Resource allocation and scheduling for scalable video streaming over wireless networks

Li, Maodong

2013

Li, M. (2013). Resource allocation and scheduling for scalable video streaming over wireless networks. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/54679>

<https://doi.org/10.32657/10356/54679>

RESOURCE ALLOCATION AND SCHEDULING FOR SCALABLE VIDEO STREAMING OVER WIRELESS NETWORKS

LI MAODONG

**SCHOOL OF ELECTRICAL AND ELECTRONIC ENGINEERING
NANYANG TECHNOLOGICAL UNIVERSITY**

A thesis submitted to the Nanyang Technological University
in fulfilment of the requirement for the degree of
Doctor of Philosophy

2013

To my parents

Acknowledgements

I have many to thank for their help and support during the development of this thesis. Foremost among them are Prof. Tan Yap-Peng and Dr. Chen Zhenzhong. To me, they have been exemplary personal and professional mentors. They always had good ideas to share, and provided easy ways to start. They were always energetic and full of passion. Whenever I met with problems in my research or daily life, they were always there ready to provide valuable comments and suggestions. They advised me how to read and write in English, how to embody research problems and solve them, everything from zero with their great patience and expertise. I would like to express my deepest gratitude and appreciation to Prof. Tan and Dr. Chen for what they have done for me. Their trust and support have been the main source of inspiration and impetus for me. They have given me valuable lessons covering wide ranges in video coding, wireless communication and etc. This thesis would not be in its current form without their constant guidance and constructive comments.

I would also like to thank the wonderful study group that discussed with me on video coding, wireless communication, signal processing, and other related topics. They are Chuah Seong-Ping, Huang Likun, Yang Gao, Lu Jiwen, and Liu Nini. Without the meaningful discussions with them, I cannot have a comprehensive understanding of the related fields. Their in-depth knowledge in these fields also teaches me a lot on how to carry out my research work.

There are also many individuals who have directly or indirectly contributed to this thesis. I am thankful to Wang Tao, Li Qiang, Hung Tzu-Yi, Yang Wei-Ting, Luo Ye, and Luo Shan for their friendship and encouragement in the past years. I would like also to thank all the anonymous reviewers who have given generously their time and expertise to review my papers and thesis.

Last but not least, I wish to dedicate this thesis to my parents who have supported me throughout all these years. Although it is a pity that I have had little time to stay with them, their unconditional love and patience gave me great power to move on.

Li Maodong

Singapore, July 2013

Abstract

Scalable video, evolved in a long way of more than 20 years, has attracted significant interest in the past decades in both academia and industry due to its versatility and potential in providing flexible storage and communication services. In contrast to the conventional video coding which encodes a video clip once and for all, the scalable video enables the composition of multiple clips in different specifications into a single bitstream and provides feasible mechanisms to adapt to the changing needs of networks or applications. With the finalization of the scalable extension of legacy H.264/AVC, i.e., Scalable Video Coding (SVC), the exploration on the potentials of coding, adaptation, scheduling and resource allocation becomes more interesting.

Especially with the era of mobile internet, the exponential growth in demand for wireless video puts the scalable video to the spotlight, considering its high coding efficiency and feasible rate adaptation capability. Due to the error-prone and resource scarcity nature of the wireless networks, it is critical that efficient and effective resource allocation and packet scheduling schemes can be developed to deal with the underlying challenges. A comprehensive scheme always lies in the joint investigations on the video characteristics and the network specifications. Video content analysis focuses mainly on coding and the optimization of rate-utility (rate-distortion), scalability-rate, scalability-quality (PSNR, MOS, QoE), and so on. Meanwhile, resource allocation and packet scheduling normally consider the specifications of wireless networks from different Open Systems Interconnection (OSI) layers.

In this thesis, scalable video streaming over wireless networks is investigated and evaluated in several typical wireless systems, through which we hope to inspire new research on this topic and foster its deployment in practical systems. The main contributions and achievements of this thesis are summarized below:

1. The first contribution of this thesis is the proposal of efficient packet prioritization and scheduling schemes for scalable video streaming over WLANs. The work provides an inspiration on how to apply the conventional unequal error protection strategy into scalable video transmissions and shows a novel way to deal with the network congestion issue in best effort networks.
2. This thesis also devotes to the fast scalable video rate adaptation. The fast adaptation scheme answers for how to obtain the maximal video quality in real-time scalable video adaptation and also provides effective tool to facilitate the video streaming applications.
3. The third contribution concerns a scalable resource allocation framework for SVC video streaming over MIMO-OFDM networks. By skillfully utilizing the feature of scalable video, the scalable framework renders a novel way to tackle the conventional dilemma in fair versus efficient design in multiuser resource allocation problems.
4. Based on a subjective video quality assessment database, the thesis summarizes various scalability adaptation tracks for QoE-aware rate adaptation. This thesis also constructs a rate-QoE model, which favors the understanding of scalable video adaptation and also facilitates the analysis in QoE-aware video streaming.

Contents

List of Figures	xvi
List of Tables	xviii
1 Introduction	1
1.1 Motivations	1
1.2 Contributions and Outline of Thesis	5
2 Background and Related Work	9
2.1 Background and Preliminaries	9
2.1.1 Scalable Video Coding	9
2.1.1.1 The Evolution of Scalable Video Compression	9
2.1.1.2 Video Coding Layer	11
2.1.1.3 Network Abstraction Layer	13
2.1.2 Video Quality Evaluation	13
2.1.3 IEEE 802.11e Networks	16
2.1.4 Multiuser MIMO-OFDM Networks	17
2.2 Related Technologies and Prior Work in Wireless Video Streaming	20
2.2.1 Resource Allocation	20
2.2.1.1 Rate Control	20
2.2.1.2 Power and Energy Efficient Transmission	20
2.2.1.3 Efficiency and Fairness Issues	21
2.2.2 Packet Scheduling	22
2.2.3 Quality of Service	23

CONTENTS

2.2.4	Quality of Experience	24
3	Packet Prioritization and Scheduling for Scalable Video Streaming over WLANs	27
3.1	Introduction	27
3.2	Packet Prioritization of Scalable Video	29
3.2.1	Multilayer prediction structure in SVC	29
3.2.2	MSE based packet prioritization	31
3.2.3	Flickering based packet prioritization	32
3.3	Packet Mapping and Scheduling	33
3.3.1	Analytical Model of EDCA	33
3.3.2	QoS Mapping	37
3.3.3	Queue Control and Packet Scheduling	39
3.4	Simulation Results	42
3.4.1	Simulation Settings	42
3.4.2	Performance Evaluation for MSE Based Prioritization and Scheduling .	43
3.4.2.1	Quality Verification I	43
3.4.2.2	Quality Verification II	47
3.4.2.3	Quality Verification III	48
3.4.3	Performance Evaluation for Flickering Based Prioritization and Scheduling	49
3.5	Concluding Remarks	50
4	A Fast Rate Adaptation Scheme for SVC Based on the Packet Dependencies	57
4.1	Introduction	57
4.2	Fast Rate Adaptation Based on Enhanced Packet Prioritization	58
4.2.1	Existing SVC Rate Adaptation Schemes	58
4.2.1.1	JSVM Basic Extraction	58
4.2.1.2	Quality Layers Based Rate Adaptation	58
4.2.2	Packet Prioritization Based on Packet Dependencies	59
4.3	Simulation Results	64
4.4	Concluding Remarks	66

5	Scalable Resource Allocation for Scalable Video Streaming over WMANs	71
5.1	Introduction	71
5.1.1	Utility of Scalable Video Coding	73
5.1.2	Multiuser MIMO-OFDM Networks	75
5.2	The Scalable Resource Allocation Framework	77
5.2.1	Problem Statement	77
5.2.2	Description of the Scalable Resource Allocation Framework	80
5.2.3	Resource Allocation for the Base Layer: Fairness	82
5.2.4	Resource Allocation for the Enhancement Layers: Efficiency	85
5.2.5	Summary of the Solution	87
5.3	Simulation Results	89
5.3.1	Simulation Settings	89
5.3.2	Performance Evaluation of the Fairness Scheme	90
5.3.3	Performance Evaluation of the Efficiency Scheme	92
5.3.4	Performance Evaluation of the Overall Scalable Resource Allocation Framework	93
5.4	Concluding Remarks	95
6	QoE-aware Scalable Video Adaptation and Streaming	97
6.1	Introduction	97
6.2	Preliminaries and Related Work	100
6.2.1	Adaptation Schemes in JSVM	102
6.2.2	Analytical Rate and Quality Modeling	102
6.3	QoE Database for SVC Adaptation	103
6.3.1	Test Materials	103
6.3.2	Evaluation Environment and Participants	107
6.3.3	Evaluation Method	107
6.3.4	Data Postprocessing	108
6.4	QoE Analysis for SVC Rate Adaptation	110
6.4.1	ANOVA Analysis	110
6.4.2	The Optimal Scalability Adaptation Track for Individual Video	111
6.4.3	The Common Scalability Adaptation Track for Grouped Videos	116
6.4.4	Rate-QoE Model and Validation	122

CONTENTS

6.5	QoE-aware Resource Allocation for SVC over Multiuser MIMO-OFDM Systems	124
6.5.1	QoE Maximization for Multiple Scalable Video Streaming	126
6.5.2	QoE-aware Resource Allocation	127
6.6	Simulation Results	130
6.6.1	Simulation Setting	130
6.6.2	Evaluations and Discussions	131
6.7	Concluding Remarks	134
7	Conclusions and Future Work	137
7.1	Main Contributions and Results	137
7.2	Future Work	140
	Author's Publications	143
	Bibliography	146

List of Figures

1.1	The instance of streaming scalable video to heterogeneous clients.	5
2.1	SVC encoding structure (based on Fig. 12, page 1116 of [10]).	11
2.2	Structure for temporal scalability coding (based on Fig. 1, page 1107 of [10]). .	12
2.3	Prediction structure for spatial scalable coding (based on Fig. 4, page 1109 of [10]).	12
2.4	Prediction structure for quality scalable coding (based on Fig. 8, page 1113 of [10]).	13
2.5	SVC NAL unit structure (based on page 15 of [19]).	14
2.6	Access categories of EDCA (based on Fig. 1, page 1666 of [35]).	18
3.1	The proposed packet prioritization and scheduling scheme.	28
3.2	Multilayer encoding structure in SVC with inter-layer prediction. (a) Prediction structure for spatial scalable coding (based on Fig. 4, page 1109 of [10]). (b) Prediction structure for quality scalable coding (based on Fig. 8, page 1113 of [10]).	31
3.3	The throughput ratio between AC2 and AC1 vs. the node number.	37
3.4	Average PSNR comparison with the reference scheme in [35]. (a) Bus Sequence received at node pair 0. (b) Foreman Sequence received at node pair 1. <i>Ref1</i> : packet prioritization, 2:1 allocation without IFQ control. <i>Ref2</i> : packet prioritization, 2:1 allocation with IFQ control. <i>Our1</i> : packet prioritization, $R(N)$ allocation without IFQ control. <i>Our2</i> : packet prioritization, $R(N)$ allocation with IFQ control.	45

LIST OF FIGURES

3.5	Average PSNR comparison with the reference scheme in [116]. (a) Bus Sequence received at node pair 0. (b) Foreman Sequence received at node pair 1.	47
3.6	Average PSNR comparison with the reference scheme in [117]. (a) Bus Sequence received at node pair 0. (b) Foreman Sequence received at node pair 1.	48
3.7	Comparison between our proposed scheme <i>Our2</i> and reference scheme <i>Ref1</i> . Left: our proposed scheme, Right: the reference scheme.	53
3.8	Comparison between our proposed scheme <i>Our2</i> and reference scheme <i>Ref2</i> . Left: our proposed scheme, Right: the reference scheme.	53
3.9	Comparison with the reference scheme in [116]. Left: our proposed scheme, Right: the reference scheme.	54
3.10	Comparison with the reference scheme in [117]. Left: our proposed scheme, Right: the reference scheme.	54
3.11	Flickering measure comparison with the reference algorithm in which no packet prioritization and QoS mapping adopted. Top: Sequence Bus of frames 800-1000. Average of the whole sequence: <i>Our</i> : 44.1, <i>Ref</i> : 758.0. Bottom: Sequence Foreman of frames 800-1000. Average of the whole sequence: <i>Our</i> : 36.64, <i>Ref</i> : 591.78.	55
3.12	Video comparison with no packet prioritization and QoS mapping scheme on flicker reduction. Left: our proposed method, Right: the reference method. . . .	55
4.1	Priority order for the JSVM basic extractor, packets are prioritized by spatial, temporal, and quality ID combinations (d, t, q) (based on Fig. 3(a), page 1188 of [138]).	59
4.3	Priority order in the given instance Fig. 4.2, (a) from JSVM Basic, (b) and (c) are from the proposed scheme: (b) quality base layer for inter-layer prediction. (c) quality enhancement layer for inter-layer prediction.	64
4.5	Snapshots of decoded video sequence for subjective assessment. (a) Video Bus adapted at 1100 kbps, scaled for easy viewing. Picture quality from left to right, JSVM Basic: 31.96 dB, Quality Layer: 32.93 dB, PD: 32.91 dB. (b) Video City adapted at 3400 kbps. Picture quality from left to right, JSVM Basic: 34.31 dB, Quality Layer: 35.21 dB, PD: 35.26 dB.	69

5.1	The SVC encoding structure for a two-quality layer configuration.	73
5.2	The system structure of multiple SVC video downlink transmissions over multiuser MIMO-OFDM networks (based on Fig. 1, page 1735 of [66] and Fig. 1, page 109 of [67]).	76
5.3	Illustration of the resource allocation.	78
5.4	Flowchart of the scalable resource allocation framework for scalable video streaming over multiuser MIMO-OFDM systems.	81
5.5	Jain's fairness index for utility gap of different video GOPs.	91
5.6	Subjective pictures from different resource allocation schemes. (a) RR, (b) LBRA [61], (c) MMSR [62], (d) ARRA [66], and (e) the Proposed.	95
6.1	SVC encoding structure (based on Fig. 12, page 1116 of [10]).	101
6.3	Snapshots of the training video sequences from JVT [175]. Left: Aspen. Right: In_to_tree.	104
6.4	Snapshots of the tested video sequences from JVT[175]. Upper: City, Crew, Harbour and Soccer. Lower: Ducks, Ice, Oldtown and Park.	104
6.5	Spatial Information (SI) and Temporal Information (TI) computed for the test sequences based on [22].	106
6.6	Stimulus presentation in the subjective assessment.	108
6.7	Average ratings for video City in all (d, t, q) combinations.	111
6.8	Average ratings for video City in all (d, t, q) combinations.	112
6.9	Instances showing the multiple choices in scalability adaptation. Upper: video City, the left one is 3.75 fps, low quality, 810.74 kbps and 2.50 in MOS; the right one is 7.5 fps, low quality, 877.68 kbps and 3.31 in MOS. Lower: video Soccer, the left one is 7.5 fps, high quality, 2519.74 kbps and 3.08 in MOS; the right one is 30 fps, medium quality, 2467.43 kbps and 5.00 in MOS.	114
6.10	Rate-MOS relationships with various (d, t, q) combinations.	115
6.12	Comparison of scalability adaptation tracks for one-set grouping $G1$ with other state-of-the-art algorithms. The (d, t, q) orders of the derived common track are: (2,4,2), (2,3,2), (2,3,1), (1,3,2), (1,3,1), (1,2,1), (1,2,0), (1,1,0), (0,2,0), (0,1,0) and (0,0,0).	120

LIST OF FIGURES

- 6.13 The transmitter structure of SVC video transmission over multiuser MIMO-OFDM systems (based on Fig. 1, page 1735 of [66] and Fig. 1, page 109 of [67]). 125
- 6.14 Sample pictures from video Crew. The upper row is from the JSVM based extraction and the bottom row is from the QoE-aware adaptation. From left to the right: ARRA, RR, MMS, QoERA and the full rate. 134

List of Tables

2.1	The Parameters of IEEE 802.11e MAC Protocol [33]	18
3.1	Notation used in Chapter 3	30
3.2	The IEEE 802.11e Duration Constants [35]	36
3.3	The Proposed Packet Prioritization and Scheduling Scheme	42
3.4	Video Quality (dB) obtained by using constant and adaptive allocation ratio among two ACs	44
4.1	Coding configuration of the tested videos	64
4.2	PSNR (dB) from different rate adaptation schemes	65
4.3	Time consumption (second) by different rate adaptation schemes	65
5.1	Notation used in Chapter 5	74
5.2	PSNR (dB) of the received videos from the fairness scheme	91
5.3	PSNR (dB) of the received videos from the efficiency scheme	92
5.4	PSNR (dB) of the received videos from the overall scalable resource allocation framework	94
6.1	(d, t, q) combination and their coding configuration	107
6.2	Five-level MOS scales	108
6.3	Four-way ANOVA on the average ratings	110
6.4	MAD deviations of the Common Tracks	119
6.5	Common scalability adaptation tracks by various groupings	121
6.6	The approximation coefficients of the proposed Rate-QoE model	124

LIST OF TABLES

6.7	MAD deviations of the Rate-QoE model by various groupings	124
6.8	List of System Parameters	131
6.9	Average bit rate (Mbps) by different resource allocation algorithms	132
6.10	Average MOS ratings by different rate adaptation and resource allocation algorithms.	133
6.11	Validation of the rate-QoE model	135

Introduction

1.1 Motivations

The increasing popularity of wireless video applications, such as mobile TV, video telephony, and wireless video multicast has led to extensive research and development activities in video coding and wireless communications. The challenges of wireless video communication come from various aspects such as limited bandwidth, heterogeneous clients, time-varying wireless medium, etc. To accommodate these challenges, advanced video coding and wireless networking techniques, as well as smart resource allocation strategies, are needed to achieve economical bandwidth utilization, fair resource distribution, efficient power and spectrum management, etc. Nowadays, wireless networking technologies, such as wireless personal area networks (WPANs) [1], wireless local area networks (WLANs) [2], wireless metropolitan area networks (WMANs) [3], and cellular networks, make people communicate easier and in more efficient way. The advancement of long-life battery, low-cost chips, and expanded storage also enriches the multimedia applications in mobile devices. Moreover, the improvement in video coding technologies such as the standardizations of MPEG-2 [4], MPEG-4 [5], and H.264 [6] have provided high compression performance, robust error resilience, and flexible functionalities for wireless video services. Last but not least, extensive research and development activities have been conducted to provide efficient resource allocation and packet scheduling schemes to enable seamless wireless video experience.

Wireless video communication can be grouped into different class: messaging, streaming, and conversational services [7]. In the messaging service, a coded video file is downloaded completely by a client and then played back in the equipment, which generally requires long

1. INTRODUCTION

downloading times and large storage spaces. Another constraint is that the video viewing can start only when all the bitstreams are received. Video delivery by streaming can overcome these shortcomings by concurrent downloading and display. It splits videos into small parts and transmits them one by one. Viewers do not need to wait for the entire video to be buffered at their side. The simultaneous transmission and playback provides significant advantages over the downloading service. Conversational or interactive video should be real-time where the transmission is two-way point-to-point or two-way multipoint. In these cases, the process of encoding, transmission and decoding is significantly time constrained. The difficulty of real-time video is that there is very little time to capture, compress, encode, transmit and decode. Efficient compression schemes have to be employed to combat the time constraint. This also means that there is reduced video quality for a given number of bits, compared to off-line compression schemes. In addition, any variations in transmission rate or transmission errors cannot easily be overcome because there is simply insufficient time.

We will consider wireless video streaming applications in this thesis. The first challenge affects the video streaming service is the limited bandwidth, especially for the video transmission which generally needs high transmission capacity. The available bandwidth is generally unknown and especially in wireless environment, it is time-varying. An effective solution should try to estimate the available bandwidth and meanwhile match the transmission rate correspondingly. It is also crucial to decide the transmission rate such that it is fair to other concurrent flows. Secondly, the variation in end-to-end delay should be overcome because video frames have to be displayed at a constant rate at the receivers. The later arrived frames will generate jerking distortion [8]. The third fundamental problem in wireless environment is losses. The wireless channels are generally contaminated by various errors. Losses give rise to significant destructive effect on the quality of received videos due to its randomness. User mobility in mobile networks always leads to the varying network topology, which necessitates a frequent reformulation of optimal routing protocol. Besides, the wireless environment is always characterized by fading, noise, attenuation, shadowing, interference, etc., which makes channel conditions change over time. Competition for channel access and bandwidth among multiple users further complicate the networking design for bursty video traffics. The negative effects caused by different factors always interweave between each other and complicate the system design. As mobile video users always seek personalized experiences, an efficient video transmission strategy should therefore be adaptive and content-aware. Unlike general data communication, video communication over wireless networks is often characterized as

bandwidth intensive, delay sensitive, but loss-tolerant. The providing of Quality of Service (QoS) for different users poses great challenges. An efficient algorithm should always try to provide trade-offs among video quality, resource utilization, and implementation complexity.

To cope with these challenges, existing solutions mainly focus on two aspects: the resource allocation scheme and the packet scheduling scheme. For the resource allocation, the primary concern is to identify what resources are available and how they should be distributed for the best streaming service. In wireless networks, resources can be any combination of time slots, frequency bands, spatial signatures, power and energy, coded blocks, or other artificially developed elements. The resource utilization diverges greatly for different types of media therefore the solutions always vary by cases. The resource management issue is always addressed based on the network types. We will concentrate on two widely interested networks, i.e., the WLANs and WMANs, to analyze the resource allocation problems. The packet scheduling can be implemented from the sense of network packets or video packets. For the former, packets are encapsulated at the transport layers. While for the latter, the packets are mainly referred to as the video packets from the encoder, which can be readily considered as a component in the application layer.

A multitude of resource allocation and packet scheduling strategies can be found in different Open Systems Interconnection (OSI) layers to enhance the video streaming services. Among the single layer strategies are scalable video coding and transcoding at the application (APP) layer, multipath and multihop adaptive routing at the network layer, priority queuing and adaptive medium access schemes at the Medium Access Control (MAC) layer, adaptive modulation and coding techniques, power control and link adaptation at the physical (PHY) layer, etc. However, single layer based solutions may not be efficient in improving the system performance. The resource management and adaptation strategies at the lower PHY layer, MAC layer, network layer and transport layer can be optimized by considering the specific video characteristics from the APP layer. On the other hand, the video and streaming strategies could be optimized by considering the scheduling and protection schemes at the lower network layers. In such cases, the concept of cross-layer design [9] is widely employed to address challenges in wireless video transmission. However, which layers should be selected for the joint optimization and what information should be coordinated across layers are still case-dependent. The trade-off between the cross-layer's performance and its complexity in system realization should always be considered.

1. INTRODUCTION

Different video coding standards perform distinctively in facilitating the video streaming services. Many video coding standards have been established to address the difficulties arising from the wireless video streaming. Among them, the emerging scalable video coding technique shows distinctive advantages in coding efficiency and rate adaptation. Although different streaming services pose different constraints and requirements, the main objectives still lie in combating time-varying bandwidth, delay jitter, and packet loss. Scalable video is competent to provide solutions to these problems by its scalability and efficiency. Scalable Video Coding (SVC) [10] encodes video into base and enhancement layers. The base layer contains the lowest level of spatial and temporal resolutions and the coarsest quality representation. The enhancement layers improve the quality and/or increase the resolutions by successive layer coding upon the base layer. Different layers are integrated into a single file which reduces the administrative expense of linking and encoding [11]. Compared to H.264/AVC [6], the SVC-encoded bitstream is only about 10% larger than the H.264/AVC non-scalable video bitstream at the same quality [12]. In addition, the SVC base layer is compatible with the legacy H.264/AVC encoding standard. With existing hardware encoders, content providers can convert their current formats to SVC compatible streams easily [11]. A distinctive feature of SVC comes from its feasible bitstream adaptation capability to accommodate heterogeneous users. For a scalable encoded video, the base layer video with the frame rate of 15 frames per second (fps) could be provided to handheld devices when the usable bandwidth is low. For broadband users with larger display size, an enhanced quality with the frame rate of 30 fps and more quality enhancement bitstreams would be more favorable. If the encoded videos are transmitted to high definition terminals, more enhancement layers which support high resolution/quality could be delivered to render the most favorable user experience. The capability of flexible rate adaptation is specially suitable for wireless networks where channel status changes frequently [13]. The rate adaptation can be easily achieved by excluding some enhancement layer packets to conform to the network capacity. The merits of adopting scalable video in wireless transmission are presented in Fig. 1.1, where a one-time encoded bitstream can be easily adapted to various network and client requirements.

Despite its flexibility and superior compression performance, to best utilize the scalable video in the wireless streaming applications is not a straightforward task. Firstly, the rate-distortion performance of the scalable video coding is not easy to be modeled. Secondly, how to perform the rate adaptation for different users needs to be exploited. Thirdly, how the packets should be scheduled to suit different network specifications has not been well investigated.

Figure 1.1: The instance of streaming scalable video to heterogeneous clients.

To address all these challenges and to answer to the questions proposed in wireless video streaming services, this thesis covers a list of comprehensive studies on the scalable video encoding, adaptation, packet scheduling, resource allocation, etc. We follow a main thread to carry out the SVC analysis from simple and coarse criteria to more advanced and comprehensive studies. We first consider the resource allocation and scheduling issue under some common objective quality metrics. Then we move to more general definition of video utility. To compensate for the insufficiency of the objective metrics in characterizing the true video quality, we also analyze the Quality of Experience (QoE) issue in SVC adaptation and streaming. The resource allocation and scheduling topics will be addressed in WLAN and WMAN environments, to consider and apply on their advanced MAC and PHY layer techniques. The contents and contributions of this thesis are summarized in the next section.

1.2 Contributions and Outline of Thesis

Aiming at addressing the aforementioned problems and designing more robust algorithms to combat the difficulties in scalable video streaming over wireless networks, we propose in this thesis several enhanced methods. The improvement comes from coding analysis and packet prioritization for scalable video, rate adaptation in time-varying network, resource allocation in multiple network layers, network model construction, problem formulation, analysis and the proposal of solutions. The analysis on SVC will be established gradually from simple to more comprehensive studies. The remaining of this thesis is organized as follows:

1. INTRODUCTION

- **Chapter 2:** In Chapter 2, a review of background and the related work are given in the field of wireless video streaming from two aspects: video coding standards and network protocols. Some relevant concepts related to resource allocation and scheduling will also be reviewed.
- **Chapter 3:** In Chapter 3, a packet scheduling scheme is proposed, which includes packet prioritization, QoS mapping, queue control, for the delivery of SVC over the IEEE 802.11e wireless networks. The prioritization is based on per-packet calculation and the packet mapping and regulation are based on an accurate modeling of the MAC, Link layers of the WLANs. The proposed structure enables the interactions among different network layers, providing differentiated treatment for SVC video packets with unequal importance. A novel and efficient solution is proposed to reduce the computational complexity of the joint optimization problem. The proposed solution can achieve enhanced video quality for the end users.
- **Chapter 4:** The prioritization scheme introduced in Chapter 3 calls for off-line calculations. To reduce the computational burden in the preprocessing, we introduce in this chapter a fast rate adaptation scheme based on an enhanced prioritization scheme, which can derive packet priority efficiently with negligible complexities. The method will be shown to be fast and efficient with no obvious quality degradation introduced. Specifically, a new algorithm to extract the SVC bitstream by an enhanced packet prioritization scheme based on the packet prediction relationship (PPR) is presented. The prioritization utilizes only some header information of SVC video packets and no additional calculation is needed. Experimental results demonstrate that the proposed scheme achieves notable quality improvement over the basic bit extraction scheme in SVC and attains comparable performance to a well established rate adaptation scheme with a significant reduction in the computational cost. The fast rate adaptation provides a feasible way for addressing the multiuser resource allocation problems, where an instance will be shown in Chapter 5.
- **Chapter 5:** A scalable resource allocation framework is proposed for scalable video streaming over multiuser Multiple-Input Multiple-Output (MIMO)-Orthogonal Frequency Division Multiplexing (OFDM) networks. We exploit the utility of scalable video, which is defined based on the fast rate adaptation scheme in Chapter 4, and consider the resource allocation issue in the multiuser MIMO-OFDM networks. Different from Chapter

3 which focuses solely on exploiting system efficiency, a scalable framework is proposed to address both the fairness and efficiency issues. The scalable framework guarantees differentiated service objectives for different scalable video layers. To provide every user with a fair opportunity to acquire a basic video viewing experience, a fairness scheme is designed to guarantee that each user is entitled to the MAXMIN fairness to have their base layer video packets received. After all users have their base layer packets successfully scheduled, resources are distributed to exploit the network efficiency. Fast and efficient solutions are proposed for the fairness and efficiency resource allocation problems. The two solutions are integrated into a unified bit loading and power allocation scheme, which verifies the rationality and practicability of the scalable framework.

- **Chapter 6:** All previous chapters focus solely on the providing of video services under some objective quality metrics. This chapter studies the Quality of Experience (QoE) issues in scalable video adaptation and streaming. The QoE-aware adaptation is addressed by constructing a subjective video quality assessment database based on the full scalability of SVC. The optimal scalability adaptation track for individual video and common scalability adaptation tracks for grouped videos are then summarized based on the database. A rate-QoE model is developed for SVC adaptation and the QoE modeling provides the foundations in designing a QoE-aware resource allocation scheme for scalable video streaming over multiuser MIMO-OFDM systems. The QoE-aware adaptation and resource allocation scheme will be shown to be significantly superior to the QoE-blind counterparts in revealing the video users' true experience. Nevertheless, the QoE exploitation needs to be further improved before it can be widely deployed as the benchmark in wireless video services, this leads to the discussion of the possible future work in Chapter 7.
- **Chapter 7:** We summarize the work in this thesis and propose possible extensions for future study in this chapter. The unsolved problems in the topic of wireless video streaming are also highlighted for further investigation.

