishnan. An analysis of short-term fairness in wireless media access protocols. In *ACM Sigmetrics*, June 2000.
- [50] T.V. Lakshman, A. Ortega, and A.R. Reibman. VBR video: tradeoffs and potentials. *Proc. of IEEE*, 86(5):952–973, May 1998.
- [51] H.-J. Lee, T. Chiang, and Y.-Q. Zhang. Scalable rate control for MPEG-4 video. *IEEE Trans. on Circuits and Systems for Video Technology*, 10(6):878–894, September 2000.
- [52] Q. Li and M. van der Schaar. Providing adaptive QoS to layered video over wireless local area networks through real-time retry limit adaptation. *IEEE Trans. on Multimedia*, 6(2):278–290, April 2004.
- [53] W. Li. Overview of fine granularity scalability in MPEG-4 video standard. *IEEE Trans. on Circuits and Systems for Video Technology*, 11(3):301–317, March 2001.
- [54] L.-J. Lin and A. Ortega. Bit-rate control using piecewise approximated rate-distortion characteristics. *IEEE Trans. on Circuits and Systems for Video Technology*, 8(4):446–459, August 1998.
- [55] C.E. Luna, Y. Eisenberg, R. Berry, T.N. Pappas, and A.K. Katsaggelos. Joint source coding and data rate adaptation for energy efficient wireless video streaming. *IEEE Journal on Selected Areas in Communications*, 21(10):1710–1720, December 2003.
- [56] C.E. Luna, L.P. Kondi, and A.K. Katsaggelos. Maximizing user utility in video streaming applications. *IEEE Trans. on Circuits and Systems for Video Technology*, 13(2):141–148, February 2003.

- [57] A. Majumdar, R. Puri, and K. Ramchandran. Distributed multimedia transmission from multiple servers. In *IEEE International Conference on Image Processing*, volume 3, pages 177–180, June 2002.
- [58] A. Majumdar, D.G. Sachs, I.V. Kozintsev, K. Ramchandran, and M.M. Yeung. Multicast and unicast real-time video streaming over wireless LANs. *IEEE Trans. on Circuits and Systems for Video Technology*, 12(6):524–534, June 2002.
- [59] S. Mao, S. Lin, S. S. Panwar, Y. Wang, and E. Celebi. Video transport over ad hoc networks: multistream coding with multipath transport. *IEEE Journal on Selected Areas in Communications*, 21(10):1721–1737, December 2003.
- [60] S. Martello and P. Toth. *Knapsack Problems: Algorithms and Computer Implementations*. Wiley, 1990.
- [61] A.E. Mohr, E.A. Riskin, and R.E. Ladner. Unequal loss protection: graceful degradation of image quality over packet erasure channels through forward error correction. *IEEE Journal on Selected Areas in Communications*, 18(6):819828, June 2000.
- [62] K. N. Ngan, T. Meier, and Z. Chen. Improved single-video-object rate control for MPEG-4. *IEEE Trans. on Circuits and Systems for Video Technology*, 13(5):385–393, May 2003.
- [63] T. Nguyen and A. Zakhor. Multiple sender distributed video streaming. *IEEE Trans. on Multimedia*, 6(2):315–326, April 2003.
- [64] J.-R. Ohm. Three-dimensional subband coding with motion compensation. *IEEE Trans. on Image Processing*, 3(5):559–571, September 1994.
- [65] J.-R Ohm. Advances in scalable video coding. *Proc. of IEEE*, 93(1):42–56, Jan 2005.
- [66] D.E. Ott, T. Sparks, and K. Mayer-Patel. Aggregate congestion control for distributed multimedia applications. In *IEEE INFOCOM*, volume 1, pages 13–23, March 2004.
- [67] IEEE P802.11e/Draft 6.0. Draft Amendment to IEEE Std 802.11, 1999 edition, medium access control enhancements for quality of service, November 2003.
- [68] A. Papoulis and S. U. Pillai. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 2001.