1. INTRODUCTION

Background and Related Work

In this chapter, we review recent work related to video streaming over wireless networks. We first give the background and preliminary of Scalable Video Coding (SVC) and wireless network protocols. Then we introduce some fundamental work dealing with video streaming over wireless networks as well as its extensions and applications in scalable video streaming.

2.1 Background and Preliminaries

2.1.1 Scalable Video Coding

2.1.1.1 The Evolution of Scalable Video Compression

Video coding has a long history and is designed to meet different requirements during its evolution stage. For example, MPEG-2 [4] was mainly designed for the coding of high quality moving pictures, while H.263 [14] was aimed at low bit rate communications. Excellence performance of context-based video coding supported by MPEG-4 [5], and efficient compression methods provided in H.264/AVC [6], can well adapt to most applications of interest, so they are the widely adopted standards in industry currently.

Scalable video coding forms the main focus of thesis. The scalability refers to removal of parts of the bitstream to adapt to diverse needs of end users and to time-varying network conditions. A similar concept of scalable coding is the layered coding, which provides quality scalability as in the existing standards. However, layered coded data may not be scalable. Scalable video can provide more options for various needs. MPEG-2 employs the data partitioning technique, which actually is not truly scalable coding but a way of partitioning a

2. BACKGROUND AND RELATED WORK

coded single layer bitstream into two parts to enhance the error resilience. Video coding standards since 1994 including MPEG-2/H.262, H.263, and MPEG-4 Part 2, all have temporal, spatial, and quality scalabilities. Nevertheless, except for temporal scalability, additional overhead and computational complexity are always needed. Meanwhile, codec generally operates on a discrete set of rates, and overheads increase greatly with more layers encoded. However, MPEG-4 makes progress as the notion of fine granularity scalability (FGS) is introduced, which surpasses previous standards since its predecessors generally create coarse levels of representation for quality scalability. FGS is advantageous in providing fine granular scalability and is more resilient to packet losses. MPEG-4 also presents the notion of object-based scalability which is different since previously scalability is often applied to the entire frame. In object-based scalability, a particular region of interest could be encoded with enhanced bitrate to acquire more vivid presentation. Nevertheless, all standards that could provide scalability up to this stage have the common shortcoming. They restrict scalability at the bitstream level to a predefined number of layers.

A powerful alternative is wavelet based scalable coding scheme [15] which allows the combination of temporal, spatial, and quality scalabilities with FGS over a broad range of data rates. The method is based on the framework of motion-compensated temporal filtering (MCTF), whose open-loop structure provides high flexibility and better error resilience. The wavelet based scalable video coding competed with the currently adopted scalable extension of H.264/AVC as candidate in MPEG's scalable proposal. H.264/SVC [10] was at last selected as the starting point for scalable video standard after close examination on their performance. The adopted SVC changed its temporal scalability structure from MCTF to conventional hierarchical B pictures. The comparison on their performance has been investigated in [16]. Although the concept of scalability has already been proposed in preceding coding standards but they have rarely been accepted. This is because their spatial and quality scalabilities commonly introduce a significant loss in coding efficiency and meanwhile increase the decoding complexity, when compared to their non-scalable counterparts. Next we will give detailed introduction on the emerging SVC coding standard and illustrate its underlying advantages in overcoming these potential shortcomings.

A generic encoder structure of SVC is given in Fig. 2.1 [10]. The encoder structure is organized firstly in spatial levels, known as dependency layers in the standard. Within each spatial layer, inter- or intra-prediction is used to encode the texture and motion information. Quality refinement and hierarchical-B structure are designed to support quality and temporal

Figure 2.1: SVC encoding structure (based on Fig. 12, page 1116 of [10]).

scalabilities, respectively. Downward compatibility is supported as the base spatial layer is compatible to H.264/AVC decoder.

2.1.1.2 Video Coding Layer

Following the AVC coding standard, the SVC consists of a Video Coding Layer (VCL) and a Network Abstraction Layer (NAL). VCL follows a block-based hybrid coding structure and is extended to support various scalabilities. The following items briefly review the realization of three typical scalabilities in SVC [10].

Temporal Scalability: An encoded bitstream provides temporal scalability by partitioning a set of access units from different time instants into temporal base and enhancement layers. Temporal layers IDs are used to label different layers. Temporal scalability with dyadic enhancement layers is realized by the hierarchical-B pictures, as shown in Fig. 2.2(a). Other two kinds of temporal scalabilities are presented in Fig. 2.2(b) and 2.2(c), respectively. Different prediction relations introduce different delay and memory requirements.

Spatial Scalability: SVC adopts the conventional multi-layer structure to realize spatial scalability. Spatial layers IDs (also known as dependency layer IDs) are used to label different layers. Within each spatial level, the adoption of conventional hierarchical structure combines the spatial scalability with the temporal scalability. The inter layer prediction enables the usage of the lower layer information to improve the high layer coding efficiency. The reconstructed

2. BACKGROUND AND RELATED WORK

Figure 2.2: Structure for temporal scalability coding (based on Fig. 1, page 1107 of [10]).

Figure 2.3: Prediction structure for spatial scalable coding (based on Fig. 4, page 1109 of [10]).

lower layer pictures are up-sampled for motion compensation. A competing alternative of inter layer prediction is temporal prediction within a single layer, which may be more applicable for low motion segments. The inter-layer prediction modes include: inter-layer motion prediction, inter-layer residual prediction and inter-layer intra-prediction [10]. An inter layer prediction structure of spatial scalability is depicted in Fig. 2.3.

Quality Scalability: SVC employs a key-picture concept [17] so realize medium-grain

Figure 2.4: Prediction structure for quality scalable coding (based on Fig. 8, page 1113 of [10]).

quality scalability (MGS). The bitstreams from quality enhancement layers can be truncated and the signaling mechanism makes the exchanging between different quality layers possible, which facilitates the rate adaptation. The key picture concept achieves the trade-off between drift and coding efficiency [10]. Combined with the hierarchical structure, the pictures can be decoded within a single loop. Fig. 2.4 illustrates the principle in quality scalability.

2.1.1.3 Network Abstraction Layer

The SVC encoded data are encapsulated into NAL units as its predecessor H.264/AVC. An NAL unit begins with a four-byte header, which includes NAL unit types, scalability IDs, discardable and truncatable flags, priority ID, and other additional information to assist bitstream manipulations. The remaining bytes are payload data. An illustration of NAL unit structure for SVC is illustrated in Fig. 2.5. The NAL units include VCL NAL units and non-VCL NAL units. The former ones include coded slices, while the later ones contain some additional information. Several consecutive NAL units form an access unit (AU), which contains the coded slices at a temporal location, and several consecutive AUs constitute a coded video sequence [10, 18].

2.1.2 Video Quality Evaluation

Video quality issue has always focused on two different aspects. One is on the initial appreciation of a video sequence like how the camera is placed and whether the exposure is adequate. While the other aspect concerns how the quality is affected when it goes through a processing or transmission procedure. The processing or the transmission may lead to different degrees

2. BACKGROUND AND RELATED WORK

Figure 2.5: SVC NAL unit structure (based on page 15 of [19]).

of quality change compared to the original video sequence. As our topic here is on the wireless video streaming, the quality issue discussed mainly refer to the second aspect. Nowadays, digital video systems have mostly replaced their analog counterparts. Hence, the video quality measurement covered in this work is only for digital video.

Subjective assessment is the most reliable way for the video quality evaluation. Because human beings are the ultimate receivers in most applications and video services are finally assessed by the acceptability of the end users. The subjective quality measurement is generally conducted on a number of subjects who provide scores after assessing the video clips under some standard procedures. A widely adopted index is the Mean Opinion Score (MOS) [20]. There are a multitude of evaluation methods. International recommendations for this aim, i.e., ITU-R BT.500-11 [21] and ITU-T P.910 [22], include specifications on how to perform different types of subjective tests. The quality ratings are generally derived from two different methods [23]. In single stimulus methods, viewers rate the quality of an impaired video sequence and provide their ratings. While in the double stimulus methods, viewers assess the change in quality by comparing the reference with the impaired videos. Instances for the former include single stimulus continuous quality evaluation (SSCQE) and Absolute Category Rating (ACR). Examples belonging to the double stimulus are implicit comparisons double stimulus continuous quality scale (DSCQS) and explicit comparisons double stimulus impairment scale (DSIS) [23]. Despite its accuracy, subjective evaluation tends to be computationally intensive, cumbersome, not repeatable, and is often hard to adapt to real time circumstance [24]. Nonetheless, subjective evaluations still serve as the benchmark for evaluating other objective quality metrics.

To cope with the difficulties in conducting subjective video quality assessment, a variety of objective metrics have been proposed to efficiently, economically and somewhat accurately to

evaluate the video quality or to predict the subjective MOS. There are a lot of classifications on the objective metrics under different criteria. By distinguishing how the reference is adopted during the evaluation, quality assessment is generally discriminated to full-reference, reduced-reference and non-reference metrics [25, 26]. The thesis will not go into the details on the classification and comparisons of different metrics, and the readers are suggested to refer to excellent work [27, 23, 28, 29] for overall introduction. The metrics that will be used in this thesis are briefly introduced below.

Peak-Signal-to-Noise-Ratio (PSNR) [30]: The most widely adopted data metric together with mean squared error (MSE). The formula to compute it is simple and it is also easy to understand from the mathematical view to maximize the PSNR (corresponding to minimize the MSE). The MSE and PSNR of a single frame are calculated as

$$MSE = \sum_{i=1}^M \sum_{j=1}^N \frac{[I(i, j) - O(i, j)]^2}{M \times N} \quad (2.1)$$

$$PSNR = 20 \lg\left(\frac{255}{\sqrt{MSE}}\right) \quad (2.2)$$

where $I(i, j)$ is the input signal at pixel (i, j) , $O(i, j)$ is the output signal, and $M \times N$ is the picture size. The result is a single number in decibels. For video sequence, the average amount over all frames is usually calculated to estimate the quality. Despite its popularity, PSNR cannot reflect the true visual experience effectively. The human visual system perceives video quality not only by the pixel fidelity but also considers the temporal and spatial relationships of the pixels.

Video Quality Metric (VQM) [31]: VQM measures on the video impairments such as noise, blurring, jerky, irregular motions, blocks and color distortion [31]. The combined effects are measured by a single metric. VQM takes the original and the processed/distorted videos as input. The quality scores are derived from a list of procedures including calibration, feature extraction, parameter calculation and VQM score calculation. VQM can be computed by differentiating applications such as: video conferencing, TV application, developer, PSNR and general purpose. The VQM uses a combination of seven parameters to make the comparisons. Four are for spatial gradient of luminance components, one is for contrast and temporal information of luminance components, and the remaining two are for chrominance components. The existing work shows that VQM has a high correlation with the results from a subjective assessment.

2. BACKGROUND AND RELATED WORK

Structural Similarity Index (SSIM) [32]: This method is not an error based scheme like the previous ones. By investigating the human vision system which is shown to be more sensitive to structural information from vision field, the SSIM extracts the structural distortion measurement instead of the error. The SSIM metric is calculated on various windows of an image as [32]

$$SSIM = \frac{\sigma_{io}}{\sigma_i \cdot \sigma_o} \cdot \frac{2\sigma_i \cdot \sigma_o}{(\sigma_i)^2 + (\sigma_o)^2} \cdot \frac{2\bar{i} \cdot \bar{o}}{(\bar{i})^2 + (\bar{o})^2} \quad (2.3)$$

where i and o denote the input and output images, $\bar{i}$ and $\bar{o}$ are their means, σ_i and σ_o are the corresponding standard deviations, and σ_{io} is the cross covariance. The three terms in Eq. (2.3) measure the loss of correlation, contrast distortion and luminance distortion, respectively. The SSIM values range from -1 to 1, and the best value is obtained only when a and b are identical. SSIM is designed to improve on traditional methods like PSNR. A measurement on structural distortion should give a better correlation to the subjective evaluation.

2.1.3 IEEE 802.11e Networks

In wireless local area networks (WLANs), users are connected to the Internet through the access point. The most widely deployed WLANs are based on the IEEE 802.11 standards. Different channel access mechanisms have been proposed to allocate the transmission opportunity to multiusers. The contention-based mechanism, Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA), is deployed in the IEEE 802.11 a/b/g standards. Although have been widely deployed, these standards can only provide services in best-effort fashions, which would restrict the QoS support in video communication. The IEEE 802.11e standard [33] introduces contention-based and controlled medium access strategies. These two strategies improve the quality of service support for multimedia communications. Among which, the Enhanced Distributed Channel Access (EDCA), belonging to contention-based scheme, provides prioritized traffic protection in different access categories (ACs). Each AC contends for medium with the following access parameters [33]:

- *AIFS*: arbitration interframe space. An AC counts down its backoff counter or attempts a transmission after some time, which is related to AIFS. For AC i , $AIFS[i] = SIFS + AIFSN[i] * slottime$, where short interframe space (SIFS) and slottime are specified as in the IEEE 802.11a standard [34].

- $CW_{\min}$: minimal contention window size. To avoid collision, AC enters a backoff stage which consists of a random number of slot times before it starts transmission. The number is initially selected uniformly in the region $[0, CW_{\min} - 1]$.
- $CW_{\max}$: maximal contention window size. After each unsuccessful transmission attempt, AC starts a new backoff procedure with a larger contention window size until $CW_{\max}$ is reached.
- $TXOP_{lim}$: transmission opportunity limitation. It is used to permit consecutive frames transmitted by each AC once it is granted the access to the medium.
- PF, RL : persistence factor and maximum retry limit. PF is the increasing factor of the size of contention window and its default value is 2. RL specifies the maximum retransmission limit of each AC before it drops the packet. Although these parameters are not specified in the protocol, they can also be modified for optimization to facilitate video transmission.

ACs are granted differentiated service priorities as different parameters are adopted. For AC_i ($i = 0, 1, 2, 3$) and AC_j ($j = 0, 1, 2, 3$), when $i < j$, $CW_{\min}$, $CW_{\max}$, and $AIFS$ have greater values for i than those for j . If one class has smaller values of the above parameters, the class has an earlier chance to obtain the wireless medium thus offers a higher priority service for its traffics. For $TXOP_{lim}$ and RL , a larger value guarantees a higher opportunity for the traffics to be successfully sent out. The parameters of four ACs used for video streaming in this work are listed in Table 2.1.

Every AC starts a backoff stage with time specified by $AIFS$. When the backoff counter gets down to zero, the AC starts to transmit if the network is available, or defers its transmission with a new backoff process if the medium is occupied. If collision happens, a virtual collision handler will give transmission opportunity to the AC with a higher priority. The other AC has to start a new process of sensing carrier and backoff counter decrement. Fig. 2.6 illustrates an outline of four ACs in the EDCA. Details of the EDCA could be found in [33].

2.1.4 Multiuser MIMO-OFDM Networks

In a Multiple Input Multiple Output (MIMO) system, multipel antennas provide substantial gains from different aspects [36, 37, 38]:

2. BACKGROUND AND RELATED WORK

Table 2.1: The Parameters of IEEE 802.11e MAC Protocol [33]

AC	AIFSN	CW _{min}	CW _{max}	TXOP _{lim}	PF	RL
AC3	2	7	15	0.003264	2	7
AC2	2	15	31	0.006016	2	7
AC1	3	31	1023	0	2	4
AC0	7	31	1023	0	2	4

Figure 2.6: Access categories of EDCA (based on Fig. 1, page 1666 of [35]).

Spatial Division Multiplexing (SDM) [39]: The SDM scheme aims at increasing the attainable multiplexing gain by exploiting the user-specific channel response. It increases the throughput of a system in terms of the number of bits transmitted within a given time/frequency resource blocks.

Spatial Division Multiple Access (SDMA) [40]: By contrast to SDM, the SDMA aims at maximizing the number of users supported on a system. The SDMA exploits the user-specific “spatial signature” to support multiple users within the same frequency band and time slot if the signatures are sufficiently different and can be accurately measured [36], which enables adaptive resource allocation enhanced system performance.

Spatial Diversity [41, 42]: In this kind of scheme, antennas are positioned as apart as possible to ensure that signals transmitted from different antennas experience independent fading. The intelligent combination of the independently faded signal can achieve the maximized diversity gain. Two of the most important techniques belonging to this subclass are Space-Time Block Coding (STBC) [41] and Space-Time Trellis Coding (STTC) [42].

Beamforming [43]: The beamforming scheme mitigates the interference introduced by users in the vicinity to increase the beamforming gain for the desired users. This is achieved by smart antennas to create spatially selected transmitter-receiver beams.

Orthogonal Frequency Division Multiplexing (OFDM) [40] is a robust multi-carrier modulation technique which provides superior link quality for high speed wireless communications. It can convert a frequency-selective channel into a set of parallel frequency-flat subchannels. These subchannels are normally orthogonal to each other so that interference between them can be eliminated. For the narrow-band, low-rate, frequency-nonselective subchannels, multiple streams can be parallelly transmitted. Each subchannel can also transmit to different users, resulting in the efficient Orthogonal Frequency Division Multiple Access (OFDMA) scheme[44]. This makes the co-transmissions of different types of traffics, such as voice, video, and text, in the same radio link a reality, and based on their individual QoE requirement. In addition, the OFDM allows different subchannels or users to employ different modulation schemes based on their specific channel or user characteristics. This can significantly enhance system performance by designing smart radio resource allocation schemes. Due to the superior performance provided by OFDM, it has been employed by a number of wireless standards, especially in the most advanced standards in WLANs and WMANs [36].

OFDM can be configured with the transmitter/receiver antenna arrays by exploring various diversities in the time-varying wireless environment, resulting in a MIMO-OFDM configuration [37, 38]. A MIMO-OFDM system is particularly suitable for adaptive resource utilization to exploit the multiple diversities in space/frequency domain. Flexible radio resource management and radio link control can be designed, which can specially enhance the resource utilization in a multiuser video transmission system.

2.2 Related Technologies and Prior Work in Wireless Video Streaming

To combat the inherent challenges posed in wireless video streaming, several existing and promising technologies need to be mentioned. They focus on different aspects to enhance the video experience for the clients. New tools and methodologies are still expected to be exploited in these techniques to adapt to new systems and emerging applications.

2.2.1 Resource Allocation

Resource allocation is of paramount importance for wireless video streaming due to the scarcity of resources such as power, frequency band, time, code, etc. The wireless channel encounters more harsh turbulence by noise, fading and shadowing, when compared to the wired networks. What makes the resource allocation problem more challenging is the resources are shared by heterogeneous terminals which are different in channel status and video requirements. The allocation of resources should always have a comprehensive consideration on rate, power, bandwidth and time resources, etc.

2.2.1.1 Rate Control

Rate control in wired networks assumes that network congestion can lead to packet delay, packet loss, or the jitter artifact [45]. However, these detrimental effects in wireless networks may also arise from radio link errors. The traditional solutions for wired network may fail to identify the congestion conditions, which may lead to unsatisfied results. Cross-layer approaches are thereby employed to achieve the rate adaptability over a wireless link. For example, an enhanced rate-distortion (R-D) model for AVC video rate control is proposed in [46]. The error control scheme in [47] also belongs to rate control class, even though the rate control is more general since it is designed not only for error reduction. Similar topic could also be found in [48], in which the authors presented a frame layer rate control for the H.264/AVC video. SVC shows its great flexibility for rate adaptation and controlling, relevant work could be found in [49, 50, 51, 52, 53, 54].

2.2.1.2 Power and Energy Efficient Transmission

Beside the R-D optimization strategy, there are also transmission power and energy optimization problems. Energy efficiency has been one of the critical issues in wireless video streaming.

Huge video traffic and computationally expensive coding algorithm consume significant portion of energy over the energy-limited mobile devices. While silicon fabrication technology and IC design strive to increase the energy efficiency of CPU for mobile platform, the overall energy performance can only be optimal if wireless transmission algorithm is energy-efficient. These contributions can be found in [55, 56, 57, 58]. The work in [55] involves collaborative minimization of power in different aspects. Authors in [59] introduced an algorithm on their cross-layer design for energy efficient video transmission. They formulated the cross-layer design problem as a resource allocation problem. A two-phase solution approach for the optimization problem was introduced in [60].

2.2.1.3 Efficiency and Fairness Issues

The exploitation on network efficiency has been extensively studied in the past decades. The efficient resource allocation schemes always dedicate to achieve the maximum throughput or minimum resource consumptions [61, 62, 63, 64, 65]. Usually, resource is measured by data transmitted in a time interval, the time ratio of a specified interval, bandwidth, packet loss rate, and others. For video applications, the objective is to achieve the highest overall users' video quality with limited radio resources [66, 67, 68]. The exploitation of efficiency can be classified into two categories, the user level and the system level. For the former one, the focus is only on how to distribute the same types of resources among users to achieve the uniform system utility such as rate or quality increment. While the latter emphasizes more on the system design, such as for heterogeneous traffics, i.e., file downloading and video streaming. The fairness criteria always focus on the same resources, but aim at providing equal resource allocation among users. The following fairness metrics have been widely adopted in wireless transmission:

Max-min fairness [65]: The philosophy is to allocate the resources, such as channels and time slots, to the user with the minimum QoS. The advantage of this kind of scheduling strategy is the delivery of extreme fairness among users in the same priority class. Because users may encounter different degrees of channel variations, this kind of fairness may not optimize the overall performance. It is an extreme case of fairness and performance tradeoff. In [69], the authors discussed the max-min fair allocation at the MAC layer. A distributed algorithm was designed to let each node to determine its link share.

Proportional fairness [65]: Proportional fairness maximizes the summation throughput and meanwhile gives all users a minimal level of service. It assigns each data flow a rate or a scheduling priority which is inversely proportional to its anticipated resource consumption

2. BACKGROUND AND RELATED WORK

[70]. To illustrate the idea, in [65] the authors define the maximal transmission rate for the i th user at time T as $C_i(T)$ and the transmission history for this user is $R_i(t)$. The proportional fairness scheduler can be expressed as follows [65]:

$$\arg \max_i \frac{C_i(T)}{\sum_{t=0}^{T-1} R_i(t) + \epsilon} \quad (2.4)$$

where ϵ is a small positive constant to prevent error of “divided by zero”. If the user is selected, $C_i(T) = R_i(T)$; otherwise $R_i(T) = 0$. Because of the ergotic characteristics, the preceding scheduler provides the estimated data rate proportional to the mean of channel conditions. So the preceding proportional fair scheduler provides a trade-off between fairness and performance.

The design of the system objective always contradicts with each other when the focus is either on fairness or efficiency. The providing of fairness among users always leads to compromised overall system performance, while the persistent emphasis on the efficiency often leads to high degrees of unfairness. In other words, a system achieving a higher level of efficiency will suffer from more unfairness. From the view of a network or system, it is of high importance to efficiently utilize the channel or other resources because the available radio resources are limited and the revenue should be maximized. From the users’ perspective, it is expected to have fair resource allocation such that they are not disadvantaged and their service requirements are guaranteed. To cope with such contradiction, there is some reference work that proposes to provide the trade-off between these two objectives by a mixed strategies [71, 72, 73, 74, 75]. In [73], the previous transmission intervals of a video GOP dedicate to provide fairness among users, while the later intervals focus on efficiency exploitation. By allocation intervals sharing between the two schemes, the trade off can be obtained based on applications. It should be noted that for heterogeneous traffics, i.e., real-time and non-real-time, the trade-off is not only on user level but also on system level. How to coordinate the competence between different types of traffics should be carefully designed [66, 75].

2.2.2 Packet Scheduling

Packet scheduling utilizes the dependency information in video coding, i.e., the packet prediction relationship to achieve the optimized packet transmission. The prediction dependency of the source encoder at the APP layer is mapped to the lower transport layer [45]. Given constraints such as transmission time, guaranteed QoS, or for certain feedback information, the

scheduling of packets can be designed to achieve enhanced video quality. The optimization can be carried out from two sides: the server-side and the client-side [45]. Adaptive media play-out [76] belongs to the former category, where a receiver can control the rate to match the play-out process. A server-side optimization scheme is proposed in [77]. The scheme utilizes the source information to schedule packets under different constraints.

The packet scheduling schemes can also be combined with some other techniques to enhance the quality of streaming service. Especially for scalable video packets which are inherently different in quality presentation, unequal treatment of video packets is helpful to obtain enhanced video quality. For example, scheduling strategies could be designed by considering how the erroneous packets are recovered.

Error Control: Error control based on feedback can also be employed for streaming services because of the tolerance to long Round Trip Time (RTT). Girod *et.al* reviewed this technique in depth in [78]. In their scheme, a feedback channel indicates the received and the lost parts of a bitstream. Intra mode can be employed for some macroblocks to stop the error propagation. Reference picture selection serves as another instance, where the feedback provides an instantaneous mechanism for the encoder to control the coding process [78]. If the encoded data can be classified into different priority considering their contribution to video quality, unequal error protection can be operated. Error control schemes can be generally classified into the “unequally partitioned bitstream, unequal protection mapping” framework [45]. The data partitioning scheme in MPEG-4 is a commonly referred example [79], which prioritizes pictures and slices into different categories considering their encoding modes and their roles in the prediction stage. The parameter sets used to distinguish picture types in H.264/AVC also belong to this class. The layered coding or to move further, the scalable coding schemes [54], [80], [81], and [82] are alternative forms of the data partitioning. A hybrid APP layer scheme considers forward error correction (FEC) and retransmissions is proposed in [47].

2.2.3 Quality of Service

Due to their flexibility and low cost infrastructure, wireless networks enable a tremendous number of applications. However, existing networks cannot render guaranteed quality of service (QoS) [83] support for the delay-sensitive, bandwidth-intense and loss-tolerant video communications. Some outstanding work could be found in [84, 85, 86, 87, 88], etc. Besides these work, we discuss below an important concept in providing QoS support for the video delivery.

2. BACKGROUND AND RELATED WORK

Cross-layer Optimization [9]: The conventional layered stack has significantly simplified the network design, but results in suboptimal video performance. Especially for video application with high bandwidth and delay requirements, the isolation between layers leads to significant performance degradation. A cross-layer approach enhances the system performance by considering multi-layers together [9]. The resulting flexibility helps to provide better QoS support given network dynamics and limited resources. The well-known joint source-channel coding [89] concept falls into this category. More general cases of cross-layer optimization are summarized in [9]. The authors classified the cross-layer design into different categories [9]: top-down approach, bottom-up approach, application-centric approach, MAC-centric approach and integrated approach. Please refer to [9] for details. How to select the best approach depends on the system goals, network protocols, application background, implementation complexity, and so forth [90, 91, 92].

2.2.4 Quality of Experience

Generally, Quality of Experience (QoE) considers how a system or service satisfies users' expectations. For multimedia applications, the QoE comprises factors such as video quality, audio quality, the synchronization of both, the presentation style, etc. Concentrating solely on our current topic, user's perception of streaming video characterized in terms of QoE is directly measured by the acceptability of the end users on the transmitted video quality [93]. QoE is relevant to but different from the aforementioned concept QoS, which mainly focuses on system related parameters. However, QoS lacks the human perception in quantifying the perceived video quality [94]. It tries to tackle the systemic issues from a network perspective rather than the users perceived effects. Of course, QoE is directly related to QoS, but the challenge is how to define the right set of tools and processes to map the QoS at the network level to the QoE at the user levels. Due to the growing mind share of QoE in academy and industry, the standardization efforts are progressing rapidly. Industry experts in International Telecommunication Union (ITU) have defined several standards related to video quality, such as [95]: ITU-T Rec. G.1070 [96], ITU-T Rec. J.246 [97] and ITU-T Rec. J.247 [98]. Meanwhile, an industry group, the Video Quality Experts Group (VQEG), has also been working on the video quality standardization since 1997 [95].

A convincing way to measure the video QoE is through the subjective MOS evaluations [20]. However, as we have discussed, the MOS evaluation tends to be time-consuming and

complex. Recently, great efforts have been put to model the QoE for wireless streaming. A non-reference evaluation framework has been presented in [99] by constructing a k -dimensional QoE space. The space is partitioned into N representative zones, each with a QoE index. Instantaneous parameter values are matched with the indices to infer QoE. Other important references include QoE modeling [100, 101], QoS to QoE mapping [102], etc. Despite the existing efforts, there are still various challenges in the QoE definition, quantification, measurement and provision due to the new research development in video quality such as perceptual video [103], multiview video [104], 3D video [105, 106], high efficiency video [107], etc. We will specially investigate in this thesis the QoE issue in scalable video adaptation and streaming.

2. BACKGROUND AND RELATED WORK

Packet Prioritization and Scheduling for Scalable Video Streaming over WLANs

3.1 Introduction

We consider an optimized scheme for scalable video streaming over wireless local area networks (WLANs). This part of work has been published in our papers [108, 109]. To support QoS for scalable video streaming over WLANs, we employ the multimedia-favored IEEE 802.11e standard as network protocol. The background of the IEEE 802.11e standard has been given in Section 2.1.3. In time-varying wireless environments, intensive competition and heavy traffic congestions are common threats to data delivery. Meanwhile, bulky video packets aggravate competition among both the internal and external EDCA queues, and the accumulative effects make queue overflow and delay negligible. As the SVC codec provides high flexibility for packets prioritization, it is possible to design an efficient packet scheduling scheme for EDCA by utilizing its differentiated service.