- [69] C.I. Podilchuk, N.S. Jayant, and N. Farvardin. Three-dimensional sub-band coding with motion compensation. *IEEE Trans. on Image Processing*, 4(2):125–139, February 1995.
- [70] J. G. Proakis. *Digital communication*. McGraw-Hill, 1995.
- [71] R. Puri, K.-W Lee, K. Ramchandran, and V. Bharghavan. An integrated source transcoding and congestion control paradigm for video streaming in the internet. *IEEE Trans. on Multimedia*, 3(1):18–32, March 2001.
- [72] H. Radha, Y. Chen, K. Parthasarathy, and R. Cohen. Scalable internet video using MPEG-4. *Signal Processing: Image Communication*, 15(1-2):95–126, September 1999.
- [73] H. Radha and M. Wu. Overlay and peer-to-peer multimedia multicast with network-embedded FEC. In *IEEE International Conference on Image Processing*, volume 3, pages 1747–1750, October 2004.
- [74] H.M. Radha, M. van der Schaar, and Y. Chen. The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP. *IEEE Trans. on Multimedia*, 3(1):53–68, March 2001.
- [75] T. S. Rappaport. *Wireless Communications: principles and practice: 2nd edition*. Prentice Hall, 2002.
- [76] E.C. Reed and J.S. Lim. Optimal multidimensional bit-rate control for video communication. *IEEE Trans. on Image Processing*, 11(8):873–885, August 2002.
- [77] A.R. Reibman and B.G. Haskell. Constraints on variable bit-rate video for ATM networks. *IEEE Trans. on Circuits and Systems for Video Technology*, 2(4):361–372, December 1992.
- [78] J. Ribas-Corbera and S. Lei. Rate control in DCT video coding for low-delay communications. *IEEE Trans. on Circuits and Systems for Video Technology*, 9(1):172–185, February 1999.
- [79] H. Rohling and R. Grunheid. Performance comparison of different multiple access schemes for the downlink of an OFDM communication system. In *Proc. IEEE Vehicular Technology Conference*, pages 1365–1369, 1997.
- [80] S.P. Sethi and Q. Zhang. *Hierarchical Decision Making in Stochastic Manufacturing Systems*. Birkhäuser Boston, Cambridge, MA, 1994.

- [81] Y. Shan, I. V. Bajic, S. Kalyanaraman, and J. W. Woods. Overlay multi-hop FEC scheme for video streaming over peer-to-peer networks. In *IEEE International Conference on Image Processing*, volume 5, pages 3133–3136, October 2004.
- [82] Y. Shan and A. Zakhor. Cross layer techniques for adaptive video streaming over wireless networks. In *IEEE International Conference on Multimedia and Expo*, volume 1, pages 277–280, August 2002.
- [83] P.G. Sherwood and K. Zeger. Error protection for progressive image transmission over memoryless and fading channels. *IEEE Trans. on Communications*, 46(12):1555–1559, December 1998.
- [84] T. Sikora. Trends and perspectives in image and video coding. *Proc. of IEEE*, 93(1):6–17, January 2005.
- [85] J. Song and K. J. R. Liu. Robust progressive image transmission over OFDM systems using space-time block code. *IEEE Trans. on Multimedia*, 4(3):394–406, September 2002.
- [86] J. Song and K. J. R. Liu. An integrated source and channel rate allocation scheme for robust video coding and transmission over wireless channels. *EURASIP Journal on Applied Signal Processing*, 2004(2):304–316, February 2004.
- [87] V. M. Stankovic, R. Hamzaoui, and Z. Xiong. Real-time error protection of embedded codes for packet erasure and fading channels. *IEEE Trans. on Circuits and Systems for Video Technology*, 14(8):1064–1072, August 2004.
- [88] K. Stuhlmuller, N. Farber, M. Link, and B. Girod. Analysis of video transmission over lossy channels. *IEEE Journal on Selected Areas in Communications*, 18(6):1012–1032, June 2000.
- [89] G.-M. Su, M. Chen, and M. Wu. Cross-path PDMA-based error protection for streaming multiuser video over multiple paths. In *IEEE International Conference on Image Processing*, 2006.
- [90] G.-M. Su, Z. Han, A. Kwasinski, M. Wu, K. J. R. Liu, and N. Farvardin. Distortion management of real-time MPEG-4 FGS video over downlink multicode CDMA networks. In *IEEE International Conf. on Communications*, volume 5, pages 3071–3075, 2004.
- [91] G.-M. Su, Z. Han, M. Wu, and K. J. R. Liu. Joint uplink and downlink optimization for video conferencing over wireless LAN. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages 1101–1104, March 2000.

- [92] G.-M. Su, Z. Han, M. Wu, and K. J. R. Liu. Dynamic distortion control for 3-D embedded wavelet video over multi-user OFDM networks. In *IEEE Global Telecommunications Conference*, volume 2, pages 650–654, 2004.
- [93] G.-M. Su, Z. Han, M. Wu, and K. J. R. Liu. Multiuser cross-layer resource allocation for video transmission over wireless networks. *IEEE Network Magazine*, 20(2):21–27, March/April 2006.
- [94] G.-M. Su, Z. Han, M. Wu, and K. J. R. Liu. Joint uplink and downlink optimization for real-time multiuser video conferencing over WLANs. *IEEE Journal of Selected Topics in Signal Processing*, submitted.
- [95] G.-M. Su, Z. Han, M. Wu, and K. J. R. Liu. A scalable multiuser framework for video over OFDM networks: fairness and efficiency. *IEEE Trans. on Circuits and Systems for Video Technology*, to appear.
- [96] G.-M. Su and M. Wu. Efficient bandwidth resource allocation for low-delay multiuser MPEG-4 video transmission. In *IEEE International Conf. on Communications*, volume 3, pages 1308–1312, 2004.
- [97] G.-M. Su and M. Wu. Efficient bandwidth resource allocation for low-delay multiuser MPEG-4 video transmission. *IEEE Trans. on Circuits and Systems for Video Technology*, 15(9):1124–1137, September 2005.
- [98] M-T. Sun and A.R. Reibman. *Compressed video over networks*. Marcel Dekker, Inc, 2001.
- [99] D. Taubman and A. Zakhor. Multirate 3-D subband coding of video. *IEEE Trans. on Image Processing*, 3(5):572–588, September 1994.
- [100] D.S. Taubman. High performance scalable image compression with EBCOT. *IEEE Trans. on Image Processing*, 9(7):1158–1170, July 2000.
- [101] D. Tse and S.V. Hanly. Multiaccess fading channels. I. Polymatroid structure, optimal resource allocation and throughput capacities. *IEEE Trans. on Information Theory*, 44(7):2796–2815, November 1998.
- [102] D. Tse and P. Viswanath. *Fundamentals of wireless communication*. Cambridge University Press, 2005.
- [103] S. Ulukus, E. Biglieri, and M.Z. Win. Optimum modulation and multicode formats in cdma systems with multiuser receivers. In *IEEE INFOCOM*, volume 1, pages 395–402, April 2001.
- [104] S. Ulukus and R. D. Yates. Iterative construction of optimum signature sequence sets in synchronous CDMA systems. *IEEE Trans. on Information Theory*, 47(5):1989–1998, July 2001.

- [105] M. van der Schaar, S. Krishnamachari, S. Choi, and X. Xu. Adaptive cross-layer protection strategies for robust scalable video transmission over 802.11 WLANs. *IEEE Journal on Selected Areas in Communications*, 21(10):1752–1763, December 2003.
- [106] M. van der Schaar and H.M. Radha. A hybrid temporal-SNR fine-granular scalability for internet video. *IEEE Trans. on Circuits and Systems for Video Technology*, 11(3):318–331, March 2001.
- [107] M. Wahlqvist, H. Olofsson, M. Ericson, C. Ostberg, and R. Larsson. Capacity comparison of an OFDM based multiple access system using different dynamic resource allocation. In *Proc. IEEE Vehicular Technology Conference*, volume 3, pages 1664–1668, May 1997.
- [108] L. Wang and A. Vincent. Bit allocation and constraints for joint coding of multiple video programs. *IEEE Trans. on Circuits and Systems for Video Technology*, 9(6):949–959, September 1999.
- [109] Y. Wang, J. Ostermann, and Y.-Q. Zhang. *Video processing and communications*. Prentice Hall, 2001.
- [110] Y. Wang, A. R. Reibman, and S. Lin. Multiple description coding for video delivery. *Proc. of IEEE*, 93(1):57–70, January 2005.
- [111] Y. Wang and Q. Zhu. Error control and concealment for video communication: A review. *Proc. of IEEE*, 86(5):974–997, May 1998.
- [112] H.-Y. Wei, C.-C. Chiang, and Y.-D. Lin. Co-DRR: an integrated uplink and downlink scheduler for bandwidth management over wireless LANs. In *IEEE International Symposium on Computers and Communication*, volume 2, pages 1415–1420, 2003.
- [113] C. Y. Wong, R. S. Cheng, K. B. Lataief, and R. D. Murch. Multiuser OFDM with adaptive subcarrier, bit, and power allocation. *IEEE Journal on Selected Areas in Communications*, 17(10):1747–1758, October 1999.
- [114] M. Wu, R. Joyce, H-S. Wong, L. Guan, and S-Y. Kung. Dynamic resource allocation via video content and short-term traffic statistics. *IEEE Trans. on Multimedia*, 3(2):186–199, June 2001.
- [115] Y. Xiao. IEEE 802.11e: QoS provisioning at the MAC layer. *IEEE Wireless Communications*, 11(3):72–79, June 2004.
- [116] J. Xu, Z. Xiong, S. Li, and Y.-Q. Zhang. Memory-constrained 3-D wavelet transform for video coding without boundary effects. *IEEE Trans. on Circuits and Systems for Video Technology*, 12(9):812–818, September 2002.