We have seen a multitude of approaches [89, 87, 110, 111, 112, 113, 114, 115] dedicated to the issue of video transmission over lossy networks. However, they mainly adopted non-scalable video coding schemes so the source adaptation is difficult. There are also some studies [35, 116, 117] focused on the SVC streaming over WLANs. In [35], the authors proposed an optimized cross-layer framework for the transmission of motion compensated temporal filtering (MCTF)-based scalable video. A differentiated ratio of bitstream distribution among AC2

3. PACKET PRIORITIZATION AND SCHEDULING FOR SCALABLE VIDEO STREAMING OVER WLANS

Figure 3.1: The proposed packet prioritization and scheduling scheme.

and AC1 was adopted. However, the ratio is fixed regardless of the network variation and the number of packets left in queues. In [116], the authors proposed a traffic prioritization scheme for SVC over the IEEE 802.11e wireless networks. It is easy to operate but only confined in simple scalability structure, and the authors failed to consider the throughput differentiation among ACs. Another work investigating SVC transmission was presented in [117]. The authors proposed an intuitive way of packet scheduling according to the SVC packets layer ID. However, the throughput of different ACs and the packet length information were not considered, leading to a suboptimal solution. There is other work aiming at designing robust wireless video transmission strategies [9, 77, 118, 119, 120], etc. They may or may not specially focus on scalable video and WLANs. Nevertheless, they provide in-depth exploitations on the problem and the ideas presented in these work are helpful for us to define and solve our specific problem.

In this chapter, we propose an efficient packet prioritization scheme to evaluate the importance of SVC packets. Based on this information, we devise an efficient packet scheduling algorithm to maximize the overall video quality at the receivers in a bandwidth-constrained network. The packet prioritization scheme assigns different priorities for packets based on their lengths and distortion impacts. It considers the layer a packet belonging to and the prediction relationship during encoding. The packet scheduling scheme maximizes the overall

video quality from two aspects. Firstly, according to length and distortion impact information obtained from the application layer and the services differentiations among different ACs in the MAC layer, a QoS mapping scheme decides packet allocation among different ACs with reference to the number of wireless nodes. Secondly, to maximize the overall video quality, an optimization model is developed based on the occupation status and the departure rate of the interface queues from the link layer. The sending order and timing are interactively decided and timely regulated to minimize the queue overflow. Fig. 3.1 shows the block diagram of the proposed scheme and Table 3.1 summarizes the notation used in this chapter.

3.2 Packet Prioritization of Scalable Video

We analyze the priority of SVC packets which is critical in designing the packet scheduling scheme. The analyses and validations are conducted on two objective quality measures, Mean Square Error (MSE) and flickering, which are widely adopted in video streaming. This analysis can be easily applied to other video quality metrics.

3.2.1 Multilayer prediction structure in SVC

The priority of an individual SVC packet depends on its frame type as well as its position in the coding structure. Because of the hierarchical structure in SVC, the error propagation is confined mostly within one GOP. Loss of the lone I frame will affect subsequent P and B frames. Loss of P frames will cause quality degradation on all the enhancement layers that are predicted directly or indirectly from them, and loss of B frames in lower layers will lead to distortions to those coded at higher layers. In addition, the intra and inter layer predictions bring complicated dependent relations between packets as three different scalabilities are supported. To better understand the dependency and the induced distortion caused by packet loss, we illustrate the multilayer prediction relations of SVC in Fig. 3.2.

As shown in Fig. 3.2, temporal scalability can be supported at arbitrary spatial layers. There are several types of hierarchical prediction structure to enable the temporal scalability. The structure depicted in Fig. 3.2 is the most common one, which is also adopted in this work. Any distorted I or P frames will lead to quality degradation to all subsequent frames of the current GOP. Loss of lower layer (close to base layer) B frames will lead to distortion on all higher layer B frames residing in the current GOP. For spatial layer packet loss, error propagation will spread to all higher layer frames which belong to enhancement layers of the

3. PACKET PRIORITIZATION AND SCHEDULING FOR SCALABLE VIDEO STREAMING OVER WLANS

Table 3.1: Notation used in Chapter 3

ACi ($i = 0, 1, 2, 3$)	access category of EDCA model
$AIFS[i]$	arbitration interframe space of ACi
$AIFSN[i]$	AIFS number of ACi
$CW_{min}[i]$	minimal contention window size of ACi
$CW_{max}[i]$	maximal contention window size of ACi
$TXOP_{lim}$	transmission opportunity limitation
PF	persistence factor of AC
$RL[i]$	maximum retry limit of ACi
Δ_k	importance of video packet k , i.e., distortion introduced when discarding k
$MSE_{k,j}$	mean square error introduced on frame j when packet k is lost
γ	percentage threshold in flickering calculation
$h \prec j$	frame h is predicted from frame j
C	frame number for a certain sequence
N	number of wireless nodes in the medium
τ_i	probability that ACi transmitting packets in a generic time slot
$b_{i,r,s}$	probability that ACi in backoff stage r with backoff delayer is s
$p_{f,i}$	the channel idle probability in deferring period between $AIFS_{min}$ and $AIFS_i$
$p_{b,i}$	the channel idle probability in a time slot after the AIFS period of ACi
d_i	the difference in time slot number between $AIFS_i$ and $AIFS_{min}$.
$p_t(N)$	probability that ACi senses the wireless medium is busy
$p_{c,i}(N)$	probability that ACi encounters collision
$p_{e,i}(N)$	probability that packet in ACi is discarded since retry limit is reached
τ	probability that a node gets access to the medium during a generic time slot
$p_{s,i}(N)$	probability that ACi can successfully transmit
$p_c(N)$	probability that one node encounters collision
$p_s(N)$	probability that one node can successfully transmit
$p_b(N)$	probability that one node is busy
$S_i(N)$	normalized throughput of ACi
$R(N)$	ratio of normalized throughput between $AC2$ and $AC1$
r_k	distortion to size ratio for packet k
T	scheduled time for transmitting one GOP
K	interface queue size
l_k	length of packet k
t_k	time allocated to transmit packet k
λ_i	application rate for ACi
η_i	service rate for ACi
ρ_i	traffic intensity for ACi
m	increment of UDP packets scheduled to transmit in current GOP
M	UDP packets scheduled to transmit in a single GOP
T_k	waiting interval for packet k
l_{ACi}	length of packet scheduled in ACi

current frame in terms of both spatial and temporal scalabilities. To be specific, distorted frames include the current frame and the corresponding ones in higher spatial levels, and their respective descendant frames as in the temporal hierarchical structure. In addition, SVC uses a

Figure 3.2: Multilayer encoding structure in SVC with inter-layer prediction. (a) Prediction structure for spatial scalable coding (based on Fig. 4, page 1109 of [10]). (b) Prediction structure for quality scalable coding (based on Fig. 8, page 1113 of [10]).

key-picture concept [10] to realize medium-grain quality scalability. Fig. 3.2(b) illustrates the combination of the hierarchical prediction structures and the key-picture concept. The loss of the coarsest temporal layer frame, which serves as the key picture, will lead to distortion on all subsequent frames before the next key picture. In other packet losses, the distortion will spread to all hierarchically derived frames at both lower and higher quality layers.

3.2.2 MSE based packet prioritization

In video streaming applications, generally the original video sequence may not be available at the server; what in hand is the encoded bitstream. If the original video sequence is available, we could directly calculate the quality degradation caused by each individual packet loss. An accurate estimation of the loss influence is necessary to provide an optimal measure of video packet priority. To avoid the situation that the original video sequence is not available at the sender node for many real applications, we calculate the distortion of packet loss by comparing with the video sequence reconstructed without any packet loss, which could be easily obtained by decoding the compressed bitstream. We resort to the mean square error (MSE) to estimate the distortion introduced by each individual packet loss. If packet k belongs to frame j in one GOP, it will generate distortion not only on frame j itself but also on those frames predicted from j . We formulate the induced MSE due to the loss of a packet k as:

$$\Delta_k = MSE_{k,j} + \sum_{h \prec j} MSE_{k,h} \quad (3.1)$$

where $MSE_{k,j}$ is the MSE value of frame j between the reconstructed video with and w/o the loss of packet k ; $h \prec j$ means frame h is predicted from frame j . The last item represents the MSE summation of all enhancement layers and subsequent frames that are induced due to error propagation. Contaminated frame h could be selected according to the prediction relationship illustrated previously. As only a few frames will be compared in distortion measurement, the extra computational demand is limited and the measure can be additionally acquired during the encoding process.

3.2.3 Flickering based packet prioritization

In a time-varying wireless environment with packets loss inevitable, another critical measure for estimating received video quality is flickering. This artefact is visible by the significantly extended differences of reconstructed values relative to pixel differences, which were originally small between two adjacent frames. Flickering is quite apparent especially in high coding efficiency standards, where loss of bitstream will give rise to significant loss of information on which temporal coherence is maintained. What matters is that the influence will spread to other pictures because of error propagation.

Fan *et. al* [121] defined a flickering measure between successive intra frames. It calculates the extended differences of two adjacent frames which are intra coded compared to the original video sequences. The mechanism that introduces flickering is different but the objective examined is the same, so we improve the measure to accommodate its efficacy for wireless video. Without the original video, we prioritize packets according to introduced flickers once one is lost. Let $O(j, z)$ indicate z th macro block of the j th frame in the reconstructed video without packet loss, and $P(j, z)$ be the reconstructed one where packet k is lost. $O(j+1, z) - O(j, z)$ and $P(j+1, z) - P(j, z)$ denote lossless and lossy reconstructed video frame difference between two successive frames, j and $j+1$. Suppose $SAD(O(j, z), O(j+1, z))$ across all macro blocks, from first to second-to-last frame, are samples of a folded normal distribution with cumulative distribution function as $F(x)$. The flickering measure is defined as:

$$\Delta_k = \frac{\text{AVG}}{j, z(SAD(O(j, z), O(j+1, z)) < F^{-1}(\gamma))} SAD(P(j+1, z) - P(j, z), O(j+1, z) - O(j, z)) \quad (3.2)$$

where γ is a percentage threshold to ensure that only small difference in collocated macro block that is extended after reconstruction is considered. It should be noted that the threshold is difficult to be decided [121] as it is content dependent. High motion video should have large

threshold to ensure serviceability, whereas low motion video should have small threshold so as to avoid over measurement. Here we present a probability model for $SAD(O(j, z), O(j + 1, z))$ to ensure its usability for different videos.

Suppose there are C frames in the reconstructed video without packet loss. We calculate $SAD(O(j, z), O(j + 1, z))$ for each individual macro block in every frame and across the whole sequence from frame 1 to $C - 1$. Denote these values as x_a , where a is sequence index and the total number is denoted as n . Suppose that the difference of two collocated pixels values in successive frames appears randomly. We formulate values of x_a as samples of random variable X , which is from a folded normal distribution. Based on sample values, we have unbiased estimation of mean and variance for X , denoted as μ_f and σ_f^2 , respectively.

$$\mu_f = \tilde{x} = \frac{1}{n} \sum_{a=1}^n x_a \quad (3.3)$$

$$\sigma_f^2 = \tilde{s}^2 = \frac{1}{n-1} \sum_{a=1}^n (x_a - \tilde{x})^2 \quad (3.4)$$

Regardless of the motion activities, the lowest percentage γ of x_a samples are considered to be potential ones generating flickering. With μ_f and σ_f estimated, $F^{-1}(\gamma)$ could be obtained directly with reference to Table II in [122]. This design ensures that for different videos regardless of their contents and motion sets, we can always set a percentage region such that macro block difference within lowest bound γ is calculated to estimate the flickering. In our simulations, we set $\gamma = 0.1$ and find it is effective in reflecting the flickering artefacts.

3.3 Packet Mapping and Scheduling

3.3.1 Analytical Model of EDCA

We consider the SVC packet scheduling issue in MAC layer. We select the IEEE 802.11e EDCA as an example for WLAN because it is specially designed for multimedia transmission. An overview of EDCA background has been given in Section 2.1.3. We describe in this subsection an analytical model of the EDCA under a saturated traffic condition, which is a representative case of concern for scalable video transmission. The adopted model provides quantitative results of the throughput comparison between different ACs, which demonstrates the performance differentiation due to the differences of parameters among four ACs. Based on this, we can differentiate payload traffic to high and low priority queues so as to maximize

3. PACKET PRIORITIZATION AND SCHEDULING FOR SCALABLE VIDEO STREAMING OVER WLANS

the performance. The results are referred to as the Markov Chain model constructed for EDCA in [123], [124] and [125]. For time-varying wireless networks, we consider throughput ratios between different ACs as a function of the wireless node number. The parameters of the ACs used for video streaming in this chapter are listed in Table 2.1.

Assume that there are N active nodes in a WLAN. For any given node, AC3 has a higher priority, followed by AC2, AC1, AC0 with lower priorities. Denote $AIFSN[i]$ as the AIFS number of AC i . Let $CW_{\min}[i]$ and $CW_{\max}[i]$ indicate the $CW_{\min}$ and the $CW_{\max}$ of AC i , respectively. Denote $RL[i]$ the maximum retry limit for AC i . According to the EDCA backoff scheme, the value of backoff counter is chosen randomly in the range of $[0, CW_{i,r} - 1]$, where $CW_{i,r}$ is the size of the contention window of AC i in the r th retransmission. With $CW_{i,0} = CW_{\min}[i]$, the $CW_{i,r}$ of EDCA increases with the number of retransmission times as follows:

$$CW_{i,r} = \min(PF^r * CW_{i,0}, CW_{\max}[i]) \quad (3.5)$$

where PF is set to the default value of 2 in our work.

Let τ_i be the probability that AC i of a wireless node is transmitting a packet in a generic time slot and p_i be the probability that AC i senses the wireless medium busy due to both internal and external collisions. According to the Markov chain model described in [123], we have:

$$\tau_i = \sum_{r=0}^{RL[i]} b_{i,r,0} = b_{i,0,0} \frac{1 - p_i^{1+RL[i]}}{1 - p_i} \quad (3.6)$$

where $b_{i,r,s}$ is the stationary distribution, i stands for AC i , and r is the backoff stage assuming value in the region of $(0, 1, \dots, RL[i])$, s is the backoff delayer taking value from set $(0, 1, \dots, CW_{i,r} - 1)$. So here $b_{i,r,s}$ is the probability that AC i stays in backoff stage r with backoff delayer s . The initial value $b_{i,0,0}$ based on stationary distribution of the Markov model is [124]:

$$b_{i,0,0} = \frac{1}{\frac{1 - p_{f,i}^{d_i}}{(1 - p_{f,i})p_{f,i}^{d_i}} ((1 - p_{b,i}) \sum_{r=0}^{RL[i]} \frac{CW_{i,r}-1}{2} p_i^r + \frac{1 - p_i^{RL[i]+1}}{1 - p_i}) + \sum_{r=0}^{RL[i]} \frac{CW_{i,r}-1}{2} p_i^r + \frac{1 - p_i^{RL[i]+1}}{1 - p_i}} \quad (3.7)$$

where $p_{f,i} = \prod_{t=i}^3 (1 - \tau_t)^N$ is the channel idle probability during a timeslot in the deferring period between $AIFS_{min}$ and $AIFS_i$; $p_{b,i} = \prod_{t=0}^3 (1 - \tau_t)^N$ is the probability that the channel is idle during a timeslot after the AIFS period of AC i ; and $d_i = AIFSN_i - AIFSN_{min}$ is the difference in time slot number between $AIFS_i$ and $AIFS_{min}$. Under a saturation condition,

with other parameters for every AC predesigned, we have the probability that AC i senses the medium busy as a function of the number of the wireless nodes in the medium:

$$p_i(N) = 1 - \prod_{t=0}^3 \frac{(1 - \tau_t)^N}{1 - \tau_i} \quad (3.8)$$

Equations (3.5) to (3.8) form a nonlinear equation set, which can be solved through numerical method. Having $p_i(N)$ available for every AC i , the probability that a wireless station gets access to the medium during a generic time slot can be deduced as:

$$\tau = 1 - \prod_{t=0}^3 (1 - \tau_t) \quad (3.9)$$

The probability that AC i experiences a collision both from the other AC in the same node and that from other nodes is:

$$p_{c,i}(N) = 1 - (1 - \tau)^{N-1} \prod_{i < t \leq 3} (1 - \tau_t) \quad (3.10)$$

When collision happens, AC i starts a new sensing process and tries to access to the medium for another time. This process will only stop when a successful transmission is acquired or the retransmission limitation is reached. The probability that a packet in AC i is discarded because the retry limit $RL[i]$ is reached could be deduced as:

$$p_{e,i}(N) = p_{c,i}(N)^{RL[i]+1} \quad (3.11)$$

Since ACs of all nodes in the medium compete for the channel access, collisions will encumber transmission from some ACs. The probability that AC i in one node can successfully transmit its packets is:

$$p_{s,i}(N) = \tau_i N (1 - p_{c,i}) \quad (3.12)$$

From the view of a wireless node, the probability that this node suffers a collision should be $p_c(N) = \sum_{t=0}^3 p_{c,t}(N)$, and the probability that this node can successfully transmit is $p_s(N) = \sum_{t=0}^3 p_{s,t}(N)$. Consequently, the probability that one node is busy is:

$$p_b(N) = 1 - (1 - \tau)^N \quad (3.13)$$

With the collision and busy probability of every AC deduced, we can normalize the throughput of AC i as:

$$S_i(N) = \frac{E(P_i)}{E(L_i)} = \frac{p_{s,i}(N)E[p]}{1 - p_b(N)\delta + p_s(N)T_s + p_c(N)T_c} \quad (3.14)$$

3. PACKET PRIORITIZATION AND SCHEDULING FOR SCALABLE VIDEO STREAMING OVER WLANS

Table 3.2: The IEEE 802.11e Duration Constants [35]

Option	Value
δ	$20\mu s$
T_s	$E[p] + 131.8182\mu s$
T_c	$E[p] + 93.1818\mu s$

where $E(P_i)$ is the average payload transmission time and $E(L_i)$ is the expected length. Parameters δ , $E[p]$, T_s and T_c are the timeslot duration, the average frame payload duration, the successful transmission time, and the time of collision, respectively [126]. Duration constants for the basic mode of 802.11e are summarized in Table 3.2. Furthermore, $E[p]$ can be deduced as in [126]:

$$E[p] = \int_0^{p_{\max}} (1 - F(x)^2) dx \quad (3.15)$$

where $F(x)$ is the probability distribution function of the packet payload size.

The IEEE 802.11e standard recommends the use of four ACs for different kinds of packets. The recommendation renders packets in different priority the opportunity to finish their service goal. To conform to our application and relieve the heavy load of multimedia traffic on the entire wireless network, we follow a setting that is adopted in [35]. We send video streams via two of the four ACs, i.e., AC2, which is recommended for video traffics, and AC1, which is suitable for best effort traffics. Important packets are sent through AC2 and the rest are sent through AC1 in a best-effort manner, so that a higher protection is provided for more important packets. This adoption could reduce the network loading as some video packets are served in best-effort manner and the feature of SVC that is easy for bitstream manipulation under various network conditions can be exploited. In formulating our proposed approach, we analyze the normalized throughput of two ACs so that we can allocate video bitstreams according to channel capacities to maximize the overall throughput. Since we only employ two ACs, AC2 and AC1, the abovementioned parameters should be revised accordingly in the deduction to conform to the case of two ACs. In addition, we use $R(N)$ to denote the ratio of normalized throughput between AC2 and AC1:

$$R(N) = \frac{S_2(N)}{S_1(N)} \quad (3.16)$$

The nonlinear equation set is solved beforehand to acquire a look-up table of $R(N)$ for different N . Based on this table, the packets are allocated into the two ACs accordingly. This

Figure 3.3: The throughput ratio between AC2 and AC1 vs. the node number.

method exploits the transmission capacities of two ACs and makes adjustment adaptively to maximize overall throughput. Fig. 3.3 illustrates the throughput ratio between AC2 and AC1 versus the node number under the network settings as listed in Tables 2.1 and 3.2.

3.3.2 QoS Mapping

Based on the analytical model of EDCA and packet priority analysis as described above, we formulate an optimal QoS mapping strategy to allocate SVC packets into two ACs according to the distortion measure of packet loss and throughput differentiation between the ACs.

Assume that the length of the video packet k is l_k . The packets in the current GOP, belonging to a set $\Phi = \Phi_1 \cup \Phi_2$, are classified exclusively into two sets Φ_1 and Φ_2 , distributed into AC1 and AC2, respectively. Our strategy of packet allocation into two ACs to minimize packet dropping due to exceeding of the maximum retry limit is formulated as:

$$\min_{a,b} \left(\sum_{a \in \Phi_1} \Delta_a p_{e,1}(N) + \sum_{b \in \Phi_2} \Delta_b p_{e,2}(N) \right) \text{ s.t. } \left| \sum_{b \in \Phi_2} l_b - R(N) \sum_{a \in \Phi_1} l_a \right| \leq \Gamma \quad (3.17)$$

where $p_{e,1}(N)$ and $p_{e,2}(N)$ are the packet dropping probabilities for AC1 and AC2, respectively. The constraint is to guarantee that the ratio of the total packet lengths within AC2 and AC1 conforming to the theoretical throughput ratio. As the packet length is discrete and packet number is finite, the ratio could not be strict $R(N)$ so an inequality is used. Let $\Gamma = \max\{l_1, l_2, \dots, l_t\}$ be the largest packet size within the current GOP. The problem could be

3. PACKET PRIORITIZATION AND SCHEDULING FOR SCALABLE VIDEO STREAMING OVER WLANS

further deduced as:

$$\begin{aligned}
& \min_{a,b} \left(\sum_{a \in \Phi_1} \Delta_a p_{e,1} + \sum_{b \in \Phi_2} \Delta_b p_{e,2} \right) \\
&= \min_b \left(\sum_{k \in \Phi} \Delta_k p_{e,1} + \sum_{b \in \Phi_2} \Delta_b (p_{e,2} - p_{e,1}) \right) \\
&= p_{e,1} \sum_{k \in \Phi} \Delta_k + \min_b (p_{e,2} - p_{e,1}) \sum_{b \in \Phi_2} \Delta_b \quad \text{s.t.} \quad \left| \sum_{b \in \Phi_2} l_b - R(N) \sum_{a \in \Phi_1} l_a \right| \leq \Gamma
\end{aligned} \tag{3.18}$$

For a coming GOP, parameters including $p_{e,1}(N)$, $p_{e,2}(N)$ and packet length of the GOP $\sum_{k \in \Phi} l_k$ are all pre-decided and thus irrelevant to the minimization. Since $p_{e,1} > p_{e,2}$, the optimization problem of (3.17) is equivalent to selecting a group of packets for Φ_2 from set Φ such that

$$\max_b \sum_{b \in \Phi_2} \Delta_b \quad \text{s.t.} \quad \sum_{b \in \Phi_2} l_b \leq L_2 \tag{3.19}$$

where L_2 should be approximately equal to $\text{ceil}(\sum_{k \in \Phi} l_k R(N) / (R(N) + 1)) + \Gamma$, and the constraint is relaxed to an inequality constraint since the equality may not hold exactly in practice. The problem (3.17) is equivalent to the classic knapsack problem.

The solution to the maximization problem can be obtained by resorting to dynamic programming. Suppose that we have n packets in the current GOP, 1 through n ; define $A(i, L)$ to be the maximum value that can be obtained with total length less than or equal to L using items up to i . $A(i, L)$ can be derived recursively from $i = 0$ and $L = 0$ as follows:

$$A(i, L) = \begin{cases} 0 & \text{if } i = 0 \\ 0 & \text{if } L = 0 \\ A(i-1, L) & \text{if } l_i > L \\ \max\{A(i-1, L), A(i-1, L - l_i) + \Delta_i\} & \text{if } l_i \leq L \end{cases} \tag{3.20}$$

The optimal solution can be found by calculating $A(n, L_2)$. Besides the optimal solution by dynamic programming for the knapsack problem, other heuristic yet efficient solutions are also available. Since we can only guarantee an approximate ratio of $R(N)$ as there are a finite number of packets within a GOP, we do not need to resort to dynamic programming solution which is computationally intensive. Instead, we obtain the results from a greedy algorithm, which is inspired by the greedy algorithm for a knapsack problem [127]. To ensure an approximate ratio of $R(N)$ between two ACs and only a small number of packets in one GOP are under consideration, the greedy algorithm is good enough for the problem at hand. That is, Δ_k

and l_k are considered as the value and weight in the knapsack problem, respectively. First for all packets within the coming GOP, we deduce the distortion to size ratio for packet k as:

$$r_k = \frac{\Delta_k}{l_k} \quad (3.21)$$

We prioritize the packets in descending order of r_k , and partition them into two sets at the point where the preceding packets that are scheduled for AC2 have a total length that is $R(N)$ times the remaining ones which are scheduled for AC1. This heuristic solution can significantly reduce the complexity at the expense of only small performance loss.

The QoS mapping strategy at the MAC layer will be combined with interface queue control in the next subsection to form our unified packet scheduling scheme.

3.3.3 Queue Control and Packet Scheduling

We have formulated an ordered sequence of packets to decide, for an in-coming GOP, how video packets should be allocated into two access categories to improve the overall performance. The formulation is conducted only based on the intrinsic attributes of EDCA. We now formulate a minimization problem for video packet dropping due to insufficient network bandwidth. Excessive packets will lead to severe packet loss because of interface queue (IFQ) overflow at the link layer, which will result in significant quality degradation. We aim to determine the number of packets that should be scheduled to match the network capacity and hence to maximize the end-to-end quality for end users. For packets scheduled into AC i ($i = 1, 2$), we assume that the time slot with length T is allocated for transmission. In insufficient network where excessive packets are scheduled into the queue, the scheduling of video packet k (that is scheduled to a certain AC i) to avoid IFQ overflow is generalized as follows:

$$\min_{k_1, k_2} \left\{ \sum_{k_1 \in \Theta_1} \Delta_{k_1} p_{d,k_1,i} + \sum_{k_2 \in \Theta_2} \Delta_{k_2} \right\} \quad \text{s.t.} \quad \sum_{k_1 \in \Theta_1} t_{k_1} \leq T \quad (3.22)$$

where Δ_k , $p_{d,k,i}$ and t_k are distortion introduced if packet k is lost, packet dropping probability, and time interval for sending video packet k , respectively. Sets Θ_1 and Θ_2 contain packets that are scheduled into queue and discarded beforehand, respectively. Set $\Theta = \Theta_1 \cup \Theta_2$ contains all packets that are scheduled into the two ACs.

Once the allocation strategy is determined, the IFQ dropping probability $p_{d,k,i}$ can be derived based on the packet arrival and departure models. Suppose the simple case that the network condition is constant during the scheduling period, packets arriving in a constant rate

3. PACKET PRIORITIZATION AND SCHEDULING FOR SCALABLE VIDEO STREAMING OVER WLANS

could to a great degree minimize the queue overflow. Here we assume the packets arrive and depart evenly in time. As $p_{d,k,i}$ is irrelevant to specific packet k , we discard the subscript k in the following discussion. For a certain number of packets, we formulate packets arriving to AC i as a Poisson Process with application rate λ_i , and the state transition in the queue of AC i as $M/M/1/K$, where K is the maximum size of the queue for AC i . Denote η_i as the service rate of AC i . Hence, the traffic intensity of AC i is:

$$\rho_i = \frac{\lambda_i}{\eta_i} \quad (3.23)$$

To investigate packet overflow in a congested queue where excessive packets are ready for transmission, we have $\rho_i > 1$. It follows that the probability of overflow for queue AC i is:

$$p_{d,i} = \frac{\rho_i^K (1 - \rho_i)}{1 - \rho_i^{K+1}} = 1 - \frac{1}{\rho_i} + \frac{1}{\rho_i^{K+1}} + \dots \approx 1 - \frac{1}{\rho_i} \quad (3.24)$$

Hence, the problem in (3.22) can be deduced as

$$\begin{aligned} & \min_{k_1, k_2} \left\{ \sum_{k_1 \in \Theta_1} \Delta_{k_1} \left(1 - \frac{1}{\rho_i}\right) + \sum_{k_2 \in \Theta_2} \Delta_{k_2} \right\} \\ & = \min_{k_1} \left\{ \sum_{k \in \Theta} \Delta_k - \sum_{k_1 \in \Theta_1} \Delta_{k_1} \frac{1}{\rho_i} \right\} \quad \text{s.t.} \quad \begin{cases} \sum_{k_1 \in \Theta_1} t_{k_1} \leq T \\ \rho_i > 1 \end{cases} \end{aligned} \quad (3.25)$$

Eq. (3.25) can be shown to be equivalent to

$$\max_{k_1} \left\{ \sum_{k_1 \in \Theta_1} \Delta_{k_1} \frac{1}{\rho_i} \right\} \quad \text{s.t.} \quad \begin{cases} \sum_{k_1 \in \Theta_1} t_{k_1} \leq T \\ \rho_i > 1 \end{cases} \quad (3.26)$$

$$\iff \max_{k_1} \left\{ \frac{\sum_{k_1 \in \Theta_1} \Delta_{k_1}}{\sum_{k_1 \in \Theta_1} l_{k_1}} \right\} \quad \text{s.t.} \quad \sum_{k_1 \in \Theta_1} l_{k_1} > T \eta_i \quad (3.27)$$

A straightforward solution to Eq. (3.27) is to select packets in descending order of $\frac{\Delta_k}{l_k}$ until the constraint is satisfied. However, the goal of minimizing packet overflow is achieved in a single queue of individual AC. To combine the EDCA model based packet allocation, we propose the following solution to jointly optimize the system performance for two ACs. This gives the whole structure of our packet scheduling scheme.

It has been shown in [128] that the TCP transmission is more suitable for home wireless. Here, we follow most of the research work using UDP in the transport layer. Nevertheless, the packet prioritization, packet mapping and IFQ control mechanisms could also be applied

in TCP transmission. The only difference is for TCP transmission, TCP ACK should be transmitted in the highest priority AC to guarantee the accurate acknowledgement. Denote the summation of the average departure rate of the queues for AC2 and AC1 within the previous GOP as d , where d is measured as UDP packets departing in a unit time and is updated GOP by GOP. T is the transmission time for the current GOP. If the departure rate is assumed to be constant in successive GOPs, the amount of UDP packets which should be sent in the current GOP is $\text{ceil}(Td)$. To avoid underestimating the network capacity thus affecting the utilization, we schedule $M = \text{ceil}(Td) + m$ UDP packets for transmission with a small increment m .