- [117] J. Xu, Q. Zhang, W. Zhu, X.-G. Xia, and Y.-Q. Zhang. Optimal joint source-channel bit allocation for MPEG-4 fine granularity scalable video over OFDM system. In *IEEE International Symposium on Circuits and Systems*, volume 2, pages 360–363, 2003.
- [118] W. Xu and S.S.Hemami. Efficient partitioning of unequal error protected mpeg video streams for multiple channel transmission. In *IEEE International Conference on Image Processing*, volume 2, pages 721–724, September 2002.
- [119] W. Xu and S.S.Hemami. Distortion optimized multiple channel image transmission under delay constraints. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 997–1000, May 2004.
- [120] X. Xu, Y. Wang, S. S. Panwar, and K. W. Ross. A peer-to-peer video-on-demand system using multiple description coding and server diversity. In *IEEE International Conference on Image Processing*, volume 4, pages 1759–1762, October 2004.
- [121] Y. Yang and S.S. Hemami. Rate control for VBR video over ATM: simplification and implementation. *IEEE Trans. on Circuits and Systems for Video Technology*, 11(9):1045–1058, September 2001.
- [122] H. Yomo and S. Hara. An uplink/downlink asymmetric slot allocation algorithm in CDMA/TDD-based wireless multimedia communications systems. In *Proc. IEEE Vehicular Technology Conference*, volume 2, pages 797–801, 2002.
- [123] H. R. Sheikh Z. Wang and A. C. Bovik. *Objective Video Quality Assessment*. in *The Handbook of Video Databases: Design and Applications*, B. Furht and O. Marqure (Editors), CRC Press, 2003.
- [124] X.M. Zhang, A. Vetro, Y.Q. Shi, and H. Sun. Constant quality constrained rate allocation for FGS video coded videos. *IEEE Trans. on Circuits and Systems for Video Technology*, 13(2):121–130, February 2003.
- [125] L. Zhao, J. Kim, and C.-C. J. Kuo. MPEG-4 FGS video streaming with constant-quality rate control and differentiated forwarding. In *SPIE Conf. Visual Comm. & Image Processing*, pages 230–241, 2002.
- [126] H. Zheng and K. J. R. Liu. Robust image and video transmission over spectrally shaped channels using multicarrier modulation. *IEEE Trans. on Multimedia*, 1(1):88–103, March 1999.
- [127] H. Zheng and K. J. R. Liu. The subband modulation: a joint power and rate allocation framework for subband image and video transmission. *IEEE*

Trans. on Circuits and Systems for Video Technology, 9(5):823–838, August 1999.

- [128] X. S. Zhou and S.-P. Liou. Optimal nonlinear sampling for video streaming at low bit rates. *IEEE Trans. on Circuits and Systems for Video Technology*, 12(6):535–544, June 2002.
- [129] X. Zhu, S Han, and B.Girod. Congestion-aware rate allocation for multi-path video streaming over ad hoc wireless networks. In *IEEE International Conference on Image Processing*, volume 4, pages 2547–2550, October 2004.