Based on the priority analysis for individual video packet, we have a descending ordered packet sequence. We truncate this sequence to keep the preceding M UDP packets and the others are discarded to avoid network congestion. The reserved packets for transmission are allocated into two ACs according to the ratio $R(N)$. With the descending ordered sequence derived according to r_k , discarding packets at the end of the whole sequence and distributing packets into ACs with ratio $R(N)$ will be superior than allocating the packets first then discarding packets at the end of each sequence in the two ACs. The number of UDP packets kept in each queue are equal but more important packets are reserved for transmission.

For each packet, a waiting interval T_k following the previously sent video packet is calculated as:

$$T_k = \frac{Tl_k}{l_{ACi}} \quad (3.28)$$

where l_{ACi} is the the summation of length of all packets reserved in the AC in which packet k resides. The regulation realizes an even arrival of video packets to conform to the Poisson assumption. Moreover, this design serves like a leaky-bucket traffic shaper [129], which regulates the packet arrival in an approximately constant rate. The increment m packets in this sense are a means to combat the inherent inefficiency of the leaky-bucket implementation [130].

The proposed scheme introduces a delay at most one GOP, which is acceptable and applicable in a video streaming scenario. In addition, we do not start discarding video packets until the IFQ has been occupied half or more. As a summary, our proposed cross-layer optimization scheme, including the packet prioritization and packet scheduling algorithms, is illustrated in Table 3.3.

Table 3.3: The Proposed Packet Prioritization and Scheduling Scheme

```

procedure ComputeLossFactor ( $\Delta_k$ )
    discarding  $k$ 
    calculate  $\Delta_k$ 
return when(all packets are traversed)
    Order packets in the current GOP in descending  $\Delta_k$ 

procedure ComputeTruncatingPoint ( $tp$ )
     $L_{tp} += L_k$ 
return max  $k$  such that  $L_{tp} \leq \text{ceil}(Td) + m$ 

procedure ComputeCutPoint ( $cp$ )
     $L_{cp} += L_k$ 
return max  $k$  such that  $L_{cp} \leq R(N)L/(R(N) + 1)$ 

procedure  $k$  from start
    waiting time for  $k$  is  $l_k/l_{AC2}$ 
return when( $k$  up to  $cp$ )
procedure  $k$  from  $cp + 1$ 
    waiting time for  $k$  is  $l_k/l_{AC1}$ 
return when( $k$  up to  $tp$ )

```

3.4 Simulation Results

3.4.1 Simulation Settings

In a typical wireless network, several nodes compete to transmit various traffics. In order to examine the performance of our strategy, we used twelve nodes, six sources and six sinks. In each node, AC2 and AC1 were employed to transmit video. For the six source nodes, four of them served with constant bit rate (CBR) sources, and the other two were used to transmit SVC bitstreams. The CBR nodes were employed as contenders for the wireless medium. The two video source nodes were designed to transmit CBR traffic at the beginning, and after five seconds, they started to transmit videos.

Our simulations were conducted on Network Simulator 2.33 (NS-2.33) [131]. The implementation of EDCA of 802.11e for NS-2 was based on the model developed by Telecommunication Networks Group at Technical University of Berlin [132]. This EDCA model was modified to add the retry limit differentiation among ACs such that AC2 had a higher prior-

ity to transmit more important video packets. Wireless nodes were placed in a 250 meters by 250 meters square area to communicate with each other. The IEEE 802.11e settings had the parameters listed in Table 2.1. The intrinsic of hierarchical coding structure resulted in a large variation in the SVC packet length, where packets from different SVC layers were different in length up to ten times or more. Video packets were partitioned into UDP packets and then delivered to subsequent network layers. The consequence was that a large number of UDP packets were scheduled for transmission at the same time. To provide sufficient buffering for bursty video traffic, we increased the default length of IFQ from 50 to 500 packets.

As for SVC codec, we used JSVM 9.8 [133]. Two one-minute video sequences, Bus and Foreman, were encoded and transmitted over the IEEE 802.11e wireless network. To make fair comparisons with different algorithms proposed in [116, 35, 117], we employed different coding structures in different comparisons. Details are given in corresponding experiments. SVC node pair 0 transmits and receives video Bus and node pair 1 is for video Foreman. We first show the comparisons of our model and reference models on traditional PSNR evaluation in Section 3.4.2. In addition, we evaluate the performance of the flickering based algorithm in Section 3.4.3.

3.4.2 Performance Evaluation for MSE Based Prioritization and Scheduling

We show the end-to-end quality performance when the MSE based criterion is adopted. Comparisons are conducted with three reference algorithms [35, 116, 117]. Within each comparison, the network parameters are identical. PSNR is in agreement with MSE in comparing video quality and meanwhile it is widely used to assess visual quality and evaluate system performance. Therefore, we present PSNR comparison of received videos in the MSE based packet prioritization simulations.

3.4.2.1 Quality Verification I

We compare first with the one proposed in [35]. In [35], a wavelet based scalable video coding scheme was adopted. An adaptive QoS mapping algorithm with micro control was proposed to improve the performance. But packets were allocated into AC2 and AC1 with a constant ratio of 2:1. As shown in the analytical model of EDCA in Section 3.3.1, the normalized throughput ratio of AC2 and AC1 should be function of the node number N in the wireless environment, owing to various load conditions. To show that the packet allocation according to $R(N) =$

3. PACKET PRIORITIZATION AND SCHEDULING FOR SCALABLE VIDEO STREAMING OVER WLANS

Table 3.4: Video Quality (dB) obtained by using constant and adaptive allocation ratio among two ACs

No. of CBR nodes	4	6	8	10	20	30	40	50
R(N)	31.93	31.88	31.57	31.16	29.86	29.53	28.62	27.71
2/1	31.90	31.73	31.46	31.00	28.94	27.67	26.99	26.41

$S_2(N)/S_1(N)$ could improve the system performance, we first designed a simple experiment to validate it. Only one sequence BUS was encoded and transmitted in an environment where different number of CBR nodes served as competing traffics. Video was encoded with three kinds of scalabilities and a total of 1800 frames were encoded. The experimental results are listed in Table 3.4.

With different numbers of CBR nodes employed, the total transmitted data are different. On the other hand, keeping the total transmission rate constant will result in rate fluctuation of a single CBR node. Therefore, the network settings under different N were different. We present the results in Table 3.4 to show that the adaptive regulation scheme always outperform the constant 2:1 strategy when any number of wireless nodes competed for the medium simultaneously. To avoid the influence from other factors, the interface queue control was not employed in this simulation. We could see that when the number N is small, $R(N)$ is near to the value of 2:1, hence the difference between the adaptive and constant allocation scheme is negligible. When a large number of nodes communicated simultaneously, the fierce competition resulted in severe packet dropping in AC1 so the throughput ratio was much larger than 2:1. In this case the IFQ control scheme could command more performance improvement.

Next, we compare to the reference strategy in [35]. Since the codec used in [35] are different, we compare specific aspects of QoS mapping strategy for SVC packet allocation. Specifically, we encoded Bus and Foreman by SVC with all three kinds of scalabilities. We present the average PSNR at different data rates by comparing two reference methods based on [35], denoted as *Ref1* and *Ref2* in Fig. 3.4. In both *Ref1* and *Ref2*, the proposed packet prioritization strategy was used since the algorithm proposed in [35] is based on a different codec. In *Ref1*, the mapping ratio 2 : 1 suggested in [35] was used and IFQ control was adopted. In *Ref2*, the packet allocation ratio was kept at 2:1 for two ACs but our proposed adaptive QoS mapping strategy was adopted. The comparison is to validate the performance of our proposed mapping ratio $R(N)$. To validate the performance of EDCA based allocation and IFQ control separately, the packets were prioritized and allocated with ratio of $R(N)$, but no IFQ control

Figure 3.4: Average PSNR comparison with the reference scheme in [35]. (a) Bus Sequence received at node pair 0. (b) Foreman Sequence received at node pair 1. *Ref1*: packet prioritization, 2:1 allocation without IFQ control. *Ref2*: packet prioritization, 2:1 allocation with IFQ control. *Our1*: packet prioritization, $R(N)$ allocation without IFQ control. *Our2*: packet prioritization, $R(N)$ allocation with IFQ control.

was implemented in our proposed scheme *Our1*. In scheme *Our2*, the packet prioritization and QoS mapping strategies were jointly employed.

By comparing with *Ref1*, we demonstrate the outperformance of our packet scheduling strategy. Firstly, without considering the IFQ overflow, the reference scheme led to unpredictable packet loss due to bursty video traffic. The allocation ratio for two ACs at fixed 2:1 cannot optimize the overall throughput under a time-varying wireless environment as packets may be overly scheduled into a particular AC. In addition, adaptively regulating the sending time and the number of packets in our scheme can maintain a smooth transmission and hence reduce the quality fluctuation. The reference model lacked of such mechanisms in maintaining a stable video quality. Especially for large video packets with great length variations like Foreman, the advantage of our proposed regulation is notable. The PSNR gains of our proposed scheme *Our2* over *Ref1* are 3.65 dB and 5.33 dB for two video sequences, respectively, as shown in Fig. 3.7.

Ref2 outperformed *Ref1* as the adaptive QoS mapping strategy was adopted. However, the comparison shows that a constant 2:1 is inferior to the allocation with accurate throughput ratio, when compared to our proposed algorithm. Hence, the service ratio between AC2 and AC1 should be adaptive to exploit the transmission capacities of ACs. Otherwise, some

3. PACKET PRIORITIZATION AND SCHEDULING FOR SCALABLE VIDEO STREAMING OVER WLANS

important packets could be scheduled into low priority AC1 and may be discarded due to the smaller maximum retry limit or queue buffer overflow. In the reference scheme [35], packets scheduled into AC2 are maintained at a lower ratio to provide adequate protection for packets in AC2 by avoiding buffer overflow. However, increasing payload in AC2 and adaptively regulating the number of packets according to network conditions can improve the overall system performance, as demonstrated in the results obtained by the proposed approach. It should be noted that at the data rate of $11Mbps$, *Ref2* achieves a close performance compared to *Our2*. This is because with the increase of data rate, the negative impact of the biased allocation in *Ref2* is reduced. Moreover, we are aiming at increasing overall quality of all transmitted videos. Considering together with the quality of Bus, *Our2* can always perform better than all the reference schemes. Fig. 3.8 shows some video frames obtained by the reference algorithm *Ref2* and our proposed *Our2* scheme at the data rate of $9Mbps$.

The improvement can be observed by comparing *Our1* and *Ref1*, where packet allocation ratio between AC2 and AC1 was based on $R(N)$ rather than 2:1. The improvement comes from the fact that packet allocation according to $R(N)$ could optimally exploit the transmission capacity of each AC under various load conditions. The performance of *Ref2* is also generally better than *Our1*, because the IFQ control contributes more than EDCA based allocation in such a setting where excessive packets are transmitted. The network bandwidth is not sufficient to support all transmitted traffics, including SVC and CBR. Too many packets only lead to severe network degradation. The IFQ control could efficiently regulate the number of scheduled packets to adapt to the network capacity. Meanwhile, according to our discarding-less-important-packet-first scheme, the more important packets are transmitted in a less congested network thus they enjoy a higher opportunity to reach the receivers. Hence, the system performance is significantly improved.

The proposed scheme *Our2* performs the best among the four schemes because of the simultaneous adoption of packet prioritization, EDCA based allocation and IFQ control. Our proposed scheme is also more practical for scalable video transmission in wireless networks than [35] because only one-time encoding is supported in our setting. Once video is encoded and stored at the sender, the rate adaptation could be achieved whenever necessary and under different network conditions. In contrast, the method in [35] involves a micro control that feedbacks the network condition and the bitrate is regulated concurrently in the encoding and transmission processes. Comparatively, a precoded and stored bitstream could be reused

Figure 3.5: Average PSNR comparison with the reference scheme in [116]. (a) Bus Sequence received at node pair 0. (b) Foreman Sequence received at node pair 1.

multiple times when requests come, which is cost-saving and more conforms to the design of scalable video for easy adaptation of the bitstream to adapt to various applications.

3.4.2.2 Quality Verification II

We compare the proposed algorithm with the one in [116]. In [116], the spatial base layer packets were considered to be more important than those in enhancement layers. Within one spatial layer, the packet priority was decided based on hierarchical structure and prediction relations. For EDCA, packets were allocated to ACs equally without considering their capacity differences. As in the [116], we encoded two spatial layers and five temporal layers within each spatial layer. The comparison results under four different data rates for sequences Bus and Foreman are shown in Fig. 3.5.

We can find that the equal usage of AC adopted in the reference scheme was inferior in exploiting network capacity when compared to our model-based allocation strategy. Firstly, the high priority AC2 was underutilized. Some relatively important packets could be overly scheduled into AC1 but dropped due to queuing buffer overflow. The loss of these packets resulted in severe quality degradation. The improvement of our scheme comes from cross-layer control of scheduled packets by discarding some least important packets. Discarding some least important packets beforehand could lead to PSNR reduction in some frames, but this loss is only confined within the frame itself or a small number of subsequent frames. In the reference algorithm, the packet priority was solely decided based on prediction relations, which

3. PACKET PRIORITIZATION AND SCHEDULING FOR SCALABLE VIDEO STREAMING OVER WLANS

Figure 3.6: Average PSNR comparison with the reference scheme in [117]. (a) Bus Sequence received at node pair 0. (b) Foreman Sequence received at node pair 1.

is suboptimal since the length information is neglected. Our scheme presents its advantage in handling packets allocation and the adaptive QoS mapping remedies aforementioned problems. The improvement can also be observed in Fig. 3.9, in which subjective video pictures obtained from two schemes at the data rate of 9 Mbps are presented. In our scheme, the loss of least important packets only led to limited quality degradation, but the pictures from the reference scheme contain severe distortion. The average PSNRs were improved from 26.12 dB to 32.31 dB and from 30.11 dB to 38.42 dB for Bus and Foreman sequences, respectively.

3.4.2.3 Quality Verification III

In this part, we compare our proposed algorithm with the one in [117]. The authors in [117] proposed to allocate video packets based on the SVC NAL unit header information, where the quality base layer packets were transmitted through AC2 and the remaining enhancement layer packets were for AC1 and AC0. For fair comparisons, we sent CBR traffic through AC2, AC1 and AC0 but kept the total data rate unchanged. To be identical to the simulation settings in [117], the two video sequences were encoded with a GOP size of 8, and temporal and quality scalabilities were supported. The PSNR results of the received videos at different data rates are shown in Fig. 3.6.

It is clear that allocating packets based on NAL unit type information in [117] can suffer from severe quality degradation, especially in a network with insufficient resources. This is because: 1) packet length information is neglected so the evaluation of packet priority is

not accurate, 2) ACs' throughput differentiation is not considered, and 3) the allocation is not adaptively regulated based on network conditions. These factors give rise to lopsided packet allocation into ACs. One AC may be underutilized but the others may be scheduled too many packets. The heavily occupied AC has to discard packets which exceed their retry limit, leading to massive packets dropping. While in our model, packets are scheduled into two ACs according to their transmission capacities. Moreover, the adaptive regulation by the IFQ control discards least important packets beforehand, to provide adequate protection to more important ones. As a result, the overall PSNR is improved.

Some decoded video pictures obtained from data rate of 9 Mbps are displayed in Fig. 3.10, where the left one is from our proposed scheme and the right one is from the reference scheme. When compared to the reference scheme, our proposed approach achieves 1.21 dB and 2.73 dB gains in average PSNR for Bus and Foreman, respectively. It should be noted that the reference scheme depends highly on the encoding structure and video content. The loss of quality enhancement layer packets will not lead to severe quality degradation, but will cause significant quality degradation for other two scalabilities. The robustness of our proposed strategy can be demonstrated in video sequences with different degrees of motion activities.

3.4.3 Performance Evaluation for Flickering Based Prioritization and Scheduling

We validate the performance of our algorithm in reducing flickering artefacts on the received videos. Network settings were kept identical with previous subsections. We ranked video packets according to their flickering impact, where those lead to larger flicker were given higher priorities for transmission.

Video sequences Bus and Foreman were encoded with two spatial layers. Within each spatial layer, five temporal layers and two quality layers were supported. Since no reference considered flickering reduction for SVC over 802.11e in the literature, we compared our proposed approach with a reference scheme in which neither packet prioritization nor QoS mapping is imposed. The allocation ratio between AC2 and AC1 was set to $R(N)$ according to the EDCA model. Fig. 3.11 shows the results, where we can observe significant reduction of flickering achieved from the proposed scheme. For clear presentation, only the results from frames 800-1000 are presented. PSNR values from the proposed schemes are 29.89 dB and 35.22 dB for sequences Bus and Foreman, respectively. Those from the reference scheme are 19.89 dB and 27.09 dB. By C++ language implemented PSNR and flickering calculator running on Ubuntu

8.04 platform, the computation time for both scheme is close to each other. For the 1800-frame video, it generally consumes around 7-8 seconds to execute the computations.

As illustrated in the flickering measure curves, the proposed scheme presents its advantage in packets allocation. The number of packets deserved for transmission is always determined based on the current network transmission capacity. The least important packets are discarded to release network congestion. It can be found from these figures, flickers maintain at a low level in our scheme through the whole simulation, regardless of network conditions. Secondly, allocating a waiting interval for incoming packets according to their lengths is helpful in reducing IFQ overflow. The merit of this disposal is distinct when bursty video packets arrive intensively. In SVC coding, there is great difference in packet length between key frames and PR frames. Without considering the bursty packet size, packet importance and buffer status, the reference scheme results in large flickering artifacts and significant quality degradation, as has been demonstrated in simulation results. Subject validation on received video quality can be found in Fig. 3.12. It is clear that our packet scheduling strategy achieves better video quality whilst the reference model is not effective for packet allocation and results in annoying flickering artefacts.

By analyzing the simulation results presented in this section, we can conclude that significant improvement can be achieved by our proposed scheme. The proposed approach adequately exploits the characteristics of SVC packets and the differentiated service capacity of EDCA. The interaction among application, link and MAC layers provides sufficient information on payloads and network environments, which is helpful for the implementation of the packet scheduling algorithm. The advantages highlighted in simulation results show its enlightenment for video streaming over the IEEE 802.11e wireless networks.

3.5 Concluding Remarks

We have introduced an efficient scalable video streaming scheme for SVC over the IEEE 802.11e wireless networks based on joint packet prioritization and scheduling. Using an accurate Markov model for IEEE 802.11e MAC protocol, we designed an efficient packet scheduling scheme where three important factors were considered, namely, (i) SVC packet information at the application layer, (ii) differentiated traffic prioritization at the MAC layer, and (iii) interface queue control at the link layer. The interaction of three components provides a novel scheduling strategy for SVC streaming. In addition, we provide SVC packet priority analysis

based on their length and distortion impact. With the cross-layer information, the proposed packet prioritization and scheduling scheme minimizes the packet loss impact on visual quality and maximizes the system performance. Simulation results demonstrate the improvement of our algorithm at different aspects. This study provides a good understanding on the using of SVC for wireless streaming and is useful for further study on traffic characteristics and QoS improvement for video delivery over WLANs.

To have a thorough analysis on the proposed scheme, we discuss several critical parameters here.

- **Delay:** The delay for packet is at most one GOP length that is around 0.5 seconds for a GOP size of 16 and a frame rate of 30 fps; this is reasonable for video streaming in wireless networks. To apply the proposed framework applications which require less delay, we could reduce the GOP size. Reduction in GOP size will lead to reduction in coding efficiency. That is, under a certain network bandwidth, a smaller GOP size will lead to quality degradation, but the delay could be reduced as well.
- **Complexity:** In packet prioritization, the calculation of video distortion is confined to one GOP. With the analysis of packet prediction relations, the complexity is limited and the calculation could be completed off-line. Computations involved in packet allocation in Section 3.3.2 and adaptive regulation of IFQ control in Section 3.3.3 are negligible.
- **Video quality consistency:** Under the variation of wireless network, we always retain the most important packets and discard the less important ones to reduce the network congestion. In this way, a better video quality could be attained and the video artifacts due to network fluctuation could be reduced, as demonstrated in Section 3.4.

Except for these highlights, the proposed scheme has room to be improved in the following aspects. Firstly, the prioritization is achieved by extensive packet-by-packet calculation. Although this can be accomplished off-line before transmission, we still hope for an easier way to handle it. We have discussed the multilayer prediction structure in this chapter to identify dependent relationship in calculating the quality impact of a single packet. By measuring the quality impact, we prioritize packets in a descendent order. A possible extension would be whether we can derive the importance order directly by analyzing the prediction relationship. This will save a great deal of time without calculation. Nevertheless, the disposal should not lead to obvious performance degradation in terms of video quality. This will be the issue to

3. PACKET PRIORITIZATION AND SCHEDULING FOR SCALABLE VIDEO STREAMING OVER WLANS

be explored in Chapter 4. Secondly, we have aimed at maximizing the overall video quality in this chapter. It has always been of research interest that how we can guarantee the fairness or achieve the tradeoff between efficiency and fairness in multiuser scenario. This issue will be addressed in Chapter 5. Chapter 5 will be based on Chapter 4, which elaborates the utility analysis by an enhanced packet prioritization scheme.

Figure 3.7: Comparison between our proposed scheme *Our2* and reference scheme *Ref1*. Left: our proposed scheme, Right: the reference scheme.

Figure 3.8: Comparison between our proposed scheme *Our2* and reference scheme *Ref2*. Left: our proposed scheme, Right: the reference scheme.

3. PACKET PRIORITIZATION AND SCHEDULING FOR SCALABLE VIDEO STREAMING OVER WLANS

Figure 3.9: Comparison with the reference scheme in [116]. Left: our proposed scheme, Right: the reference scheme.

Figure 3.10: Comparison with the reference scheme in [117]. Left: our proposed scheme, Right: the reference scheme.

Figure 3.11: Flickering measure comparison with the reference algorithm in which no packet prioritization and QoS mapping adopted. Top: Sequence Bus of frames 800-1000. Average of the whole sequence: *Our*: 44.1, *Ref*: 758.0. Bottom: Sequence Foreman of frames 800-1000. Average of the whole sequence: *Our*: 36.64, *Ref*: 591.78.

Figure 3.12: Video comparison with no packet prioritization and QoS mapping scheme on flicker reduction. Left: our proposed method, Right: the reference method.

3. PACKET PRIORITIZATION AND SCHEDULING FOR SCALABLE VIDEO STREAMING OVER WLANS

A Fast Rate Adaptation Scheme for SVC Based on the Packet Dependencies

4.1 Introduction

Scalable video packets have inherent priority differentiation in realization of layered coding which enables flexible rate adaptation [134]. However, how to adapt encoded bitstream in terms of the best quality (or the minimal distortion) at a given bitrate still needs to be further exploited. There are some excellent references devoting to this issue but they mainly deal with only partial scalability [135, 136]. The SVC coding standard has proposed two reference methods, the JSVM basic bit extraction [137] and the Quality Layer information based rate adaptation [138]. However, the basic extractor cannot ensure a minimized distortion. The Quality Layer based adaptation needs the availability of original videos to carry out complex computation, and the effective range of rate adaptation is quite limited. A recent study [139] employs the Integer Linear Programming technique to redefine the priority information. The work in [54] proposes a novel distortion modeling method for optimized bit extraction. Since these two methods still involve high computational complexity, we address this problem in a more efficient manner. The improvement of the proposed scheme will be verified in Section 4.3. An enhanced packet prioritization scheme is proposed based on the analysis of packet dependencies (PD) in full scalability, without resorting to the original video sequence. This enhanced method makes

4. A FAST RATE ADAPTATION SCHEME FOR SVC BASED ON THE PACKET DEPENDENCIES

improvement from the prioritization scheme in Chapter 3 where computationally intensive pre-processing is needed. Another work [140] proposed to search for a near optimal rate adaptation and has summarized some criteria. It dedicates to heterogeneous resolution adaptation and is different from ours. We consider a more common case where the target resolution is decided by terminals and we want to decide for a specific spatial resolution, how to optimize the rate adaptation in order for the maximal video quality. Experimental results demonstrate that the proposed scheme achieves significant improvement over the basic bit extraction scheme in SVC and attains comparable performance to the Quality Layer based rate adaptation scheme with the significant deduction on the computational cost. This part of work has been published in our paper [141].

4.2 Fast Rate Adaptation Based on Enhanced Packet Prioritization

4.2.1 Existing SVC Rate Adaptation Schemes

4.2.1.1 JSVM Basic Extraction

A basic extraction process, implemented in the JSVM for rate adaptation, prioritizes packets by the high-level syntax elements. The basic extractor progressively orders spatial/temporal/quality levels until the target bit rate is reached. The prioritization order is shown in Fig. 4.1 [138]. Each block in the figure represents a layer (d, t, q) , where d is the spatial layer ID, t the temporal layer ID and q the quality layer ID. SVC can always be encoded with the largest resolution and the best quality. The prioritization order is elaborated in [138]. By excluding low priority packets first, rate adaptation can be achieved.

4.2.1.2 Quality Layers Based Rate Adaptation

The Quality Layers [138] scheme adopts a Rate-Distortion (R-D) optimization method to extract packets. The Quality Layer information is associated with every NAL unit. The information is derived by calculating the R-D value of each picture. With the Quality Layer information available, either a Mono-Resolution or Multi-Resolution Extraction scheme could be employed to achieve the rate adaptation. The details on the quality layer calculation and corresponding rate adaptation process are explained in [138].

Figure 4.1: Priority order for the JSVM basic extractor, packets are prioritized by spatial, temporal, and quality ID combinations (d, t, q) (based on Fig. 3(a), page 1188 of [138]).

The prioritization policy in JSVM basic follows a fixed pattern and neglects the possible diversities in encoding configurations. Hence it is not optimal in deciding packet priorities. The Quality Layer information based scheme calls for high computational complexity and needs the availability of original video sequences at the media gateway, which makes it not feasible for real-time SVC rate adaptation. Below we propose an efficient packet prioritization scheme based on packet dependencies which can achieve comparable performance but needs only negligible computations. The optimized distortion derived by our fast scheme is more suitable for the applications in video transmissions [13].

4.2.2 Packet Prioritization Based on Packet Dependencies

To analyze the priority of SVC packets, we propose in this chapter an efficient solution by investigating the packet dependencies in encoding. Please note that the assumption in this chapter is that we aim at maximizing the quality of a video at a specific spatial resolution. This means that when adapted to any given rate by the proposed scheme, the decoded video is always in a fixed target spatial resolution. We investigate a general case where at most three spatial levels are encoded together. Before we come up with a general prioritization guideline based on the dependencies, we have conducted an extensive learning process to acquire some priori information on packet importance.

We selected eight videos which have been widely used in academic studies for our testing. The selected video sequences were encoded with different coding configurations to showcase combined spatial, temporal and quality scalabilities. Videos encoded in QCIF and CIF resolutions contained Akiyo, Coastguard, Mother & Daughter and News. Those including three spatial sizes from QCIF to 4CIF contained Aspen, Ducks_take_off, Ice and In.to.tree. It should

4. A FAST RATE ADAPTATION SCHEME FOR SVC BASED ON THE PACKET DEPENDENCIES

be noted that the 4CIF videos have been cropped from high definition (HD) videos in the Joint Video Team (JVT) test sequence pool as in [142]. We extracted the trace file of every video using JSVM bit extraction tool. We excluded video packets in the extracted trace file individually and for each discarded packet, we reconstructed a distorted video and calculated its quality degradation with respect to the error free video. For JSVM, the loss of certain packets will make its high layer descendants undecodable. We discard these packets altogether to obtain a decodable sequence. The loss of certain packets will also lead to frame loss; we recovered it by frame copy. Finally, for each packet within a single video sequence, we obtain its quality impact on the whole sequence. Base on the quality degradation evaluations, we acquire a basic knowledge on the packet importance which helps us summarize a general packet prioritization scheme. For all selected videos, we have generated around 15 thousand video segments for evaluations. We conducted the massive learning on a Linux Server computer. The server was equipped with Linux operating system Ubuntu 11.04. Intel Xeon(R) CPUs E5420 worked at 2.50 GHz and four GB memory was installed on the system. For all eight tested video sequences, we spent about 16 days for the whole testing.

By analyzing these massive learning data, we propose our packet prioritization scheme. Fig. 2.2 presents a typical prediction structure for multiple-scalability coding, where two spatially dependent layers are depicted within one GOP. This figure presents how quality scalability is realized by employing the key-picture concept in Medium-grain Quality Scalability (MGS). In MGS, quality enhancement layer packets are used to predict the lower quality layer packets. This design improves the coding efficiency and drift errors can be stopped at the key pictures, which are from the coarsest temporal layer. Spatial scalability is realized by the inter-layer prediction. To provide high coding flexibility, the dependent spatial and quality layers for current spatial level can be optionally selected. A discardable flag in the NAL unit is employed to indicate which layer is referred to for inter-layer prediction. Temporal scalability is realized by the typical hierarchical-B structure. A target video with layer specification (D, T, Q) is retrieved from the network client, which is generally determined by the terminal specifications or the applications. By analyzing the dependencies in realization of combined scalability, we progressively prioritize packets, each is characterized by scalability IDs (d, t, q) , such that packets appear first contribute more on the distortion minimization for the target video. This achieves optimized video quality under a fixed rate constraint once packets are selected with priority.

Theoretically, spatial layer can be predicted from a quality layers in the lower spatial levels, which is marked by setting the discardable flag $x = 1$. If no quality refinement layer of a lower

Figure 4.2: Multilayer encoding structure for SVC with GOP size 4. a and $a + k$ are spatial layer IDs.

The accentuated frames are key frames from the coarsest temporal layer. Hierarchical-B pictures realize temporal scalability. Inter-layer prediction structure enables spatial scalability. The key-picture concept and MGS favor quality scalability. We assume the quality base layer (containing $(a, 0, 0)$ and $(a, 1, 0)$) is adopted for inter-layer prediction in spatial scalability.

spatial layer is related to the target resolution D , this spatial layer could be excluded first. For those spatial layers that are employed for predicting high dependency layer, the quality base layer packets account for the existence of a video frame at the target resolution. The loss of any one of these packets leads to a missing frame so they should be classified into a higher priority set $\mathcal{P}_1$. When inter-layer prediction for a spatial enhancement layer is executed, the reference spatial layer ID d and quality layer ID q are stated in the NAL unit. Packets that have equal and lower quality refinement levels q in spatial layer d , and those further related in lower dependency layers, are directly referred by the target video frame so they are group into the next priority set $\mathcal{P}_2$. The reference layer can be differently designated in encoding therefore the number of packets in this set varies a lot for different layers/videos. Therefore the fixed prioritization scheme cannot effectively cope with it. Within these two sets, temporal scalability should dominate because once the lower temporal layer video frames are received or enhanced, the higher layer frames can be constructed easily by error concealment. Because temporal scalability is realized by Hierarchical-B structure, the concealed higher layer frames will generally resemble its original copy.

Quality enhancement layers at the target spatial level D are progressively placed into the

4. A FAST RATE ADAPTATION SCHEME FOR SVC BASED ON THE PACKET DEPENDENCIES

set $\mathcal{P}_3$. Because the temporal enhancement layer can be useful in improving the flowing experience produced by the accumulated temporal decomposition, therefore in this set, packets are ordered first by quality layers and then the temporal layers. There are also lower spatial layer packets which are not employed for inter layer prediction, but they still contribute for the quality increasing because of the adoption of the MGS. As can be seen from Fig. 4.2, higher quality layer packets are employed to predict lower layer packets (for example, $(a, 0, 1)$ to $(a, 1, 0)$ in the figure). Therefore, these packets indirectly influence the quality of target video, they are prioritized in the last priority set $\mathcal{P}_4$. Temporal layer dominates in this set because in SVC, the key pictures are combined with hierarchical prediction structures as could be seen in Fig. 4.2. Spatial layer dominates over the quality layer because quality refinement is only realized within a spatial layer. If packets from all quality enhancement layers in the lower spatial layer are employed for inter-layer prediction, then no packets exist in this set, therefore the number of packets in this set also varies. In addition, there are also some packets, designated as StreamHeader or ParameterSet in the standard, are vital for the decodability of the whole sequence. They are placed in the highest priority set $\mathcal{P}_0$ for the first provision. On the basis of these observations, packets are prioritized based on the following orders:

- If packets contain information of StreamHeader or ParameterSet, or they are the first encoded packet of an intra frame, they are assigned to the highest priority set $\mathcal{P}_0$ due to loss of these packets will lead to an undecodable video.
- If packets belong to the quality base layer ($q = 0$), they are assigned to $\mathcal{P}_1$. Within this group, they are prioritized first following the temporal prediction order (t). Those having the same temporal layer ID are ordered following the spatial order (d). Taking Fig. 4.2 as example, the quality base layer packets will be ordered as: $(a, 0, 0), (a + k, 0, 0), (a, 1, 0), (a + k, 1, 0), (a + k, 2, 0)$.
- If quality enhancement layer packets of the lower spatial layer (lower than the target spatial layer, $d < d^{max}$) are used for prediction of target spatial levels ($x = 0$), they are assigned to $\mathcal{P}_2$. Within this group, they are prioritized first following temporal prediction order (t), then following the spatial order (d), and at last following the quality order (q).
- For quality enhancement layer packets in the target spatial layer ($d = d^{max}$), such as the uppermost packets in Fig. 4.2, they are assigned to $\mathcal{P}_3$. Within this group, they are prioritized first following quality order (q), then following the temporal order (t).

Algorithm 1: Description of the Proposed Packet Prioritization Scheme

```

1 if Packets are StreamHeader or ParameterSet then
2   | they are assigned to  $\mathcal{P}_0$ 
3 else
4   | if  $q = 0$  then
5   |   | They are assigned to  $\mathcal{P}_1$ 
6   |   | Packets prioritized first by TID  $t$ , then by DID  $d$ 
7   | if  $d < d^{max}$  &  $x = 0$  then
8   |   | They are assigned to  $\mathcal{P}_2$ 
9   |   | Packets prioritized first by TID  $t$ , then by DID  $d$ , and at last by QID  $q$ 
10  | if  $d = d^{max}$  &  $q > 0$  then
11  |   | They are assigned to  $\mathcal{P}_3$ 
12  |   | Packets prioritized first by QID  $q$ , then by TID  $t$ 
13  | else
14  |   | They are assigned to  $\mathcal{P}_4$ 
15  |   | Packets prioritized first by TID  $t$ , then by DID  $d$ , and at last by QID  $q$ 

```

- If quality enhancement layer packets of the lower spatial layer ($d < d^{max}$) are not used for prediction of higher spatial levels ($x = 1$), they are assigned to $\mathcal{P}_4$. Within this group, they are prioritized first by temporal order (t), then the spatial order (d), at last the quality order (q). For example in Fig. 4.2, $(a, 0, 1)$ is prioritized before $(a, 1, 1)$.

To summarize, an algorithmic description is given in **Algorithm 1**. As exemplified by Fig. 4.2, the packets there will be prioritized in a decreasing order as shown in Fig. 4.3(b). Packets sharing the same layer IDs are ordered by their frame index. In Fig. 4.3(c) we also show the prioritization result if the quality enhancement layer at the lower spatial level is employed for inter layer prediction. However, the JSVM basic solution just follows a fixed pattern and will not distinguish these cases. It will always follow a fixed pattern as in Fig. 4.3(a). With the prioritized packets, we achieve the rate adaptation by discarding the lower priority packets first to conform to the rate constraint. The derived Rate-Distortion relationship is actually the lower bound of the distortion at the adapted rate point. This fast adaptation provided by the proposed method can be efficiently used in tackling resource allocation for video transmission. We will show its usage in next chapter.

4. A FAST RATE ADAPTATION SCHEME FOR SVC BASED ON THE PACKET DEPENDENCIES

Figure 4.3: Priority order in the given instance Fig. 4.2, (a) from JSVM Basic, (b) and (c) are from the proposed scheme: (b) quality base layer for inter-layer prediction. (c) quality enhancement layer for inter-layer prediction.

Table 4.1: Coding configuration of the tested videos

Video	Bus	Football	Foreman	Mobile	City	Crew	Harbour	Soccer
Spatial layer	2	2	2	2	3	3	3	3
Quality Layer in QCIF	3	4	4	3	3	3	2	2
Quality Layer in CIF	2	4	3	3	3	2	3	2
Quality Layer in 4CIF	-	-	-	-	2	3	2	2
Highest frame rate	30	30	30	30	60	60	60	60

4.3 Simulation Results

We validate the performance of the proposed packet priority analysis in this section by extensive simulations. To make thorough studies on the proposed work, we encode reference videos recommended in [142] with different configurations as shown in Table 4.1. The proposed packet prioritization scheme based on packet dependencies is denoted as PD.

The first frame of each video is intra coded and GOP size is 16. The derived Rate-Quality (PSNR) curves from different rate adaptation schemes for two representative videos are shown

Table 4.2: PSNR (dB) from different rate adaptation schemes

Video	Bus	Football	Foreman	Mobile	City	Crew	Harbour	Soccer	Average
JSVM Basic	31.60	32.71	31.89	28.24	32.89	35.27	32.85	34.47	32.49
Quality Layer	32.13	33.50	32.83	28.76	33.52	35.45	33.17	34.54	32.99
PD	32.10	33.66	32.99	28.79	33.72	35.40	33.20	34.48	33.04

Table 4.3: Time consumption (second) by different rate adaptation schemes

Video	Bus	Football	Foreman	Mobile	City	Crew	Harbour	Soccer	Average (CIF/4CIF)
JSVM Basic	0.0013	0.0108	0.0106	0.0086	0.0279	0.0353	0.0280	0.0189	0.0078/0.0275
Quality Layer	63.12	255.82	202.36	180.97	1468.11	1704.32	1461.62	1011.98	175.57/1411.51
PD	0.0013	0.0110	0.0103	0.0087	0.0274	0.0348	0.0290	0.0191	0.0078/0.0276

in Fig. 4.4. The figures in the left show the comparisons in the full range of rate adaptation. The JSVM basic scheme performs the worst as it only prioritizes packets in a fixed pattern but ignores the possible differences in changed coding configurations. The proposed scheme acquires comparable performance as the Quality Layer based scheme.

To have a quantitative measure of the difference among different schemes, we calculate the average PSNR and compare their differences as in [143]. In [143], the difference of two groups of data is evaluated by the mean absolute difference. Because the Quality Layer based scheme can only be effective for partial of adaptation points and in a limited range as shown in Fig. 4.4, the average PSNR values are only calculated within that effective interval. Results on PSNR are shown in Table 4.2. The comparison demonstrates that the proposed scheme achieves comparable performance with the Quality Layer based scheme but outperforms the JSVM basic more than 0.5 dB. The improvement can be consistently found for all tested videos with different content characteristics, which demonstrate the efficacy of the proposed scheme. We highlight in the right parts of Fig. 4.4 the comparisons in the partial effective range of the Quality Layer based scheme for easy of viewing. The four selected videos, Bus, City, Foreman and Football are characterized in different contents. Nevertheless, the proposed scheme always achieves the consistent performance for all videos. The Quality Layer based scheme is not optimal (our proposed scheme surpasses it for some videos) because in its Rate-Distortion calculations, only the dependent distortion from temporal decomposition is accounted. In fact, the dependency also comes from other quality refinement layers, as we have discussed in Section 4.2.2. To check the comparisons of the three schemes from another angle, we show in Fig. 4.5 snapshots of the decoded pictures from two video sequences Bus and City.

To evaluate the performance of the three rate adaptation schemes on the computational complexity, the consumed time in these schemes are summarized in Table 4.3. We use a com-

4. A FAST RATE ADAPTATION SCHEME FOR SVC BASED ON THE PACKET DEPENDENCIES

puter equipped with Intel Core 2 Duo CPU E6750 at the 2.66 GHz for these evaluations. The memory installed is two GB and the operating system is Ubuntu 8.04. For Quality Layer based scheme, the time spent includes the Quality Layer information computation and corresponding rate adaptation. The time spent in the JSVM basic and the proposed schemes is for the packet prioritization and the corresponding rate adaptation. It is easy to find that the time consumed for the proposed scheme is trivial when comparing to the Quality Layer based scheme.

As a summary, the proposed scheme, i.e., PD, performs averagely 0.5 dB better than the JSVM basic bit extraction scheme but with the close computational burden. It also outperforms the Quality Layer based rate adaptation scheme in the following aspects whilst achieves the same PSNR in average: 1) much lower computational cost, 2) no need of the original video sequences for rate-distortion calculations, and 3) wider range to support rate adaptation. In addition, the proposed rate adaptation derives closely adjoined truncation points which, is enough for a wide range of applications calling for fine-granular rate adaptations. We can also indirectly validate that our method also performs better than a recent study [139]. The reference [139] utilizes an average of 1/400 time of the Quality Layer calculation to guarantee only a slight loss in quality. However, our proposed method can ensure almost the same quality with the Quality Layer based scheme but can save even more time than the reference. The proposed scheme is more than 2.2×10^4 times faster than the Quality Layer based scheme. The algorithm to solve an Integer Linear Programming problem in [139] needs some calculations. However, the performance is still not matchable to our proposed scheme, which requires much less computational complexity. Compared to the work in [54] which focuses on single-resolution video, our scheme is applicable to multi-resolution videos. Moreover, we avoid the complex computations to derive the distortion, which is quite desirable for real-time services.

4.4 Concluding Remarks

We propose in this chapter a fast rate adaptation scheme for the SVC standard. The fast adaptation is achieved by an enhanced packet prioritization scheme, which is based on the investigation of the packet dependencies in combined scalability. By selecting the more important packets first, a distortion minimized rate adaptation can be achieved for the target bitrate. This method makes improvement from Chapter 3 where off-line calculations are needed. The proposed scheme is feasible in low computational complexity and provides a wide range and fine-granular rate adaptation, which makes it efficient in applications of rate adaptation and

resource allocation. To highlight its usage in SVC transmission, we consider the resource allocation issue in Chapter 5.

The executable programs for the evaluated bit extraction schemes, as well as some sample bitstreams/videos obtained in this chapter, can be accessed at [144].

4. A FAST RATE ADAPTATION SCHEME FOR SVC BASED ON THE PACKET DEPENDENCIES

Figure 4.4: Performance comparison (verified by PSNR) on different rate adaptation schemes. Left: rate adapted in the full rate range. Right: the highlighted rate adaptation in the effective rate range of the Quality Layer based scheme. (a) Bus. (b) City. (c) Foreman. (d) Football.

(a)

(b)

Figure 4.5: Snapshots of decoded video sequence for subjective assessment. (a) Video Bus adapted at 1100 kbps, scaled for easy viewing. Picture quality from left to right, JSVM Basic: 31.96 dB, Quality Layer: 32.93 dB, PD: 32.91 dB. (b) Video City adapted at 3400 kbps. Picture quality from left to right, JSVM Basic: 34.31 dB, Quality Layer: 35.21 dB, PD: 35.26 dB.

4. A FAST RATE ADAPTATION SCHEME FOR SVC BASED ON THE PACKET DEPENDENCIES

Scalable Resource Allocation for Scalable Video Streaming over WMANs

5.1 Introduction

We have focused on maximizing the overall quality for multiple video users in Chapter 3. It is of high interest how resources should be allocated fairly among users. To move further, how the trade-off between efficiency and fairness can be achieved in a multiuser resource allocation problem. We will address this issue in this chapter by proposing a scalable resource allocation framework to handle the subject in a masterly way.

The next-generation wireless networks are anticipated to have boosted system throughput and miscellaneous services. The core component of the ubiquitous wireless access is the multiantenna-multicarrier technology. Among which, the Multiple Input Multiple Output-Orthogonal Frequency Division Multiplex (MIMO-OFDM) technique can provide high speed connections by fully exploiting various diversities [36, 145]. Although MIMO-OFDM has been deployed in many latest wireless standards, we will select the representative Wireless Metropolitan Area Network (WMAN) specification, the IEEE 802.16 [3] as an example in this work. Please note that the analysis and the framework are applicable to other systems with MIMO-OFDM setup, whereas the experimental analysis will only be presented based on the IEEE 802.16 specifications.

5. SCALABLE RESOURCE ALLOCATION FOR SCALABLE VIDEO STREAMING OVER WMANS

For an Orthogonal Frequency Division Multiple Access (OFDMA) system, various resource allocation schemes can be found in the literature [65, 61, 62, 66, 73]. Conventional strategies intended for single antenna systems include round-robin (RR) [65] and water filling (WF) based [61] solutions. The RR resource allocation alternately distributes resources to users without considering the contents and channels. Hence, it is generally inefficient in resource utilization compared to the content and network aware schemes, as will be demonstrated in the experimental analysis. A Linkgain-Based Resource Allocation (LBRA) [61] achieved the rate maximization in a single antenna system by assigning an OFDM subcarrier to a user with the best channel and power is allocation in water-filling fashion. However, this scheme could not maximize the performance of multiantenna systems as cochannel user interference was not considered. A Multicarrier Maximum Sum Rate (MMSR) bit removal scheme was proposed in [62]. This scheme could not achieve the optimal video transmission without utilizing the characteristics of the video sources. Moreover, the throughput maximization is not equivalent to the video utility maximization, particularly due to the dependence of video packets. More critically, the rate maximization schemes always lead to a lopsided resource allocation where users with undesirable channel conditions are always compromised. Such a lopsided allocation will always lead to severe quality degradation for some users and result in a loss in the overall system performance. Tsai *et al.* [66] investigated the delivery of multimedia traffic over OFDMA/SDMA networks and presented an efficient Adaptive Radio Resource Allocation (ARRA) algorithm. The ARRA does not consider the video packet differences in length and their dependency; therefore, it will only lead to degraded system performance. A pioneering work on scalable framework has been proposed for single antenna OFDM system in [73]. The major differences between our framework and the existing work lie in the following aspects. Firstly, the existing work distinguishes the fairness and the efficiency schemes by artificially divided transmission intervals. The differentiation of the objective functions is subjective and is application unaware. To accommodate heterogeneous clients in our framework, we attempt to provide a fair and efficient solution for different scalable video layers. Secondly, the employed video codec is different with us therefore the video utility analysis differs from the previous one. Last but not least, existing solutions mainly focus on a single antenna system. Based on the work in [73], we formulate a more general scalable framework and propose an efficient solution to accommodate the multi-antenna systems in this chapter.

Up to now, the majority of resource allocation schemes devote themselves to bandwidth utility optimization. However, it is of high importance to explore the video utility in constrained

Figure 5.1: The SVC encoding structure for a two-quality layer configuration.

networks in order for the improved viewing experience [146, 147, 68]. Due to the uniqueness of the multiuser MIMO-OFDM systems, the existing solutions cannot be directly applied. A direct implementation of the existing solutions cannot provide an answer for how resource should be distributed in these multidimensional resource reservoirs.

To achieve an optimized system utility for multiple scalable video downlink over multiuser MIMO-OFDM systems, we propose a scalable resource allocation framework to distribute limited radio resources. The utility analysis in Chapter 4 can guarantee a maximized utility under a fixed data rate by distinguishing packets of different priorities, while the scalable resource allocation framework presented in this chapter, can ensure that under the fixed data rate, resources can be efficiently distributed to users for different optimization objectives. To better utilize the inherent superiority provided by SVC and offer each user a provision of a basic viewing of the video, the scalable framework distributes radio resources on two different aspects. For the receiving of the base layer bitstream for every user, a fair scheme is designed to achieve the MAXMIN fairness among users. When all users have their base layer packets received, an efficient scheme is employed to explore the network efficiency. A combined analysis is presented where the fair and efficient schemes are integrated into a unified resource allocation framework. The proposed joint bit loading and power allocation solution for either fairness or efficiency refrains from the high computational complexity in solving the combinatorial problems. A filling fashion of power and bit allocation scheme ensures that the proposed scheme is effective and can be easily applied in practical systems. Table 5.1 summarizes the notation used in this chapter.

5.1.1 Utility of Scalable Video Coding

There are various possibilities in encoding an SVC bitstream based on different combinations of scalabilities. To facilitate the analysis in this chapter, we consider a simple case where two quality layers are encoded within a spatial level as shown in Fig. 5.1. Temporal scalability is

5. SCALABLE RESOURCE ALLOCATION FOR SCALABLE VIDEO STREAMING OVER WMANS

Table 5.1: Notation used in Chapter 5

k	interchangeably used as video or user index
U_k	utility of video k
D_k	distortion of video k
R_k	accumulated transmission rate of video k
$R_k^{[a]}$	accumulated transmission rate by all previous system frames
$R_k^{[c]}$	rate allocated in current system frame
γ_k	rate of video k
ξ_k	packet selection vector
$I^{[j]}$	prioritization index
t and q	temporal and quality layer IDs
$R_k^{[P]}$	rate of prioritized packets for video k
$D_k^{[P]}$	distortion of prioritized packets for video k
K	user number
N	number of subchannels
L	symbols in a system frame
Q	number of transmit antennas
b	number of subcarriers in a subchannel
BER_{tgt}	target bit error rate
$i, k, n,$ and l	subcarrier, user, subchannel and symbol indices, respectively
$x_{k,i}^{(l)}, y_{k,i}^{(l)}, \mathbf{H}_{k,i}^{(l)}, \mathbf{w}_{k,i}^{(l)}$ $p_{k,i}^{(l)}, m_{k,i}^{(l)}, SNR_{k,i}^{(l)}, Z_{k,i}^{(l)}$	transmitted symbol, received symbol, channel gain, BF weight vector, power, modulation, SNR and noise for user k in the i th subcarrier of the l th symbol
σ^2	variance of noise
$c_{k,n,l}, p_{k,n,l}, m_{k,n,l}$	user, power and modulation allocation indices for user k in unit (n, l)
$\mathbf{U}$	utility vector
$f(\mathbf{U})$	objective function of utility
$\mathbf{C}, \mathbf{P}, \mathbf{M}$	vector of $c_{k,n,l}, p_{k,n,l},$ and $m_{k,n,l}$
P_T	power constraint in a symbol
$\mathcal{M}$	set of transmission bits by different modulations
M	cardinality of set $\mathcal{M}$
$\tilde{R}_k^{[B]}$	rate of base layer packets of video k
$e_{k,n,l}$	power cost for user k in unit (n, l)
$\mathcal{K}_{n,l}$	set for users admitted into unit (n, l)

realized within a quality layer. For more details about SVC, please refer to the introductory knowledge in Chapters 1 and 2.

In Chapter 4, we have proposed a fast rate adaptation scheme. With the prioritized SVC packets, we can selectively choose high priority packets in resource-constrained scenario to achieve the best video utility (or the minimized distortion). For the progressively selected packets, we derive point-wise Rate-Distortion (R-D) relationship. We can construct a continu-

ous R-D function on the discrete R-D points by a linear interpolation as in [65]

$$D_k^{[P]}(\gamma_k) = \frac{D_k^{[P]}[i+1] - D_k^{[P]}[i]}{R_k^{[P]}[i+1] - R_k^{[P]}[i]}(\gamma_k - R_k^{[P]}[i]) + D_k^{[P]}[i] \quad (5.1)$$

for $R_k^{[P]}[i] \leq \gamma_k \leq R_k^{[P]}[i+1]$ and $i = 0, 1, \dots, T_k$

where γ_k is the allocated rate for user k and T_k is the number of video packets. As we consider the downlink of videos to multiple users where each user receives a single video, k is interchangeably used for the user and video in this chapter. The derived R-D relationship based on prioritization of packets is actually a lower bound of the distortion at a given data rate. In this work we define the utility as the minus distortion by the direct reference to the conventional video rate-distortion (R-D) analysis

$$U_k(R_k) = -D_k^{[P]}(R_k) \quad (5.2)$$

where the superscript $[P]$ here indicates its derivation based on the packet prioritization scheme in Chapter 4.

5.1.2 Multiuser MIMO-OFDM Networks

MIMO systems can be implemented in different ways for diversity gain, capacity gain, or interference reduction. In SDMA, independent data are transmitted simultaneously from an individual antenna to increase the system throughput. The SDMA scheme exploits the unique, user-specific “spatial signature” to support multiple users within the same frequency band and/or time slot, provided that the spatial signatures are different and can be accurately measured [36]. With the availability of OFDMA, resources could be effectively allocated by exploiting both spatial and frequency diversities to enhance the overall system performance.

The downlink SDMA/OFDMA system adopted in this work is depicted in Fig. 5.2. We assume that K single antenna receivers request videos simultaneously from a base station which is equipped with Q antennas. As the coherence time of the wireless channel is much larger than the duration of an OFDM symbol, we let L OFDM symbols compose a system frame and the channel is considered to be invariant during a system frame. Please note that the system frame here refers to a time slot in the MIMO-OFDM system. As grouping of adjacent subcarriers will result in higher multiuser diversity [66] and can reduce the implementation complexity, we follow the same configuration that adjacent b subcarriers are organized to form

5. SCALABLE RESOURCE ALLOCATION FOR SCALABLE VIDEO STREAMING OVER WMANS

Figure 5.2: The system structure of multiple SVC video downlink transmissions over multiuser MIMO-OFDM networks (based on Fig. 1, page 1735 of [66] and Fig. 1, page 109 of [67]).

a frequency subchannel. A subchannel is considered as an integrated unit in the resource allocation. This design can reduce overhead information and complexity in resource allocation. At the beginning of each system frame, the resource allocation scheme is executed to allocate the time-frequency unit, modulation order, and power to users to obtain the optimized video utility. A corresponding illustration of the resource allocation block is depicted in Fig. 5.3.

We assume that the transmitter has full knowledge of channel state information (CSI), which can be acquired in a WiMAX similar system via a duplex feedback. A simple yet efficient transmit beamforming scheme, zero-forcing beamforming (ZFBF) [148] weight vector can be configured such that users are orthogonal in the space domain. It should be noted that there exist other advanced techniques, such as optimal dirty paper coding and other beamforming schemes. We employ ZFBF because: 1) We always arrange insufficient number of users in a resource unit so the ZF-based precoding can achieve close to the optimal capacity. 2) We focus on the resource allocation. The proposed framework and the solutions apply to other coding scheme. 3) It is simple to implement. We first consider the relationship of power and modulation in different subcarriers. The received signal $y_{k,i}^{(l)}$ of user k in the i th subcarrier and l th OFDM symbol is

$$y_{k,i}^{(l)} = \mathbf{H}_{k,i}^{(l)} \mathbf{w}_{k,i}^{(l)} \sqrt{p_{k,i}^{(l)}} x_{k,i}^{(l)} + Z_{k,i}^{(l)} \quad (5.3)$$

where $\mathbf{H}_{k,i}^{(l)}$ is channel gain, $Z_{k,i}^{(l)}$ is the thermal noise which is assumed to be in complex Gaussian distribution with zero mean and variance σ^2 , $p_{k,i}^{(l)}$ is the power allocated, $x_{k,i}^{(l)}$ denotes

the corresponding data symbol, and $\mathbf{w}_{k,i}^{(l)}$ is the beamforming weight vector. The SNR of the received signal is

$$SNR_{k,i}^{(l)} = \frac{|\mathbf{H}_{k,i}^{(l)} \mathbf{w}_{k,i}^{(l)}|^2 p_{k,i}^{(l)}}{\sigma^2} \quad (5.4)$$

To satisfy a target bit error rate BER_{tgt} , sufficient power $p_{k,i}^{(l)}$ should be allocated for reliable reception of $m_{k,i}^{(l)}$ bits for a selected modulation-coding scheme. It should be noted that $p_{k,i}^{(l)}$ is calculated differently when Adaptive Modulation and Coding (AMC) is employed. Following the existing work [66, 67, 149], we will not consider various coding schemes in this work and will focus specially on the study of the resource allocation problem. We will simply mention the selection of modulation schemes in the following context¹. If M -ary quadrature amplitude modulation (MQAM) is employed, the power is related to the target BER and channel as [150]

$$p_{k,i}^{(l)} = \ln \left(\frac{1}{5BER_{target}} \right) \frac{\sigma^2 (2^{m_{k,i}^{(l)}} - 1)}{1.5 |\mathbf{H}_{k,i}^{(l)} \mathbf{w}_{k,i}^{(l)}|^2} \quad (5.5)$$

5.2 The Scalable Resource Allocation Framework

We show how to optimize the system performance for scalable video delivery over multiuser MIMO-OFDM wireless networks. Given the mixed integer and nonlinear natures of the problem, we propose practical bit loading and power allocation schemes to effectively resolve the optimization problems. We present the proposed scalable framework to transmit the base layer bitstream with guaranteed fairness and then the enhancement layer packets to exploit the network efficiency.

5.2.1 Problem Statement

We carry out our resource allocation in the time-frequency unit (OFDM symbol-subchannel unit). We use only subscript to distinguish from the subcarrier-based relationship. For example, $p_{k,n,l} = \sum_{i \in n} p_{k,i}^{(l)}$ is the power summation for all subcarriers in the subchannel n and the

¹The coding schemes have not been considered in existing work [66, 67, 149] because of the following justifications: 1) The focus of these work is on resource allocation. 2) The proposed resource allocation framework is still applicable when coding schemes are considered, only different in power and rate calculations. 3) Packet loss due to transmission errors can be reduced by coding schemes. However, it is quite limited when compared to the massive packet loss due to insufficient resources. As the focus of this work, as well as those of [66, 67, 149], is on resource allocation, we do not explicitly the coding schemes. We will incorporate them into consideration in our future work.

5. SCALABLE RESOURCE ALLOCATION FOR SCALABLE VIDEO STREAMING OVER WMANS

Figure 5.3: Illustration of the resource allocation.

symbol l to user k . $m_{k,n,l}$ is the transmitted bits in subchannel n . Let $c_{k,n,l}$ be the allocation indicator; $c_{k,n,l} = 1$ means that the unit of the n th subchannel and the l th OFDM symbol is allocated to the user k for packet transmission. Otherwise, $c_{k,n,l} = 0$. Then, the transmission during an OFDM symbol yields a total power consumption of $\sum_{k=1}^K \sum_{n=1}^N c_{k,n,l} p_{k,n,l}$. The data transmitted to user k in a system frame are $\sum_{n=1}^N \sum_{l=1}^L c_{k,n,l} m_{k,n,l}$. As a unified framework, for different scalability level, we formulate the resource allocation problem as maximizing the system utility by allocation of unit assignment vector $\mathbf{C}$, power distribution vector $\mathbf{P}$ and modulation selection vector $\mathbf{M}$, where $\mathbf{C}$, $\mathbf{P}$, and $\mathbf{M}$ indicate the value sets in which $c_{k,n,l}$, $p_{k,n,l}$, $m_{k,n,l}$ are chosen in each combination (k, n, l) . As stated in Section 5.1.1, we characterize the utility of video k by a utility function $U_k(R_k)$, where the R_k corresponds to the sum rate of all packets actually transmitted to user k . R_k includes the accumulated rate transmitted from all previous time slots ($R_k^{[a]}$) and the one allocated in current system frame ($R_k^{[c]}$). Based on the accumulated rate from all previous system frames, we try to maximize system utility by allocating radio resource in current system frame. The maximization problem is defined as an objective function of $\mathbf{U}$. $\mathbf{U}$ is a K -element utility vector where

$$\mathbf{U} = [U_1, U_2, \dots, U_K]^T.$$

$$\max_{\mathbf{C}, \mathbf{P}, \mathbf{M}} f(\mathbf{U}) \quad \text{s.t.} \quad \begin{cases} \sum_{k=1}^K \sum_{n=1}^N c_{k,n,l} p_{k,n,l} \leq P_T \\ 0 \leq \sum_{k=1}^K c_{k,n,l} \leq Q \\ p_{k,n,l} \geq 0 \\ c_{k,n,l} \in \{0, 1\} \\ m_{k,n,l} \in \mathcal{M} \end{cases} \quad (5.6)$$

where P_T is the power constraint in each symbol and $\mathcal{M}$ is an integer set containing the possible number of transmission bits in each unit under different modulation schemes.

The video utility optimization problem is designed under the following constraints: 1) the total power consumption within each OFDM symbol should not exceed the maximum power supply and power allocation should not be negative; 2) the number of users admitted into a unit should not exceed the number of transmitter antennas Q , to guarantee that the spatial multiplexed signals can be correctly decomposed at the receivers; 3) the modulation scheme for each user in a unit is selected from a candidate set $\mathcal{M}$. $\mathbf{U}$ is a general form and can be defined in different mathematical forms for the expected optimization objective.

We will show that the problem (5.6) is NP-hard for different optimization definitions $f(\mathbf{U})$. The mixed integer programming problems with nonlinear constraints and general form of utility definition are not readily to be solved optimally via the existing programming techniques [151], [152]. The optimal solution calls for high computational complexity because of the nonlinear and non-convex constraints. It is not practical to exhaustively search for the optimal solution in a practical system owing to the following statement: If we denote cardinality of $\mathcal{M}$ is M , the allocation in the resource space requires $\mathcal{O}(\binom{K}{Q}(M+1)^{LN})$ times of exhaustive search.

We will develop algorithms in the following sections, which are preferable for streaming applications, to efficiently resolve the resource allocation problems. By exploiting the inherent priority differentiations of video packets, we employ the optimized rate-utility analysis framework for SVC presented earlier. According to the video utility analysis, a two-stage algorithm is designed to guarantee fairness for the base layer and exploit efficiency for the enhancement layer, respectively. Although we consider single antenna receivers in this work, the proposed formulation Eq. (5.6) can be extended to multi-antenna case on the basis that we calculate the data received by a user within a system frame differently.

5.2.2 Description of the Scalable Resource Allocation Framework

In a multiuser system, several users compete for limited radio resources when requesting the same video in different resolution/quality or just simply different videos. The successive encoding structure imposes that base layer packets should be preferably delivered to every user. The receiving of only the base layer packets can provide a coarser quality and lower frame rate video that is acceptable for a basic viewing experience. Every subscribed user could earn a basic provision of service, regardless of his/her channel condition. To provide every user with the minimum requirement, no user is deemed to compromise. Hence in this course, fairness among users should be guaranteed.

After every user is satisfied with a basic provision of video viewing, the remaining resources should be distributed differently. For videos in different motion and content characteristics, the consumption of the same number of resource will contribute to different degree of quality increment. For users experiencing different network conditions, different resource allocations also lead to distinct network efficiency. Therefore in this course, efficiency exploitation should be carried out to better explore the overall system capacity.

We propose a scalable resource allocation framework derived from the optimization problem proposed in Section 5.2.1 and, provide fairness for the base layer packets and efficiency for the enhancement layer packets. It should be noted that practically in SVC, multiple types of scalabilities are jointly realized so sometimes it is inconsistent in classifying packets to base and enhancement layers. The cut-off point for the scalable framework can be defined according to applications, but the proposed scheme always applies well. A flowchart of the proposed framework is presented in Fig. 5.4. The $\tilde{R}_k^{[B]}$ denotes the total data from the base layer for the user k . The fair scheme and the efficient scheme are executed exclusively by judging whether all admitted users have their base layer packets successfully scheduled. The resource allocation is carried out until the power resources in an OFDM symbol are used up; then the procedure proceeds to the next symbol. Once the deadline for currently served GOP is reached, the overtime packets are discarded and the resource allocation is executed for the next GOP. Transmission loss may occur sometimes. A retransmission before the GOP deadline may improve the performance, especially for the important base layer packets. However, we only focus on the resource allocation here, and will incorporate retransmission into consideration in our future work. To ensure that the proposed scheme can be efficiently carried out, GOP synchronization is needed. The comparisons of video utility for different video sequences are

$$e_{k,n,l} = \begin{cases} \infty & \text{if } m_{k,n,l} = \lceil \mathcal{M} \rceil \\ \infty & \text{if } \sum_{k=1}^K c_{k,n,l} = Q \\ p(\mathcal{K}_{n,l}, m_{k,n,l}^+) - p(\mathcal{K}_{n,l}, m_{k,n,l}) & \text{if } c_{k,n,l} = 1 \text{ \& } m_{k,n,l} < \lceil \mathcal{M} \rceil \\ p(\mathcal{K}_{n,l}^+, m_{k,n,l}^+) + \sum_{k' \in \mathcal{K}_{n,l}} (p(\mathcal{K}_{n,l}^+, m_{k',n,l}) - p(\mathcal{K}_{n,l}, m_{k',n,l})) & \text{if } c_{k,n,l} = 0 \end{cases} \quad (5.7)$$

Figure 5.4: Flowchart of the scalable resource allocation framework for scalable video streaming over multiuser MIMO-OFDM systems.

valid if the transmission of GOPs starts simultaneously. This disposal also answers for the situation if user entry/exit occurs during streaming. It delays at most a GOP time to initiate the new allocation.

We intend to design efficient bit filling strategies for the resource allocation problem (5.6). To progressively utilize limited resources and design different resource allocation strategies for different video layers, a bit filling scheme is designed to allocate power resources first for fairness and then for efficiency, until the power is used up. This bit filling scheme overcomes the shortcomings of the existing bit removal solution [62] which is not suitable for a scalable framework. The core idea of the bit filling scheme is summarized as follows: increase one order

of modulation for a user in a subchannel where the target video utility $f(\mathbf{U})$ can be maximized.

We accomplish such a filling based bit loading scheme by calculating the power cost at each step. The cost is the required power to increase the modulation order. The power cost $e_{k,n,l}$ for the user k in the unit (n, l) is expressed in Eq. (5.7). A set $\mathcal{K}_{n,l}$ is constructed to record all users that have been admitted into the unit (n, l) and have selected one modulation order from the set $\mathcal{M}$. The power cost is calculated in Eq. (5.7) by following cases. 1) If a selected user in a unit reaches the largest modulation order or the number of other admitted users in a unit reaches Q , the current user will be excluded in future modulation increasing. 2) If a user has previously selected a modulation scheme from the candidate set $\mathcal{M}$ and the current modulation order is not the highest one, the power cost is the difference between the power consumption when selecting the next higher order of modulation and that when selecting the current modulation. 3) If a user has not been selected in a unit, including a new user needs to calculate the beamforming vectors again. By differentiating this, all increasing cost is included in calculating $e_{k,n,l}$ in Eq. (5.7). At the beginning of each system frame, every selected user is initialized by an initial cost. This is derived on the assumption that only one single user is admitted to the unit (n, l) and the user is entitled to the first order of modulation. As we assume the channel is invariant within a system frame, the initial costs are only estimated at the first symbol and reused for the remaining ones.

We analyze the computational cost of power cost calculations. If a user is ready to be the first time included into certain subchannels, the complexity is bounded by $\mathcal{O}(KN)$ of beamforming vector calculation and power estimation. Otherwise if a user has been assigned with a modulation order in previous steps, the calculation complexity is bounded by $\mathcal{O}(QN)$ times of subtraction. In any case, the power cost calculations need only polynomial times of linear operations so the computational complexity involved is relatively low.

5.2.3 Resource Allocation for the Base Layer: Fairness

Fairness has long been a research interest in multiuser resource allocation problems. In guaranteeing fairness for the base layer packets, $f(\mathbf{U})$ should be designed such that the optimization problem in Eq. (5.6) is to grant every user an impartial opportunity for transmission. Since we aim at granting all admitted users fair opportunity to obtain a basic view experience, a widely used fairness metric, the MAXMIN fairness, is adopted. It should be noted that there are various fairness metrics proposed in the literature, we adopt the MAXMIN fairness based on the following reasons: 1) our utility is counted on the accumulated rate, hence it is not necessary

to consider the long term fairness in current system frame; 2) the involved users are all SVC users and no one is deemed to compromise for its base layer videos. We tend to guarantee the extreme fairness in this scenario so the MAXMIN utility metric is a reasonable choice. To transmit the base layer video packets, we define $f(\mathbf{U})$ as the negative of minimal utility gap among users. The MAXMIN fairness is achieved by minimizing the maximum utility gap compared to the utility when all base layer packets are transmitted (equivalent to maximizing the minimum of negative gap). The objective function can be defined as

$$f(\mathbf{U}) = \min_k \left(U_k(R_k) - U_k(\tilde{R}_k^{[B]}) \right) \quad (5.8)$$

where $\tilde{R}_k^{[B]}$ and $U_k(\tilde{R}_k^{[B]})$ are the total data and corresponding utility from the base layer. Hence, the optimized resource allocation problem for fairness among users comes to

$$\max_{\mathbf{C}, \mathbf{P}, \mathbf{M}} \min_k \left(U_k(R_k) - U_k(\tilde{R}_k^{[B]}) \right) \quad \text{s.t.} \quad \begin{cases} \sum_{k=1}^K \sum_{n=1}^N c_{k,n,l} p_{k,n,l} \leq P_T \\ 0 \leq \sum_{k=1}^K c_{k,n,l} \leq Q \\ p_{k,n,l} \geq 0 \\ c_{k,n,l} \in \{0, 1\} \\ m_{k,n,l} \in \mathcal{M} \end{cases} \quad (5.9)$$

We show the intractability the problem in Eq. (5.9), in the following theorem.

Theorem 5.2.1. *The optimization problem (5.9) is NP-hard even for $L = 1$ and $Q = 1$.*

Proof: We show that a special instance of the optimization problem (5.9) is a H_4^{fdma} problem in [153] even if $L = 1$ and $Q = 1$. The H_4^{fdma} problem for two user cases has been proved to be NP-hard in [153] by polynomial transformation to the known NP-complete equipartition problem. Only in this proof we ignore the subscript l . If we simplify the utility definition as the rate acquired at current symbol, the optimization problem can be reduced to a rate MAXMIN problem as

$$\max_{\mathbf{C}, \mathbf{P}, \mathbf{M}} \min_k \sum_{n=1}^N m_{k,n} \quad \text{s.t.} \quad \sum_{k=1}^K \sum_{n=1}^N c_{k,n} p_{k,n} \leq P_T \quad (5.10)$$

The problem (5.10) is known to be NP-hard in [153] therefore the original problem (5.9) with generally nonlinear and nonconvex utility definition must be NP-hard. $\square$

Theorem 5.2.1 suggests that finding a global MAXMIN solution is computationally intractable when the number of symbols and subchannels are large, even with small number of users. Given the pessimistic declaration, we are naturally led to an efficient polynomial time

5. SCALABLE RESOURCE ALLOCATION FOR SCALABLE VIDEO STREAMING OVER WMANS

algorithm. In what follows, we propose an iterative ascent algorithm by bit filling to approach the optimal allocation with polynomial order complexity.

A practical solution using iterative descent procedure to achieve MAXMIN distortion of video coding has been proposed in [154]. Based on a similar idea, a bit filling strategy is designed to deal with the power and modulation allocation which is efficient for practical applications. At the initialization stage, the power and modulation orders are set to be zero for all users in every unit. Then two sequential steps are carried out iteratively.

Step 1: The user selected to increase one order of its modulation is determined by

$$k^* = \arg \min_k (U_k(R_k) - U_k(\tilde{R}_k^{[B]})) \quad (5.11)$$

Step 2: For the selected user k^* in current symbol l^* , the modulation order is increased by one in the unit (n^*, l^*) which requires the least power cost

$$n^* = \arg \min_n e_{k^*, n, l^*} \quad (5.12)$$

Then the power resources p_{k^*, n^*, l^*} are allocated such that a target BER_{tgt} is guaranteed by referring to Eqs. (5.4) and (5.5). This procedure is carried out iteratively until the power resources are used up or $\tilde{R}_k^{[B]}$ for every user k is satisfied. The proposed solution maintains fair steps for users approaching their target utility $U_k(\tilde{R}_k^{[B]})$. The unit for transmission is selected by taking into consideration the channel variations. The design better achieves fairness and meanwhile, sufficiently takes advantage of the network efficiency. A description of the bit loading and power allocation scheme for fairness is given from Line 5 to 9 in **Algorithm 2**.

We next analyze the performance and complexity of the proposed solution for the fairness scheme. Our optimization objective is a special case of [146] where utility MAXMIN fair sharing in multi-link system has been investigated. When we come to the single bottle link system with resource constraints, the proposed solution is efficient in the sense of achieving fairness, as well as utilizing network efficiency. We select the least satisfied user in (5.11) to increase its utility. The utility difference only comes from the negligible rate difference in modulation schemes therefore we can anticipate the utility can be increased in almost the identical pace and fairness can be maintained constantly, as will be validated in the experimental evaluations. On the other hand, the channel efficiency is exploited in (5.12) where, with the least power the candidate user receives the same utility revenue. The proposed solution is simultaneously beneficial to the network efficiency and user diversity. Last but not least, the complexity involved in a system frame is upper bounded by $\mathcal{O}(LN(M+1))$ times of calculation on the utility gap, which is acceptable.

5.2.4 Resource Allocation for the Enhancement Layers: Efficiency

Exploiting network efficiency is also a major concern in designing resource allocation schemes. In this case, $f(\mathbf{U})$ has been commonly defined as the rate summation or other similar objectives. For a video transmission system, the rate maximization is not rational as the overall video experience cannot be well reflected. To design an accurate measure to estimate the benefits of resource allocation strategy in reflecting the visual experience for multiple users, $f(\mathbf{U})$ should be defined as a function of utility of all users with the consideration of video content and coding characteristics.

To proceed from our fair scheme to transmit the base layer packets, we seek to exploit the network efficiency by distributing resources to users to achieve the utility maximization for the entire network. If all users could successfully have their base layer packets delivered, the resource allocation calls a stop if all power resources have been used up. Otherwise, we consider the algorithm to fully exploit network efficiency. In this case, $f(\mathbf{U})$ is defined as the summation of video utilities of all users

$$f(\mathbf{U}) = \sum_{k=1}^K U_k(R_k) \quad (5.13)$$

As all users have already had their base layer packets delivered so every received video bit-stream could acquire a basic video viewing. The remaining radio resources should be allocated by maximizing $f(\mathbf{U})$ in Eq. (5.13). Hence, the resource allocation problem comes to

$$\max_{\mathbf{C}, \mathbf{P}, \mathbf{M}} \sum_{k=1}^K U_k(R_k) \quad \text{s.t.} \quad \begin{cases} \sum_{k=1}^K \sum_{n=1}^N c_{k,n,l} p_{k,n,l} \leq P_T \\ 0 \leq \sum_{k=1}^K c_{k,n,l} \leq Q \\ p_{k,n,l} \geq 0 \\ c_{k,n,l} \in \{0, 1\} \\ m_{k,n,l} \in \mathcal{M} \end{cases} \quad (5.14)$$

We state the hardness of the problem (5.14) in the following theorem.

Theorem 5.2.2. *The optimization problem (5.14) is NP-hard even for $L = 1$.*

Proof: We show that the optimization problem is a classical knapsack problem [155] even if $L = 1$. Only in this proof we ignore the subscript l . The knapsack problem can be mathematically expressed as: for k kinds of items, 1 through K , each kind of item k has a value v_k

5. SCALABLE RESOURCE ALLOCATION FOR SCALABLE VIDEO STREAMING OVER WMANS

and a weight w_k , fit in a knapsack by

$$\max \sum_{k=1}^K x_k v_k \quad \text{s.t.} \quad \sum_{k=1}^K x_k w_k \leq W \quad (5.15)$$

If we simply define the video utility as the rate allocation to the user in current OFDM symbol, $U_k = \sum_{n=1}^N c_{k,n} m_{k,n}$, the optimization problem in (5.14) is readily to be a knapsack problem as

$$\max \sum_{k=1}^K \sum_{n=1}^N c_{k,n} m_{k,n} \quad \text{s.t.} \quad \sum_{k=1}^K \sum_{n=1}^N c_{k,n} p_{k,n} \leq P_T \quad (5.16)$$

Because the presented instance is only a subset of the original problem and it is NP-hard already, a more general case with nonlinear utility function must also be NP-hard. The simplified problem is more complex to solve than the classical knapsack problem due to the existence of the nonlinear rate-power function, as well as the cochannel user interference which are countered in (5.4) and (5.5). Therefore, our one symbol level utility optimization is NP-hard, and so is the problem (5.14). $\square$

Due to the computational intractability of the NP-hard problem, a multitude of greedy solutions can be found from the literature. However, most of existing schemes are not applicable to our utility maximization framework as the rate maximization is their key concern. However, useful lessons can be learned from the existing work on how the network efficiency can be exploited in a similar multiuser multi-slot system. Theoretically, bit loading can optimally solve the rate maximization problem in single user multisubcarrier systems [63]. For multiuser multisubcarrier systems, the greedy algorithm for the single-user case can be extended in a natural way as long as the power cost due to multiuser interference is counted [64], with computational complexity increasing as well. Unfortunately, the problem becomes more complicated in the multiantenna multisubcarrier environment. As users are allowed but restricted to share the same frequency band, allocating bits to a subcarrier for a user essentially leads to cochannel interference which relies on the user pairing as well as the technique to split coexisting signals. Nevertheless, the optimality analysis in [64] promises a bit loading solution for our efficiency problem. When there is low interference among users in all subchannels, the multiuser bit loading algorithm obtains the optimal result. By restricting the user number in a unit no more than the transmitter antenna number, the spatial precoding can achieve a convincing channel separation. Inspired by these hints, we anticipate an efficient bit loading solution to solve the problem (5.14) with only slight deviation from the optimality, but rewarded with highly reduced complexity.

To proceed from the previous bit filling scheme for fairness, another bit filling strategy is designed, as has been presented from Line 10 to 12 in **Algorithm 2**, to improve the overall efficiency for all users. This scheme is more suitable for the SVC packet transmission and can also be tied with the fairness scheme to constitute a complete solution for scalable resource allocation framework for SVC video delivery. The efficiency scheme tries to increase the modulation order by one for the user unit group (k^*, n^*) at the OFDM symbol l^* such that the utility increment by unit of the power resource can be maximized. At each filling step, consumption of each unit power resource will lead to the maximal utility increasing, so the bit filling strategy increases the overall video utility.

Step 1: The video utility increment by increasing one order of modulation is given by

$$\begin{aligned}\Delta U_k(R_k) &= U_k(R_k + \varepsilon b) - U_k(R_k) \\ &= D_k^{[P]}(R_k^{[P]}) - D_k^{[P]}(R_k^{[P]} + \varepsilon b)\end{aligned}\tag{5.17}$$

where εb comes from the bit increment for multiple subcarriers in a subchannel. ε is the bit increment in an OFDM symbol by proceeding to the next modulation order, and b is the number of subcarriers in a subchannel, as stated in Section 5.1.2.

Step 2: For all users ready for transmission into every unit, the user-unit pair to increase one order of modulation is selected by

$$\{k^*, n^*\} = \arg \max_{\{k, n | c_{k, n, l^*} = 1\}} \frac{\Delta U_k(R_k)}{e_{k, n, l^*}}\tag{5.18}$$

The utility to cost measure step-wisely approaches to the optimal solution by counting in each filling decision: the higher incremental video utility (video diversity), and lower incremental power costs (network diversity). This procedure is carried out iteratively until the power resource is used up. The computational complexity is bounded by $\mathcal{O}(LN(M + 1))$, for calculating the utility increment within the whole system frame.

5.2.5 Summary of the Solution

Our scalable framework is an integration of the fairness and efficiency schemes.

As the packet prioritization ensures that the base layer packets are invariably prioritized before the enhancement layer packets, we can easily demarcate the fairness and efficiency schemes at the transmission of base layer packets $\tilde{R}_k^{[B]}$ for user k . Before every user has the provision of fairness, $R_k \geq \tilde{R}_k^{[B]}$, the fairness scheme is carried out. Otherwise, the efficiency

5. SCALABLE RESOURCE ALLOCATION FOR SCALABLE VIDEO STREAMING OVER WMANS

Algorithm 2: The Joint Bit Loading and Power Allocation Solution for the Proposed Scalable Framework

```

1 for  $l^* = 0; l^* < L; l^*++$  do
2    $m_{k,n,l^*} = 0, p_{k,n,l^*} = 0$ 
3   Calculate initial cost  $e_{k,n,l^*}$  if  $l^* = 0$ , valid within  $L$ 
4   while True do
5     if  $\exists k: R_k < \tilde{R}_k^{[B]}$  then
6       Calculate minus utility gap
7        $\Delta U_k(R_k) = U_k(R_k) - U_k(\tilde{R}_k^{[B]}) = D_k^{[P]}(\tilde{R}_k^{[B]}) - D_k^{[P]}(R_k^{[P]})$ 
8       Select user that has a minimum gap by  $k^* = \arg \min_k \Delta U_k(R_k)$ 
9       Select a subchannel which has the least power cost for user  $k^*$  by
         $n^* = \arg \min_n e_{k^*,n,l^*}$ 
10    else
11      Calculate utility increment  $\Delta U_k(R_k) = D_k^{[P]}(R_k^{[P]}) - D_k^{[P]}(R_k^{[P]} + \varepsilon b)$ 
12      Select user-subchannel pair by  $\{k^*, n^*\} = \arg \max_{\{k,n\}} \frac{\Delta U_k(R_k)}{e_{k,n,l^*}}$ 
13    Modulation update
14    Beamforming vector update if  $k^*$  is newly added
15    Power update to satisfy a target bit error rate (BER) by referring to Eqs. (5.4)
        and (5.5)
16    If power is used up,  $\sum_{k=1}^K \sum_{n=1}^N c_{k,n,l^*} p_{k,n,l^*} > P_T$ , break
17    Rate update for user  $k^*$ 
18    Power cost  $e_{k,n,l^*}$  update for every user-unit pair  $(k, n)$ 

```

scheme is executed. The proposed schemes for fairness and efficiency are all designed in the bit filling strategy and are summarized in **Algorithm 2**. The switchover between the two schemes can be easily achieved. Lines 5-9 are for fairness and Lines 10-12 are for efficiency scheme, respectively. This greatly simplifies the implementation of the proposed schemes into practical applications. The analysis of fairness and efficiency explorations can be independently designed for a service objective of interest. The cut-off point can also be defined differently for specific applications.

5.3 Simulation Results

5.3.1 Simulation Settings

In this section, we evaluate the performance of the proposed scalable resource allocation framework. The downlink system is set according to the IEEE 802.16 standard [3]. The test setup is based on a MIMO-OFDM system simulator implemented in C programming language in [66] and the parameters at the PHY layer are selected referring to [156]. Three types of modulation schemes, QPSK, 16-QAM and 64-QAM, are adopted in the experimental evaluations. Other channel characteristics are referred to [66]. Users' channels are generated independently but with the same statistics. Users are placed and moved randomly within a 1.6km cell therefore the specific channel for a user depends on its geographical location and varies with time. Heterogeneous users requested different videos from a two-antenna base station simultaneously.

We adopt the JSVM 9.8 [133] to encode 14 different resolution videos which are retrieved by different clients. Videos encoded in QCIF (176×144) include: Akiyo (Aki), Bridge-Close (Bri), Carphone (Car), Container (Con), Hall-Monitor (Hal) and Mobile (Mob). Videos in CIF resolution (352×288) include: Bus, Coastguard (Coa), Foreman (For), Mother-Daughter (Mot), News (New) and Waterfall (Wat). Standard definition (640×480) videos contain Stockholm (Sto) and Pedestrian (Ped). Each video is encoded with a GOP size of 16 and frame rate 30 frames per second (fps). Each video GOP starts with an intra-coded frame. One base and one quality enhancement layers are encoded and within each quality layer, five temporal layers are realized in the hierarchical-B structure. We hypothesize that every user has a minimum requirement to receive a basic video for viewing, that is, quality base layer packets should be delivered to guarantee the fairness. Once every user successfully has its quality base layer received, the remaining radio resources are distributed to exploit efficiency. We show the video performance by using a widely adopted quality metric PSNR. As we usually measure the PSNR of the received videos with reference to the original sequence which has the maximum attainable frame rate, we will employ a simple error concealment strategy, frame copying to recover the frame loss. This strategy can apply in the situation where sequences are encoded with varying frame rate. We only need to do the frame copy to the maximum attainable frame rate and calculate the PSNR frame by frame. We can see from the subsequent experiments that the guarantee of the basic video viewing could also lead to the PSNR increasing. This may or may not be always true once advanced error concealment techniques are used, but the scalable framework can always be applicable, especially for multiple scalable video streaming.

5. SCALABLE RESOURCE ALLOCATION FOR SCALABLE VIDEO STREAMING OVER WMANS

We validate the proposed framework and its solution from three aspects: First, we verify the fairness strategy and its solution via a commonly used fairness index. Then we validate the efficiency scheme. To show that such a scalable framework is applicable for scalable video streaming, we distribute resources first for fairness and then for efficiency in the third validation. For every simulation, the proposed solution will be compared with the following state-of-the-art resource allocation schemes which can be applicable in multiuser MIMO-OFDM systems: 1) The conventional Round-Robin (RR) resource allocation scheme. 2) Linkgain-Based Resource Allocation (LBRA) scheme [61]: It was designed only for a single antenna system and was not for SDMA. We modify here to select the Q best users for the multiantenna system. 3) Multiantenna Multiuser Maximum Sum Rate (MMSR) scheme in [62]. 4) Adaptive Radio Resource Allocation (ARRA) scheme in [66].

We conduct 30 runs of experiments for each simulation. Because the location of users and their movements are random, the channel realizations are uncertain at each run. It is the aim of this simulation to analyze different schemes' efficiency in resource distribution among users. However, it is impossible to get a view on the resource distribution by the average results because the average effect obliterates the distinction between users. To overcome this, we select a representative result from one of the 30 runs to analyze the difference. The one is selected by a minimum deviation criterion: for the average bitrates from all five schemes (the proposed and four references), the one leads to the minimum mean absolute difference with respect to the averages of all 30 runs is selected as a representative. This is valid due to the following justifications: 1) the average rates of the selected run are almost the same as the averages of all 30 runs, and 2) the detailed results from different users reflect the performance of different schemes in one of the normal channel realizations.

5.3.2 Performance Evaluation of the Fairness Scheme

We first evaluate the performance of the fairness scheme. To guarantee a fair resource allocation for the reception of the quality base layer packets, we adopt only the scheme presented in **Algorithm 2** between Lines 5-9. We assume that there are not sufficient resources to successfully transmit all the quality base layer packets, so the achieved utility by the resources delivered is less than the utility when all base layer packets are received.

To make comparisons on the fairness fulfillment for different resource allocation schemes, we measure the utility gap between that obtained by the received bitstream with the one when all base layer packets are received. To have a quantified analysis, we calculate the Jain's fairness

Table 5.2: PSNR (dB) of the received videos from the fairness scheme

Res.	QCIF						CIF						SD		AVG
Video	Aki	Bri	Car	Con	Hal	Mob	Bus	Coa	For	Mot	New	Wat	Sto	Ped	
RR	41.10	32.64	25.37	21.31	24.35	17.00	11.81	16.18	14.49	29.21	21.80	19.40	14.02	15.23	21.71
LBRA [61]	41.10	29.91	18.24	42.05	35.35	34.58	11.85	16.17	16.88	24.69	20.06	19.61	36.47	24.16	26.51
MMSR [62]	41.10	34.71	24.27	42.05	39.92	34.58	15.85	30.90	20.43	23.94	36.81	23.60	36.47	35.69	31.45
ARRA [66]	41.10	29.54	22.92	42.05	36.76	34.58	15.46	22.82	22.26	23.94	26.51	22.11	36.26	28.34	28.90
Proposed	32.23	29.84	33.57	29.76	30.38	26.74	26.41	26.50	33.38	32.97	30.73	28.96	28.69	34.04	30.30
Efree	41.10	35.43	40.84	42.05	39.92	34.58	31.86	31.86	39.77	38.14	37.26	34.73	36.47	39.19	37.37

Figure 5.5: Jain's fairness index for utility gap of different video GOPs.

index [157] for the utility gap for videos in different GOPs, obtained from different resource allocation schemes. The index obtains its maximum value 1 when the utility gaps are identical for different videos. The indices with respect to consecutive GOPs are presented in Fig. 5.5. We can observe that the proposed fairness scheme can always maintain the utility gap among different videos approaching 1, while all reference schemes perform unsatisfactorily in guaranteeing fairness. The RR scheme, which is the best one among references, obtains an index that is less than 0.5. In particular for those like LBRA and MMSR, which dedicate to network efficiency exploitation, will always distribute resources to favorable users, leading to a high degree of unfairness. The scores below 0.4 indicate a high degree of unfairness by the reference schemes.

Experimental results on video quality are presented in Table 5.2. The validations show that the reference schemes which set throughput maximization as the objective function are not applicable to video streaming. The PSNR of the error free (Efree) videos are listed for reference. We can observe from the table that the proposed scheme in **Algorithm 2** can ensure

5. SCALABLE RESOURCE ALLOCATION FOR SCALABLE VIDEO STREAMING OVER WMANS

Table 5.3: PSNR (dB) of the received videos from the efficiency scheme

Res.	QCIF						CIF						SD		AVG
Video	Aki	Bri	Car	Con	Hal	Mob	Bus	Coa	For	Mot	New	Wat	Sto	Ped	
RR	41.10	35.43	35.88	42.05	39.92	12.79	15.55	19.87	14.49	38.14	20.32	20.58	14.03	14.54	26.05
LBRA [61]	41.10	35.43	40.84	42.05	39.92	34.58	29.44	31.86	14.49	38.14	21.93	20.40	36.47	14.54	31.51
MMSR [62]	41.10	35.43	40.84	42.05	39.92	34.58	31.86	31.86	13.25	38.14	36.76	34.73	36.47	15.67	33.76
ARRA [66]	41.10	35.43	40.84	42.05	39.92	34.58	25.80	31.86	22.76	38.14	25.52	19.88	36.47	22.39	32.62
Proposed	39.49	32.35	39.78	42.04	39.79	34.50	29.85	31.45	36.85	37.83	36.95	34.72	33.89	39.05	36.32
Efree	41.10	35.43	40.84	42.05	39.92	34.58	31.86	31.86	39.77	38.14	37.26	34.73	36.47	39.19	37.37

fairness among multiple users. Every user is entitled to a fair opportunity to have their base layer packets received and quality degradations among multiple users are close to each other. In the reference schemes, some users can receive high quality videos at the expenses of others. The deprived users cannot have adequate resources to receive the important base layer packets, while users with favorable channel conditions can even have enough resources to receive all their enhancement layer packets. From the system view point, not only the fairness is compromised but the network utility is not optimized. The loss of the base layer packets leads to severe quality degradation so we can see that all reference schemes except for the MMSR perform worse than the proposed scheme in the average video quality. The MMSR performs the best in average PSNR of videos, but the fairness is markedly compromised, e.g., Car, Bus, For and Ped which are too distorted to have a satisfactory viewing. Take the videos Bus and New as an example, video New obtains sufficient resources to transmit most of its video packets, whereas the video Bus only has a few packets being successfully delivered. The quality degradation diverges so high that no conspicuous visual distortion can be perceived from video New but on the other hand, video Bus is greatly contaminated and little useful information could be obtained. The outcome is the proposed scheme, succeeding in guaranteeing fairness, loses only around 1 dB in average video quality compared to the best reference scheme MMSR.

5.3.3 Performance Evaluation of the Efficiency Scheme

We validate the efficiency scheme where only the efficiency solution, Lines 10-12 in **Algorithm 2**, is enabled. Table 5.3 summarizes the received video quality under different resource allocation schemes.

The RR scheme only distributes resources alternately to users without considering the video utilities and channel status, so it leads the worst performance. The LBRA, superior in single antenna systems, allocates resources only based on channel gain experienced by the individual user so it cannot handle cochannel user interference in a multiantenna system. The MMSR

considers the cochannel user interference and always distributes resources to users that can obtain the most data rate increment; therefore it performs the best in the average rate. However, the average PSNR from all reference schemes is less than that from our proposed efficiency scheme. This validates that the throughput maximization is not equivalent to video utility maximization. In wireless video applications, not only the channel states but also the characteristics of video packets should be taken into account.

The proposed efficiency scheme transmits less amount of bitstream compared to MMSR because resources allocated to the user that can achieve more utility increment but the channel gain may not be the highest at the same time. As a result, it sacrifices on throughput but can acquire increased overall video quality. The reference schemes are incompetent for packet dependent SVC video transmission as they either ignore the channel status or the video packet characteristics. The lopsided mechanisms lead to an unbalanced resource distribution. Users with superb channels consume a lot of resource for their low priority video packets. Consequently, no sufficient resources are left for users in the cell edge to transmit their base layer packets, giving rise to notable quality degradation. Taking the reference scheme ARRA as an example, it can perfectly receive all QCIF videos. The inadequate resources for videos For, Wat and Ped lead to severe quality degradation more than 10 dB. Consequently, although the reference scheme can get perfect pictures for certain videos, the average quality degradation is still 3.5 dB compared to the proposed scheme. The proposed scheme considers the packet differentiation in quality contributions. It allocates resources to the users who can achieve more utility increasing by a unit of power in each loading procedure. This guarantees that users with more important packets ready for transmission can win high chances of transmission even if they are in poor network condition. The successful delivery of more important packets from all users leads to notable increasing in the summation of video utilities. From the view point of the whole system, the overall network efficiency is increased.

5.3.4 Performance Evaluation of the Overall Scalable Resource Allocation Framework

To convince that the proposed scalable framework is applicable to the streaming of scalable videos, we design an experiment to exploit fairness and efficiency at the different transmission stages. To provide every admitted user a basic video viewing, we explore fairness before every user has received its quality base layer video. After the basic video viewing is satisfied for every user, resources are distributed to take advantage of the network efficiency. Listed in

5. SCALABLE RESOURCE ALLOCATION FOR SCALABLE VIDEO STREAMING OVER WMANS

Table 5.4: PSNR (dB) of the received videos from the overall scalable resource allocation framework

Res.	QCIF						CIF						SD		AVG
Video	Aki	Bri	Car	Con	Hal	Mob	Bus	Coa	For	Mot	New	Wat	Sto	Ped	
RR	41.10	35.43	39.67	42.05	39.92	20.83	18.41	15.64	19.88	38.14	37.26	34.66	20.61	17.53	30.08
LBRA [61]	41.10	35.43	40.84	35.73	39.92	34.58	31.86	31.86	20.25	38.14	37.26	34.73	36.47	16.62	33.91
MMSR [62]	41.10	35.43	40.84	42.05	39.92	34.58	31.86	31.86	34.10	38.14	37.26	34.73	36.47	23.20	35.82
ARRA [66]	41.10	35.43	40.84	40.28	39.92	34.52	31.86	31.86	28.52	38.14	37.26	34.73	36.47	24.62	35.39
Proposed	41.04	35.40	40.83	41.52	39.89	34.18	31.81	31.82	36.85	38.10	37.25	34.72	36.38	36.47	36.87
Efree	41.10	35.43	40.84	42.05	39.92	34.58	31.86	31.86	39.77	38.14	37.26	34.73	36.47	39.19	37.37

Table 5.4 are the summary of the experimental results for transmission of SVC videos under different resource allocation schemes.

Experimental results show that the proposed scalable framework ensures that all users can be gratified with their minimum requirements. Whereas all reference schemes lead to different degrees of unbalance; some users cannot obtain the desired video viewing. Nearly all reference schemes cannot meet the basic requirements for the motion-intensive videos, such as For and Ped, which are generally encoded in large packet sizes. They also cannot satisfy users that experience severe channel conditions, such as Con in this instance. However, we can find that for every reference scheme, there are clients that can perfectly receive all packets. The receiving of the enhancement layer packets for some users cannot achieve manifest quality improvement, whereas the loss of base layer packets for other videos will lead to severe quality degradations. For example, the SD videos are commonly encoded with very large packet size. The MMSR scheme spends nearly all available resources to send quality enhancement layer packets to come up with an undistorted video Sto. But the consequence is that the other SD video Ped can hardly get an opportunity to have its packets delivered. The video Ped itself is also encoded with large packet sizes, but only a small percentage of packets have been successfully received. A degradation of 16 dB in PSNR greatly affects the overall system performance. The loss is so high and the overall system performance cannot be made up, even though some low resolution videos (QCIF in this case) have been perfectly received. A group of received pictures are presented in Fig. 5.6 for subjective evaluation. Note that the QCIF and CIF pictures have been scaled for ease of viewing. It can be clearly found that the distorted SD pictures from reference schemes cannot present its scene faithfully. This introduces unsatisfactory viewing experience.

To summarize, the blindfold RR scheme ignores both the video utility and network condition so it could not be an efficient solution. The LBRA and MMSR focus only on the channel

Figure 5.6: Subjective pictures from different resource allocation schemes. (a) RR, (b) LBRA [61], (c) MMSR [62], (d) ARRA [66], and (e) the Proposed.

condition user encounter but lose sight of the characteristics of different video. The ARRA comes out only on the deadline information but neglects the inherent packet difference. The proposed scalable framework is applicable in handling different situations which makes it useful in a practical wireless system. Especially we can find that the reference schemes are less efficient in handling heterogeneous clients which lead to heavy packet loss when packet sizes are large for some video sequences. The utility measure presented in this chapter well reflects the packet utility and serves as an accurate measure in multiuser video transmissions. As we focus on guaranteeing fairness at the start of each GOP, the efficiency is compromised in the scalable framework, as compared to the gain obtained in Table 5.3. By comparing to the best reference scheme MMSR, we get 2.56 dB of quality improvement in efficiency exploitation in the Table 5.3. Since we consider the fairness guaranteeing in this combined scenario, the improvement decreases to 1.05 dB as can be found in Table 5.4. Nevertheless, the fairness scheme is critical for an impartial provision to meet the minimum requirement of all users, which is especially useful in ubiquitous wireless networks where terminals gradually diverge.

5.4 Concluding Remarks

We propose a scalable resource allocation framework to optimize system performance for scalable video transmissions over the multiuser MIMO-OFDM systems. We aim at transmitting scalable video packets by differentiating the service goal at different transmission stages. This

5. SCALABLE RESOURCE ALLOCATION FOR SCALABLE VIDEO STREAMING OVER WMANS

chapter is based on the optimized rate-utility analysis of the SVC video from Chapter 4. To provide every admitted user a fair opportunity to obtain a basic viewing experience, we first distribute resources via a fair scheme. After that, resources are allocated by maximizing the overall system utility. The scalable framework is effective as it offers a fair opportunity for heterogeneous video terminals to have a basic view of video programs, and upon this, network efficiency can be exploited if resources are abundant. The cut-off point could also be defined by applications. Our framework provides the trade-off between user experience and system performance. To validate the performance of the proposed framework, we make comparisons with several state-of-the-art resource allocation schemes. Notable outperformance by the proposed scheme is achieved when scalable videos are transmitted through the multiuser MIMO-OFDM networks.

Up to now, we have focused on the video provision with some objective video quality metrics. The objective metrics provide fast yet efficient way to measure the video quality in streaming services. However, it has always been argued that these objective metrics fail to accurately reflect the user experience. A more advanced and accurate concept that can characterize the acceptability of the end users is Quality of Experience (QoE). It is natural to think about how the resource allocation and scheduling schemes should be designed for wireless streaming when QoE serves as the target object. The research will be addressed in Chapter 6.

QoE-aware Scalable Video Adaptation and Streaming

6.1 Introduction

We have discussed advanced resource allocation and packet scheduling schemes in previous chapters which are devoted to providing video service under some objective quality metrics. However, media service providers nowadays are keen on improving the user's perception of quality, commonly known as the Quality of Experience (QoE) [93]. User's perception of streaming video characterized in terms of QoE is directly measured by the acceptability of the end users [158]. QoE is more related to but differs from the extensively studied concept Quality of Service (QoS) [83], which mainly focuses on system related parameters. However, QoS lacks the human perception in quantifying the perceived video quality [94]. The QoE measurement has been shown to be quite different from the objective quality metrics such as MSE and PSNR [23, 29, 93]. Researchers have argued that structured approach to QoE can be designed based on multiple essential attributes from the user and technical aspects [26, 159] and validated by the mean opinion score (MOS) [20]. MOS evaluations tend to be computationally intensive, cumbersome, not repeatable, and are often hard to adapt to real time quality assessment [24]. Recently, there are increasing efforts on modeling QoE for wireless streaming. A non-reference evaluation framework has been presented in [99] by constructing a k -dimensional QoE space. The space is partitioned into N representative zones, each with a QoE index. Instantaneous parameter values are matched with the indices to infer QoE. Other

6. QOE-AWARE SCALABLE VIDEO ADAPTATION AND STREAMING

important references include QoE modeling [100, 101], QoS to QoE mapping [102], etc. However, the design of the QoE-aware video streaming is still an open issue. Following the general guidelines in [159] that QoE improvement should be tied to combined factors, we introduce in this chapter a QoE-aware video streaming scheme over wireless networks.

Considering that the scalable video coding is an ideal choice for wireless video applications due to its flexible rate adaptation and bitstream regulation mechanisms to cope with bandwidth scarcity and network fluctuation, we again employ the SVC [10] in the study of video QoE. SVC is encoded with a most important base layer whilst quality refinement or resolution increase is realized by successive encoding of more enhancement layers. As wireless terminals are confined by diverse hardware capabilities and network conditions, scalable video can serve heterogeneous clients with a one-time encoded bitstream. The rate adaptation and bitstream manipulation can be easily achieved by discarding some enhancement layer packets to conform to the resource capacity. Multidimensional scalability of SVC enriches the selection of methods for rate reduction but on the other hand, it complicates the selection of the most favorable scalability adaptation track in terms of QoE maximization. Here, a key concern is how to carry out the rate tailoring, which scalability should be adapted first, such that we can always obtain the best QoE at a given rate. Some earlier attempts have studied this issue but they mostly focused on partial scalability [160, 161, 162, 163, 164, 165]. A recent work [166] investigated the impacts of spatial, temporal and amplitude resolutions (STAR) on the bit rate of a compressed video. The authors proposed a rate model where the rate was expressed as the product of power functions of the quantization step size, frame size and frame rate. This model is analytically tractable and has been validated in both scalable and non-scalable coding. However, the model needs content-dependent calculations for all spatial/temporal/quality subsets of videos to estimate the coefficients, which is time-consuming. Moreover, no quality or QoE model related to the three scalability factors has been derived. Comprehensive evaluation has been conducted in [167] to investigate the content distribution of scalable video coding. Some useful guidelines have been summarized by comparing two to six instances in different scalability options. Some general conclusions were drawn through the comparisons, but no explicit rate-quality model has been summarized. Consequently, it is not feasible to achieve the full-range rate adaptation and hence cannot be applied to analyze resource allocation problem, especially for multiple SVC video transmission. To sum up, these innovations have introduced some insights for understanding quality evaluation in SVC video, and motivate us to further

investigate the QoE-aware rate adaptation and resource allocation considering full scalability of SVC.

We will consider scalable video streaming over the same multiuser MIMO-OFDM systems as in Chapter 5, but with focus on the provision of QoE for the end users. In this chapter, exploiting spatial, frequency, time and user diversities to increase the end users' overall QoE is of interest. We have reviewed some resource allocation schemes for MIMO-OFDM systems in Chapter 5. We will discuss some reference work in this chapter to highlight their inefficiency in QoE provision. An intuitive and simple implementation is the round-robin resource allocation, where multi-users are served alternatively in resource blocks. Fairness is beyond control and resource is not fully utilized since the network and data characteristics are not taken into consideration in such scheme. Some adaptive transmission schemes have recently been proposed for the multi-antenna systems. Nevertheless, most existing schemes dealt with normal data streams are focused on the physical (PHY) layer only. The proposed strategies often fail in video transmission since video packets have rather stringent delay constraint and packet dependency. In two representative work [62] and [66], the authors considered the transmission of different types of traffics over MIMO-OFDM networks. They employed simple traffic models, but failed to exploit the intrinsic properties of scalable video packets such as differentiated packet priority, bursty packet sizes and dependencies. Nevertheless, all these existing methods fail to exploit the multiuser diversities and system capacities, especially the human factor QoE. Therefore, user experience cannot be guaranteed in these scalable video streaming schemes.

To address the QoE issues in scalable video adaptation and streaming, we present a QoE-aware resource allocation framework for SVC video transmission over multiuser MIMO-OFDM networks in this chapter. Our contributions can be summarized as follows:

1. A QoE assessment database has been developed specifically for SVC adaptation. The QoE impact according to full scalability, i.e., spatial, temporal, and quality scalabilities, is investigated. Based on the QoE ratings for all extracted video clips from the SVC bitstreams, the optimal scalability adaptation track for each individual video source and the common scalability adaptation tracks for grouped videos are derived and evaluated. These analyses provide general guidelines for SVC video adaptation and for streaming scalable video over communication systems.
2. A rate-QoE model is derived to approximate the rate-MOS relationship exploited in our QoE assessment database for SVC adaptation. The model not only can be employed as

a predictor of QoE, but more importantly, it provides the foundations for the QoE-aware resource allocation in the multiuser video streaming systems.

3. A QoE-aware resource allocation (QoERA) scheme is proposed to maximize users' QoE for multiple SVC video streaming over the multiuser MIMO-OFDM systems. Referring to the rate-QoE model derived by scalability adaptation, the proposed scheme integrates the spatial multiplexed time-frequency unit assignment for multiple users, power allocation and modulation selection for each unit to maximize the summation of QoE for all users.

The proposed QoE-aware resource allocation scheme overcomes the shortcomings in existing methods which neglect either the QoE characteristics or the users' network status. As we will demonstrate in this chapter, superior performance is achieved by the proposed scheme. This work has been published in [168] and [169].

6.2 Preliminaries and Related Work

SVC consists of a Video Coding Layer (VCL) and a Network Abstraction Layer (NAL). VCL follows a block-based hybrid coding structure and is extended to support three types of scalabilities, i.e., spatial, temporal and quality scalabilities. The NAL provides each packet the temporal, spatial and quality layer IDs and other related information. In this chapter, we use d , t and q to indicate the dependency layer ID (Did), temporal layer ID (Tid) and quality layer ID (Qid), respectively. For details on SVC, please refer to [10].

Due to the intrinsic prediction relationship of video packets from different layers, SVC facilitates flexible rate adaptation mechanisms for time-varying wireless channel and heterogeneous mobile clients. The flexible scheme provides various possibilities for rate regulation due to the existence of multiple scalabilities. To show the multi-dimensional scalability adaptation of SVC, we present again the generic SVC encoding structure in Fig. 6.1 and a typical prediction structure for the full scalability coding in Fig. 6.2, where two spatially dependent layers are depicted within one Group of Pictures (GOP). The lower part of the Fig. 6.2 explains how the quality scalability is realized by employing the key-picture concept in medium-grain quality scalability. The upper part provides an illustration of the spatial scalability by the inter-layer prediction. For simplicity, we omit the quality prediction in the higher spatial layer in

Figure 6.1: SVC encoding structure (based on Fig. 12, page 1116 of [10]).

Figure 6.2: Multilayer structure in SVC with inter-layer prediction, in which each packet corresponds to a (d, t, q) layer. Upper: Prediction structure for spatial scalable coding. Lower: Prediction structure for quality scalable coding. Spatial layer $a + k$ is predicted from layer a , where discardable flag $x = 0$ indicates that layer a is employed for inter-layer prediction.

this figure. Temporal scalability is realized by the hierarchical-B structure which can be easily envisaged in any a spatial or quality layer.

6.2.1 Adaptation Schemes in JSVM

As three types of scalabilities are jointly implemented, how to optimally adapt the rate by properly discarding enhancement layers to reach the best user experience is a major concern. In other words, regulating the rate by finding an optimal scalability adaptation track is of our focus. SVC provides a bitstream extraction scheme [170] for scalability adaptation but it is less efficient in guaranteeing QoE in SVC video adaptation, as we will show later. The scheme only aims at maintaining a good peak signal to noise ratio (PSNR), which is not suitable to measure QoE [93], for the highest spatial level pictures. Specifically, quality base layers at different spatial layers are prioritized first, upon which packets from lower-than-the-highest spatial layers are followed with temporal layer orders. Finally, packets within the highest spatial resolution are ordered in quality layer orders. The prioritization scheme is fixed and is not suitable for different coding configurations, resulting in worse video quality. In addition, a Quality Layer based scheme in [138] proposes to use effective Rate-Distortion (R-D) optimization to extract packets from the SVC bitstream where the Quality Layer information is associated with each NAL unit. Quality Layer information is derived by calculating the Rate-Distortion value of each picture. This method is computationally expensive and only has a confined bitrate range for adaptation. More importantly, these mechanisms are not designed based on the QoE criterion. This motivates us to find a QoE-aware adaptation solution which can also be utilized in resource allocation for SVC video streaming.

6.2.2 Analytical Rate and Quality Modeling

A series of pioneering work, [163, 166, 171, 172, 173, 174], etc., has been proposed to comprehensively study the optimized rate and quality modeling. For example in [163], the authors explore the impact of frame rate and quantization parameter on the perceptual quality. A content-dependent quality model has been proposed and validated by subjective tests. In [171, 172], the authors extend the quality model to include all scalability factors in SVC, i.e., spatial resolution (SR), temporal resolution (TR) and quantization stepsize (QS). Subjective tests validate their quality model where the joint impact of SR, TR and QS is revealed each by a content-dependent parameter. Besides, the work in [166, 173, 174] models the rate as function partially or completely of SR, TR and QS. By combining the quality and rate models, the optimized rate adaptation can be readily formulated as to select the best combination of SR, TR and QS under a given rate constraint. The rate-quality modeling studies all SR, TR and QS

combinations. Nevertheless, these methods need some pre-processing to derive some relevant parameters. We aim to seek an alternative way to tackle this problem. We seek to summarize some general guidelines for direct video adaptation. We do not intend to derive rate and quality as function of SR, TR and QS on all (d, t, q) combinations. Instead, we aim to discover only the combinations lying in the optimal adaptation track and based on them, to summarize the guidelines for rate adaptation. The proposed method is especially attractive in some real-time video streaming applications where finding a fast yet efficient mechanism is more important than seeking an optimal solution. The reference work in [163, 166, 171, 172, 173, 174] studies all possible adaptation points in an SVC bitstream and can be employed for the most complete investigation under various applications. We only focus on the optimal adaptation track since the focus of this work is on resource allocation during streaming. The disposal is adequate for the problem at hand. On the other hand, the study of the SR, TR and QS tradeoff is somewhat different in our work. In [174], an optimal Q(R) model (quality with respect to the rate) has been constructed. The optimal adaptation scheme in [174] suggests that with the rate increases, spatial and temporal resolutions will increase while amplitude resolution should decrease, and the achievable best quality continuously improve. This analytical results do not comply with the inherent SVC codec. In SVC, spatial layers are successively and “independently” encoded. The inter-layer prediction can choose any quality layer at the lower SR. Hence, it is possible to increase QP between spatial layers with increased rate. Therefore, we will consider for a complete SVC encoded bitstream, how to progressively exclude scalability layers to achieve rate adaptation. It is still meaningful to explore more on this topic in an alternative way. Nevertheless, the reference work is the pioneering study in this topic and provides the most complete and fruitful results. We will, based on their work, set up our system, carry out the data processing and present a simplified yet efficient solution to solve the problem from another angle.

6.3 QoE Database for SVC Adaptation

In this part, we construct a QoE assessment database for our proposed QoE-optimized SVC adaptation scheme.

6.3.1 Test Materials

We design subjective video quality assessment tests and the videos are selected from the Joint Video Team (JVT) site [175]. To produce the test materials, we considered ten video sequences

6. QOE-AWARE SCALABLE VIDEO ADAPTATION AND STREAMING

Figure 6.3: Snapshots of the training video sequences from JVT [175]. Left: Aspen. Right: In.to.tree.

Figure 6.4: Snapshots of the tested video sequences from JVT[175]. Upper: City, Crew, Harbour and Soccer. Lower: Ducks, Ice, Oldtown and Park.

in raw progressive format. All videos are encoded using the SVC reference module JSVM 9.19 [170] with three spatial layers containing 4CIF, CIF and QCIF resolutions. We cropped the high definition (HD) videos down to 4CIF resolution then to CIF and QCIF resolutions by the cropping tools provided in the JSVM, the same as in [142].

The subjective assessment is composed of two sessions. A training session is to let the subjects accustom themselves to the video specifications in terms of different types of scalabilities and rating procedures. The subjects are encouraged to raise questions relating to the quality assessment in the training session. The training video sequences, including Aspen and In.to.tree were chosen to expose subjects to the types and spatial, temporal and quality ranges. The snapshots of the training sequences are shown in Fig. 6.3. To investigate the QoE impact of three types of scalabilities, we first included four typical video sequences, City, Crew, Harbour and Soccer as suggested in [142]. To get a full view of various video sequences and make the database more complete, we included another four videos, Ducks.Takes.Off (Ducks

for short in later sections), Ice, Old_Town_Cross (Oldtown for short) and Park_Joy (Park for short). The snapshots of the selected test video sequences are shown in Fig. 6.4. The inclusion of eight videos ensures the accordance with the suggestion in [22] which claims that at least four different types of scenes should be chosen in order for a minimum reliability of the results. The selected video sequences are representative of different levels of spatial and temporal complexities, as computed by means of the Spatial Information (SI) and Temporal Information (TI) indices in the recommendation [22] over the 4CIF videos. In [22], the SI is computed based on the Sobel filter. Each luminance video frame at time $n(F_n)$ is first filtered with the Sobel filter $[Sobel(F_n)]$. The standard deviation over the pixels in each Sobel-filtered frame is then computed. The maximum value in the time series is chosen to represent the spatial information content of the scene, which can be mathematically expressed as [22]:

$$SI = \max_{time} \{std_{space}[Sobel(F_n)]\} \quad (6.1)$$

The TI information is calculated based on the motion difference [22]. A value $M_n(i, j)$ is calculated first as the luminance difference between two consecutive frames for frame n at pixel (i, j) . The TI is defined in [22] as “the maximum over time ($\max_{time}$) of the standard deviation over space (std_{space}) of $M_n(i, j)$ over all i and j ” (Section 5.3.2, page 5 of [22]). The mathematical form can be shows as [22]:

$$TI = \max_{time} \{std_{space}[M_n(i, j)]\} \quad (6.2)$$

In Fig. 6.5, we show the SI-TI values of the selected videos. The broad distributions of the indices in the whole space show an ample coverage of videos in different spatial and temporal characteristics. Therefore the selected videos in our database are considered to be sufficient for the QoE study.

We employed JSVM 9.19 [170] to generate three quality layers in 4CIF and CIF spatial levels. Within the QCIF resolution, we only encoded two quality layers considering the limited display size. The Quantization Parameter (QP) in SVC encoding configuration was set to 40, 33, 26 for three quality layers, respectively. Following the same disposal as in [163], we fixed the QP within a GOP for different temporal layers. GOP size was set to 16 for each video, and one encoded bitstream was composed of at most five temporal layers corresponding to frame rates of 60, 30, 15, 7.5, 3.75 frames per second (fps). Two adjacent spatial layers were in dyadic spatial and temporal resolutions. Spatial layers were tagged with three *Dids* 0, 1, and 2, from the base to the highest enhancement layer. Within the spatial layer 2, the

Figure 6.5: Spatial Information (SI) and Temporal Information (TI) computed for the test sequences based on [22].

largest Tid and Qid were 4 and 2, respectively. Those in the spatial layer 1 were 3 and 2, and those in QCIF resolution were 2 and 1. Within each spatial layer d , for temporal and quality layers ranging from ID 0 to the largest number, we extracted all (d, t, q) combinations using the Bitstream Extractor in JSVM [170] and decoded them into different specifications of video clips. Together with three uncompressed videos (a total of three in all spatial resolutions) in our experiments, 36 video clips were employed in the test for each video source. For the 8 video sources we included, which we would refer to as Original Videos (OVs) in the following context, a total of 288 video clips were generated in our database, which we would be referred to as Extracted Videos (EVs) in the following context. The EVs were displayed in random order to every subject for evaluation. It should also be noted that training videos were also shown at different qualities and frame rates, depending on which quality and temporal layers they are in. We showed the QCIF and CIF videos to subjects for evaluation in the 4CIF size as in [164] and [167].

Because we will frequently use various (d, t, q) combinations in the following context to interpret our scalability adaptation schemes, we summarize in Table 6.1 the congruent relationships of (d, t, q) combinations and their coding configurations.

Table 6.1: (d, t, q) combination and their coding configuration

d	2	4CIF spatial size
	1	CIF spatial size
	0	QCIF spatial size
t	4	60 fps
	3	30 fps
	2	15 fps
	1	7.5 fps
	0	3.75 fps
q	2	QP 26 for $d > 0$
	1	QP 33 for $d > 0$, QP 26 for $d = 0$
	0	QP 40 for $d > 0$, QP 33 for $d = 0$

6.3.2 Evaluation Environment and Participants

The viewing conditions were set up following the recommendations in [21] and [22]. Some recommended specifications are summarized as follows. The viewing distance was set to four times of CIF picture height to guarantee that all displayed videos were within the recommended range of 1-8 times of the picture height in [22]. We used the professional EIZO FlexScan SX3031W 29.8-inch wide screen LCD monitor. The display resolution was set to WQXGA (2560×1600). Display brightness and contrast were set up according to [21]. The peak luminance was within the range of 100-200 cd/m^2 . The ratio of the luminance of the screen, when displaying only black level in a dark room to that corresponding to peak white, was around 0.1. The background luminance was ensured to be lower than 20 lux in testing. As test videos had a smaller size compared to the monitor screen resolution, the videos were displayed at the center of the screen and the color of the background was set to 50% grey. As it is time-consuming to evaluate all 288 video clips throughout the one-time assessment, we split the assessed videos into two subgroups and conducted the evaluation independently. In each subgroup, we recruited 22 non-expert viewers, aging from 20 to 35, with normal or corrected-to-normal vision acuity to participate in our subjective quality assessment.

6.3.3 Evaluation Method

We adopted a single-stimulus (SS) test for evaluation based on the Adjectival Categorical Judgment Methods in Rec. ITU-R BT.500 [21], where the EVs were presented and rated independently on a category scale. The subjects were asked to evaluate the QoE of the video by

Figure 6.6: Stimulus presentation in the subjective assessment.

Table 6.2: Five-level MOS scales

Score	1	2	3	4	5
Quality	Bad	Poor	Fair	Good	Excellent

MOS [93]. The evaluation procedure is illustrated in Fig. 6.6. The presentation time for each test video was around 8-10 s as the videos adopted have standard 480-600 frames and were displayed maximally at 60 fps. The viewers voted and recorded their scores during the 5 s mid-gray post-exposure period. We recommended the viewers to give five-level MOS rating as shown in Table 6.2. The reviewers were asked to take a 10 minutes' break after testing half of the videos in each subgroup.

6.3.4 Data Postprocessing

We adopted, with some modification, the screening method recommended by BT.500-11 [21] and the method in [163] designed for single stimulus quality evaluation to remove outliers of subjects' QoE rating. The modification was required because the tested videos were in different specifications in our assessment. The outlier removal is executed in the following steps. First, we carry out a normalization procedure on the raw rating data because different viewers' rating might fall into different sub-ranges due to their personal preference or understanding. We first find the maximum and minimum scores given by each subject for an OV. The average maximum score for all OVs is taken as the new maximum bound in normalization, and so is the minimum score. All subjects' ratings for a single OV are normalized by the new maximum and minimum bound. Second, it must be ascertained as in [21] whether this distribution of scores for the test EV is normal or not using the β_2 test. For each EV the score of each viewer will be compared with the associated mean value, plus the weighted standard deviation. Every time an observer's voting falls into a range, counter P_i or Q_i will be increased. Finally, the

Algorithm 3: Description of data screening

```

1 For each viewer  $i$  and all EVs from an OV, start from  $P_i = 0$  and  $Q_i = 0$ ;
2 for  $j = 1; j \leq J; j++$  do
3     if  $2 \leq \beta_{2j} \leq 4$  then
4         if  $u_{ij} \geq \bar{u}_j + 2S_j$  then
5              $P_i = P_i + 1$ 
6         if  $u_{ij} \leq \bar{u}_j - 2S_j$  then
7              $Q_i = Q_i + 1$ 
8     else
9         if  $u_{ij} \geq \bar{u}_j + \sqrt{20}S_j$  then
10             $P_i = P_i + 1$ 
11        if  $u_{ij} \leq \bar{u}_j - \sqrt{20}S_j$  then
12             $Q_i = Q_i + 1$ 
13 if  $P_i > 3$  or  $Q_i > 3$  then
14      $\text{reject viewer } i \text{ for the current OV}$ 
    
```

values in counters P_i and Q_i are used to decide whether a user's voting should be accepted. Specifically, let u_{ij} denote the score of EV j by viewer i . We first calculate the mean, standard deviation, and Kurtosis coefficients for each EV as $\bar{u}_j$, S_j , and β_{2j} , respectively. The total number of viewers is I and the number of EVs for an OV is J . The Kurtosis coefficients are derived as [21]:

$$\beta_{2j} = \frac{m_{4j}}{m_{2j}^2} \quad (6.3)$$

where

$$m_{xj} = \frac{\sum_{i=1}^I (u_{ij} - \bar{u}_j)^x}{I} \quad (6.4)$$

For each view i and a specific OV, we then use **Algorithm 3** to identify viewers whose scores are far from the average score by all viewers and reject them. We select the threshold for P_i and Q_i as 3, the number of spatial layers as in [163].

Table 6.3: Four-way ANOVA on the average ratings

Source	SS	DoF	MS	F Stat.	p -value
v	6.50	7	0.93	10.09	4.23×10^{-11}
d	72.832	2	36.42	395.57	$< 10^{-15}$
t	71.34	4	17.83	193.72	$< 10^{-15}$
q	29.18	2	14.59	158.50	$< 10^{-15}$

6.4 QoE Analysis for SVC Rate Adaptation

6.4.1 ANOVA Analysis

After the data screening, we get an average rating for each EV. We now have a table of records where each (d, t, q) combination from a test video corresponds to an average MOS score. To understand whether there are differences in rating with respect to different d (spatial size), t (frame rate), q (quality) and v (different videos), we conduct the classic Analysis of Variance (ANOVA) test [176] on these data for validation. It should be noted that in the subjective assessment we have included three intact uncompressed videos for an OV. In the following context, we will focus on scalability adaptation analysis which investigates only on those compressed videos. Those uncompressed videos will be excluded in ANOVA and the following analyses. We conduct a four-way ANOVA to investigate how these factors influence the QoE of extracted videos. The results are reported in Table 6.3. The first column indicates the four factors considered: different video v , different spatial layer d , different temporal layer t , and different quality layer q . The second column is the sum of squares (SS) between different treatments of a factor. The third column is associated degree of freedom (DoF), which is defined as the number of treatments minus one. The forth column is the mean squares (MS) of the treatments which is equal to the SS divided by the DoF. MS is a measure of the variability of the group means around the grand mean. The fifth column shows the F statistics (F Stat.) and the last column is the p -value. The p -value for the test is represented as the area under an F -distribution to the right tail of the F Stat. We will reject the null hypothesis when the p -value is below the significance level α [176].

The value of α used in this work is set to be 0.01 according to [164]. We observe from the Table that the p -value of v is greatly lower than the significance level. We conclude that there is significant difference in the ratings for different videos, which implies that the ratings from the subjective assessment are video dependent. Besides the difference in video rating, the OVs are

Figure 6.7: Average ratings for video City in all (d, t, q) combinations.

also different from each other in its coding rate. This strengthens the inference the scalability adaptation track is video content dependent. This is a useful guideline for developing some general rules in scalability adaptation for different videos. It should be noted that for the tested videos, Soccer and Ducks significantly influence the results of the ANOVA analysis. Exclusion of these videos will increase p -value of v by several orders of magnitude. This implies that videos with specific characteristics, such as high motion video Soccer (large value of TI) or spatially complex video Ducks (large value of SI) will potentially be differently scored in the subjective assessment. It is also possible that these videos, characterized by unique SI and TI values, will have different scalability adaptation tracks. We will discuss further on this in the next section. We can also find from the ANOVA analysis that three types of scalability (different d , t and q) influence the QoE greatly (with p -value $\leq 10^{-15}$). This is persuasive because viewers should have preferences for high resolution and quality videos.

6.4.2 The Optimal Scalability Adaptation Track for Individual Video

With the ratings of all (d, t, q) combinations for a video source available, a quality model can be developed as in [171, 172]. Moreover, a rate model corresponding to every (d, t, q) combination is developed in [174]. By combining the two individual models, the authors can study the rate-quality relationship at all adaptation points [174]. Based on their rate and quality modeling, it is possible to investigate the changes of the rate and the quality on individual factors of spatial size, frame rate and quantization parameter. To show that our test data can

Figure 6.8: Average ratings for video City in all (d, t, q) combinations.

endorse the existing findings, we apply the Q-STAR model [172] on the tested data from video City in Fig. 6.7.

It can be seen from the figure that quality model generally fits well to our test data. Although the reported deviations are slightly larger than those in [172], it is still acceptable since a lot of factors impact on the quality assessment results and it is hardly to be identical for independent database setup. Because the focus of this work is on resource allocation, we are only interested in establishing the rate-QoE function in the optimal adaptation track. This analysis can be independent from the individual rate and quality modeling and thus can save a lot of time and efforts. To show an alternative way to tackle this problem, we determine the optimal scalability adaptation track directly from the test data for individual video source. By analyzing the optimal tracks, we intend to extend and summarize some general guidelines for fast video adaptation under QoE criterion. The guidelines are tremendously useful when new videos need to be adapted. And meanwhile, the rate-quality model can be constructed directly from the derived tracks. This ignores the modeling of points outside the tracks but the solution is adequate for the problem at hand. The readers can refer to [171, 172] to check how the d , t , and q will affect the subjective quality individually on all adaptation points, and can also refer to [174] to find the effect on the rate.

The optimal track is founded by tracing the upper bound of the rate-MOS data in our work. The optimal track ensures that each truncated scalable video will have the highest QoE at a given bitrate. Meanwhile, the optimal scalability adaptation track for the individual video source is the upper bound that is used for evaluation and validation of other adaptation tracks. To show that QoE indices decrease differently in three scalability directions, we take video City as an example and show the average ratings of all (d, t, q) combinations in Fig. 6.8. We show ratings from one spatial resolution in a specific color. Ratings from different (d, t, q) combinations can be compared by the height of the bars. This figure shows that MOS degradations are quite different in three scalability trajectories. Moreover, the degradation in one scalability direction always varies in different ranges. At a given rate, there could be multiple choices to carry out the rate adaptation. Two instances are shown in Fig. 6.9, where videos adapted to a close rate differ greatly in their perceptual experience. How to achieve the maximal QoE for viewers when carrying out the scalability adaptation is of our concern. We start by analyzing individual video. We plot the rate-MOS (or rate-QoE) relationship of each video linked by (d, t, q) combinations in Fig. 6.10. The rate can be easily derived after bit extraction of a certain (d, t, q) layer. Average MOS for an EV from subjects is derived after screening out outliers.

We derive the optimal scalability adaptation track for MOS ratings at different rates as highlighted in red solid lines in Fig. 6.10. If the scalability adaptation follows the track which is the upper bound, the resulting EVs can always provide the maximum QoE at a given adaptation rate.

We want to first point out some shortcomings in the existing work to support the necessity of deriving a useful scalability adaptation track. In [160, 161], subjective tests have been carried out on full length movies. The results in the short paper [160] reveal that users generally prefer to adapt by reducing frame rate first. In the long paper [161], the authors included more in-depth discussion and the experimental results reveal that for full-length movie, most of the subjects rated frame freeze (results of temporal scalability adaptation) as more annoying than blocks (similar but not totally corresponding to quality scalability adaptation). Nevertheless, these tests only try to figure out which scalability (between temporal and quality) is preferred in the general case. Hence the provided adaptation is coarser. Therefore, when fine-granular adaptation is possible, scalability adaptation may be achieved by alternately switching between different types of scalabilities. Another work in [162] investigates only spatial and temporal scalabilities. By quantifying the behavior of video QoE using two subjective video quality

6. QOE-AWARE SCALABLE VIDEO ADAPTATION AND STREAMING

Figure 6.9: Instances showing the multiple choices in scalability adaptation. Upper: video City, the left one is 3.75 fps, low quality, 810.74 kbps and 2.50 in MOS; the right one is 7.5 fps, low quality, 877.68 kbps and 3.31 in MOS. Lower: video Soccer, the left one is 7.5 fps, high quality, 2519.74 kbps and 3.08 in MOS; the right one is 30 fps, medium quality, 2467.43 kbps and 5.00 in MOS.

metrics, Structural Similarity Index Metric (SSIM) and Video Quality Metric (VQM), the results suggested that bandwidth savings have to be achieved by reducing spatial resolution first following by the frame rate. The shortcoming of these reference methods lies in the following aspects:

1. They are confined only to partial scalability;
2. The adaptation schemes are too coarse to support a fine-granular rate adaptation;
3. They are not optimal in terms of QoE.

To verify these arguments, we compare the reference schemes with the optimal track in Fig. 6.10. The Ref1-S indicates the preference on temporal scalability in [160] and Ref1-L

Figure 6.10: Rate-MOS relationships with various (d, t, q) combinations.

prefers the quality scalability adaptation first as in [161]. The Ref2 is from the reference [162]. We also have another reference JSVM basic extraction [170], which aims only at maintaining a good quality for the largest spatial resolution video. Since it is not achieved by the successive removal of enhancement layers, most of the adaptation points cannot be displayed in the point

set in Fig. 6.10. Nevertheless, the optimal track shows that it is inferior to adapting only at temporal and quality scalability directions therefore the scheme in [170] is not QoE-optimized.

6.4.3 The Common Scalability Adaptation Track for Grouped Videos

As we have shown in the previous subsection, the optimal scalability adaptation track can be found by tracing the upper bound of the rate-MOS space. The ANOVA analysis has revealed that the video ratings are content dependent. But the optimal scalability adaptation tracks summarized for the tested videos are only useful for the individual video and cannot be directly applied to other videos for their rate adaptation. We can anticipate that the ratings for a small group of videos could be considered to be similar when the content characteristics of these videos show a certain degree of similarities. In this case, we can summarize a common scalability adaptation track for the grouped videos and derive some adaptation rules for these videos. The common track is useful for rate adaptation if new videos have the similar content characteristics within the certain group.

Before we introduce the grouping strategy, we first elaborate how to derive a common scalability adaptation track for a group of videos. Within a predefined group, the optimal track for video i contains a set S_i which includes (d, t, q) combinations indicating the adaptation track. We propose a low complexity strategy to find the common track by searching only in a finite set. We define a temporary set U which searches from the intersection set $U_i = \cap S_i$ to the union set $U_u = \cup S_i$, and stepwisely includes one more element in the searching space. To follow the successive coding structure in SVC, we impose a restriction that no temporal and quality ID increase is allowed within a spatial layer with the augmented bitrate. Therefore we will only search possible common track within those U which does not violate this condition. We name this case *successive violation* for future reference. For each U , by substituting it as the common track for a video within the group, a MOS degradation will be calculated in comparison to its own optimal track according to [143]. The degradation is calculated as the mean deviation of MOS in the unit of rate; therefore it is also a measure of the mean absolute difference (MAD). If the inclusion of one more element into set U will achieve only negligible decrease (or increase) in the average MOS degradations, the searching is stopped and the elements in current set U are adopted to be the records of the common scalability adaptation track. As the total number of adaptation points is limited and the set difference of $U_u - U_i$ is small, the time spent on searching is negligible.

Figure 6.11: Video grouping in the SI-TI plane. (a) Two-set grouping. (b) Four-set grouping.

To find which videos can be potentially grouped for a common track, we adopt an exchanging-and-including strategy. Specifically, we first exchange the optimal track between each video pair. By substituting another video's optimal track, a MOS degradation will be calculated compared to its own optimal track as in [143]. If the degradations of a pair of videos are less than a threshold ϵ_1 , these videos are considered to be resembling each other in the adaptation track. Several subgroups can be found if videos that are pair-wise similar are classified into a group. Besides the videos that have been assigned into groups, there are also videos that share no similarity with others. We employ another including procedure to identify possible candidates and include them into the existing groups. We use the aforementioned strategy to find common tracks for all groups that include more than one candidate. If the MOS deviation of an isolated video, by substituting the common track of a grouped videos, is less than another threshold ϵ_2 , the video can be added into the group. Videos share similar adaptation track are classified after the two steps of grouping. The parameters ϵ_1 and ϵ_2 are decisive factors for how the grouped videos resemble each other and how many groups can be finalized. The accuracy of the classification differs for various applications. In some circumstance, the finer grouping is not necessary and sometimes the meticulous split leads to reduced computations. In this work, we show three types of groupings and compare their performance: the four-set grouping, two-set grouping, and one-set grouping.

Four-set grouping (G4): To impose a strict restriction in the grouping procedure, we can classify the tested videos into four groups. This grouping scheme gives a finer classification and therefore can achieve less deviation from the upper bound. By showing the pigeonholing

6. QOE-AWARE SCALABLE VIDEO ADAPTATION AND STREAMING

videos on the SI-TI plane in Fig. 6.11(b), it is noteworthy that videos are split distinguishably by their content characteristics. We refer the four separated groups as $G4-1$ (Soccer, meaning the first subgroup in the four-set grouping in the following context), $G4-2$ (Ducks), $G4-3$ (Crew, Harbour and Park), and $G4-4$ (City, Ice and Oldtown) in the following context. It is clear that videos characterized with high or low temporal/spatial content characteristics are split into different groups. Grouped videos can be bounded by square region. We employ a linear projection method to split the whole plane. The derived boundary resides in the middle of the maximized margin between the two sets thereby can lead to the best separation of subgroups. The cut-off rule derived from our database is $SI=12.1$ and $TI=25.5$.

Two-set grouping ($G2$): In certain applications, we may not want to impose so strict discrimination among videos. The videos can be coarsely distributed into two groups for ease of manipulations. By showing the pigeonholing videos on the SI-TI plane, it is noteworthy that videos are split distinguishably by their spatial and temporal contents. Specifically for the eight tested videos in our database, Videos City, Ice and Oldtown are classified into a group $G2-1$ and the remaining are grouped into another one $G2-2$. The classification clearly states that videos with lower SI and TI values resemble each other in the optimal scalability adaptation track, and the same holds for the videos with higher SI and TI. To formalize the video grouping, we split the groups in the SI-TI plane by an elliptical curve, as highlighted in Fig. 6.11(a). The elliptical curve resides at the best separated circle after the normalization by the maximum SI and TI values. The curve parameters are: centered ellipse with elliptical semi-axes as 40.1 and 16.6.

One-set grouping ($G1$): Except for the two and four set groupings, there are also applications which need only coarse adaptation without attention to specific video content characteristics. To simplify the problem, one may prefer a content-independent scalability adaptation scheme. All videos are classified into a single group and a general scalability adaptation track will be preferred for all the videos. We exemplify this case altogether and show its feasibility by its performance study in the following analysis.

For the grouped videos, we obtain a common track for different sources whilst guarantee that the deviation from the upper bound of each video source is minimal. The common track is favorable since we can achieve a general rule for SVC adaptation but not confined for an individual video. It is useful to have qualitative and quantitative evaluations on how the common track deviates from the optimal one. We show, by the cyan dashed lines in Fig. 6.10, the derived common tracks compared to the upper bound for eight videos when the tested videos

Table 6.4: MAD deviations of the Common Tracks

Video	City	Crew	Harbour	Soccer	Ducks	Ice	Oldtown	Park	Average
<i>G1</i>	0.020	0.046	0.006	0.076	0.081	0.009	0.118	0.068	0.053
<i>G2</i>	0.022	0.048	0.046	0.035	0.035	0.044	0.104	0.053	0.048
<i>G4</i>	0.022	0.041	0.002	0.022	0.022	0.044	0.104	0.063	0.040

are classified into four groups. The rate-MOS relationship derived from the common scalability adaptation track is shown in cyan dashed lines. The derived common tracks are shown to be close to the upper bound. We can further decrease the gap by employing finer grouping strategy. To validate the rationality of the grouping strategy and the effectiveness of the common tracks from another angle, we show in Table 6.4 the calculated MAD values. The average MOS degradation for all the eight video sources is only 0.048 for two-set grouping and is 0.040 for four-set grouping, which are quite negligible. We show the results of one group altogether in Table 6.4. It is noteworthy that the mean deviation 0.053 is still acceptable for certain applications. By employing the common adaptation track for videos with specific content characteristics, the designer can always acquire good revenue in terms of QoE for SVC video rate adaptation. This validates the practicability of employing a common track for scalability adaptation of different video sources.

The common track derived for grouped videos clearly demonstrates how the rate regulations can be conducted at different spatial, temporal and quality layers. The previous studies, [160], [161] and [162], can only provide coarser adaptation mechanisms and/or are only confined with partial scalability. It is clear in Fig. 6.10 that the reference methods [160], [161] and [162] perform differently from the optimal and common tracks under four-set grouping. Videos in any way prefer a content-related adaptation strategy rather than a fixed scheme to deal with all cases. We see, by an instance, the video Soccer with high motion, have a different track due to its special content nature. The second largest rate point (corresponding to (d, t, q) combination $(2, 3, 2)$), even serves as a valid adaptation point. However, because its rating is less than the forth valid point (correspond to $(2, 4, 1)$), it is excluded from the common track. This is just because high motion video prefers a high frame rate sequence to maintain its flowing experience. But in other group the same truth does not hold. Even for the one-set grouping scheme, our proposed common track can still achieve better QoE than the existing work. This is because the common track is always derived from the upper bound by a certain

Figure 6.12: Comparison of scalability adaptation tracks for one-set grouping $G1$ with other state-of-the-art algorithms. The (d, t, q) orders of the derived common track are: $(2, 4, 2)$, $(2, 3, 2)$, $(2, 3, 1)$, $(1, 3, 2)$, $(1, 3, 1)$, $(1, 2, 1)$, $(1, 2, 0)$, $(1, 1, 0)$, $(0, 2, 0)$, $(0, 1, 0)$ and $(0, 0, 0)$.

degree of compromise. The common track in any way is derived based on perceptual video experience and permits finer adaptation direction. To validate this point, we show the comparison in another way in Fig. 6.12.

In Fig. 6.12, we show the comparison between the adaptation trajectories of the common track derived by one-set grouping and the references. The references show that one kind of scalability is always preferred than the other one, which is not optimal. We can find from the common track derived at the every spatial layer that in Fig. 6.12, temporal downsampling and quality degradation are alternately executed for rate adaptation. Therefore, it is not straightforward to see that one kind of scalability is overwhelmingly superior to the other one for rate adaptation. The comparisons in Fig. 6.10 show that the references lead to worse MOS ratings. In addition, their adaptation step sizes are too coarse for applications and cannot achieve fine-granular rate tailoring.

We have derived common scalability adaptation tracks for different video groups. It is helpful to have some general guidelines to follow once a new SVC video is ready to be adapted. We accomplish this by studying the common tracks derived for different groups, which are summarized in Table 6.5.

We focus on the finest four-set grouping in which videos are generally characterized by

Table 6.5: Common scalability adaptation tracks by various groupings

Group	Common Track		
<i>G1-1</i>	(2,4,2), (2,3,2), (2,3,1)	(1,3,2), (1,3,1), (1,2,1), (1,2,0), (1,1,0)	(0,2,0), (0,1,0), (0,0,0)
<i>G2-1</i>	(2,4,2), (2,3,2), (2,3,1)	(1,3,2), (1,3,1), (1,2,1), (1,2,0)	(0,2,0), (0,1,0), (0,0,0)
<i>G2-2</i>	(2,4,2), (2,3,2), (2,3,1)	(1,3,2), (1,3,1), (1,2,1), (1,2,0), (1,1,0)	(0,2,1), (0,1,1), (0,1,0), (0,0,0)
<i>G4-1</i>	(2,4,2), (2,4,1), (2,3,1)	(1,3,2), (1,3,1), (1,2,1), (1,2,0)	(0,2,1), (0,2,0), (0,1,0), (0,0,0)
<i>G4-2</i>	(2,4,2), (2,3,2), (2,3,1), (2,2,1), (2,1,1)	(1,3,2), (1,3,1), (1,2,1), (1,2,0), (1,1,0)	(0,2,1), (0,2,0), (0,1,0), (0,0,0)
<i>G4-3</i>	(2,4,2), (2,3,2), (2,3,1)	(1,3,2), (1,3,1), (1,2,1), (1,1,1), (1,1,0)	(0,1,1), (0,1,0), (0,0,0)
<i>G4-4</i>	(2,4,2), (2,3,2), (2,3,1)	(1,3,2), (1,3,1), (1,2,1), (1,2,0)	(0,2,0), (0,1,0), (0,0,0)

their temporal and spatial contents. We distinguish them by low or high SI and TI information. Specifically, four groups *G4-1* to *G4-4* represent *LSHT* (low SI and high TI), *HSLT* (high SI and low TI), *HSHT* (high SI and high TI) and *LSLT* (low SI and low TI), respectively. It is clear that scalability adaptation in terms of optimized Quality of Experience differs greatly from some existing solutions where some arbitrary conclusions declare which kind of scalability is preferred over the others. The adaptation scheme under QoE criterion can critically overturn some inherent stereotypes. For example, a total of fifteen points exist in the 4CIF resolution. However, we see at most five points are employed in QoE-optimized adaptation. When rate decreases to a certain level, one may prefer to extract a low resolution video rather than seeking rate fulfillment by temporal downsampling or quality reduction. This explains the necessity of adapting scalable stream from QoE criterion. It should be noted that the common track for individual group is derived from an optimized criterion, by searching the minimum average QoE degradation in the candidate set; therefore we can to a certain extent relax the trajectory selection in scalability adaptation. By a careful study on the common track derived for different group videos, we can summarize the following observations for reference in scalability adaptation. Please check Table 6.1 for the coding configurations.

1. Low quality is not preferred in high resolution videos. For example, a larger QP is not suitable for 4CIF videos. Users otherwise prefer a spatially downsized video in this case. The case implies that an interpolated downsized video present better video experience than a coarse quantized video.
2. Except for the *HSLT* videos, a lower frame rate in the higher resolution video is not preferred. For the *HSLT* videos, it is still possible to obtain a satisfied experience at a low frame rate 7.5 fps. But in any cases, QP is expected to be smaller to achieve a good user experience.

3. On condition that the previous cases are met, for the *LSHT* videos, quality scalability degradation is preferred first for rate adaptation. For other videos, reducing temporal resolution is a better way to save bandwidth.
4. For medium level resolution such as the CIF resolution, video should be adapted first by quality degradation and then the temporal downsampling, when both the quality and frame rate are high. More specifically in our database, the lowest acceptable frame rate at this resolution is 15 fps for low SI videos and 7.5 fps for high SI videos, respectively. Beyond these constraints, users prefer a spatially downsized video.
5. At the lower frame rate region for QCIF videos, rate adaptation is generally realized first by quality degradation to the base layer then temporal reduction to the lowest frame rate.

6.4.4 Rate-QoE Model and Validation

The common scalability adaptation track derived in Section 6.4.3 illustrates how to adapt the scalable video and meanwhile guarantee the maximal achievable QoE at each rate point. However, the track is not applicable for direct adoption in studying resource allocation problem or others as the track itself cannot reflect the corresponding rate information. We intend to derive a QoE model to approximate the common scalability adaptation track and make the use of scalability adaptation tracks in resource allocation more mathematically tractable. Because MOS ratings derived in the scalability adaptation are a relative index, we aim at approximating the QoE degradation across the adaptation process. Mathematically, the QoE degradation can be approximated as

$$\Delta Q(R, \mathcal{R}) = \mathcal{G}(\mathcal{R}, SI, TI) - \mathcal{Q}(R, \mathcal{R}) \quad (6.5)$$

where R and $\mathcal{R}$ are respectively the adapted rate and the maximal rate when all SVC video packets are counted. The first item $\mathcal{G}(\mathcal{R}, SI, TI)$ is the maximal QoE rating when the maximal rate $\mathcal{R}$ is achieved. The rating will also differ from videos due to their spatial and temporal content natures. The study of the maximal achievable QoE is out scope of this research and in this dataset, the full bitstream always gets a rating which is close to the maximal MOS value. The second item $\mathcal{Q}(R, \mathcal{R})$ relates the adaptation rate R with the QoE rating by another function, and it is also video-dependent with the maximal rate $\mathcal{R}$. By investigating our dataset, the QoE degradation in the common scalability adaptation track can be approximated to a downward

logarithm function as follows

$$\Delta Q(R, \mathcal{R}) = \mathcal{G}(\mathcal{R}, SI, TI) - Q(R, \mathcal{R}) = -\ln(f(R, \mathcal{R})) \quad (6.6)$$

The graphical illustration of rate-MOS relations show a uniform QoE degradation to the normalized rate $R/\mathcal{R}$. Hence, the QoE modeling can be generalized as

$$Q(R, \mathcal{R}) - \mathcal{G}(\mathcal{R}, SI, TI) = \ln(f(R/\mathcal{R})) \quad (6.7)$$

where $f(\cdot)$ is a polynomial function; R and $\mathcal{R}$ are respectively the adapted rate and the maximal rate. We calculate the absolute approximation error and conclude that the second order polynomial serves as a good approximation of $f(\cdot)$. Because using a third order polynomial will only reduce the deviation by about 0.01 in MOS which is negligible. Hence, we propose to approximate the QoE at a given rate for a specific video as

$$Q(R, \mathcal{R}) = \ln\left(a_0 + a_1 R/\mathcal{R} + a_2 (R/\mathcal{R})^2\right) + Q_m \quad (6.8)$$

where the $Q_m = \mathcal{G}(\mathcal{R}, SI, TI)$ is the maximal achievable QoE for the non-truncated video. To comply to the real data, the polynomial function $f(\cdot)$ should be restricted to a range to make sure the calculated MOS lies in the effective range $[1, 5]$. To this, we restrict $f(\cdot) \in [e^{1-Q_m}, e^{5-Q_m}]$. The ceiling or flooring values will be selected once the polynomial falls outside the range. To validate the proposed rate-QoE model, we calculate the MAD between the approximated function and the common track based on our QoE assessment database. We approximate the coefficients in Eq. (6.8) with MOS ratings and normalized rate from all eight video sources. The approximation coefficients in our simulations are summarized in Table 6.6. The Rate-QoE model has small MAD deviations from the common scalability adaptation track for the four video sources which are reported in Table 6.7. These deviations are so small that they seldom influence the viewer's ratings. The average deviation is only around 0.2 which would seldom influence the user rating in a five-level score. Therefore we can draw the conclusion that the proposed model faithfully reflects the rate-QoE relationship for different SVC videos in scalability adaptation. Another evidence can be seen in the rate-QoE model approximation, which is shown in green dash-dot lines in Fig. 6.10.

The rate-QoE model provides a quantized index to analyze the scalability adaptation which makes resource allocation more mathematically tractable. In addition, the small deviation of the rate-QoE model promises its usage in estimating the QoE value under a given rate, which

Table 6.6: The approximation coefficients of the proposed Rate-QoE model

Groups	a_0	a_1	a_2
<i>G1-1</i>	0	2.02	-1.05
<i>G2-1</i>	-0.03	2.14	-1.14
<i>G2-2</i>	0.02	1.97	-1.02
<i>G4-1</i>	-0.01	2.68	-1.67
<i>G4-2</i>	0.03	2.40	-1.46
<i>G4-3</i>	0.02	1.75	-0.79
<i>G4-4</i>	-0.03	2.14	-1.14

Table 6.7: MAD deviations of the Rate-QoE model by various groupings

Video	City	Crew	Harbour	Soccer	Ducks	Ice	Oldtown	Park	Average
<i>G1</i>	0.179	0.333	0.137	0.140	0.104	0.240	0.392	0.140	0.209
<i>G2</i>	0.233	0.351	0.125	0.136	0.084	0.239	0.372	0.127	0.208
<i>G4</i>	0.223	0.282	0.102	0.088	0.064	0.239	0.372	0.173	0.193

will save a great deal of time without resorting to the time-consuming subjective quality assessment. By using the rate-QoE model, we propose in the next section a QoE-aware resource allocation scheme for multiple SVC video transmissions over the emerging multiuser MIMO-OFDM systems. Considering the small performance gap between four-set and one-set grouping, we will employ the model derived by one-set grouping in the following resource allocation.

6.5 QoE-aware Resource Allocation for SVC over Multiuser MIMO-OFDM Systems

We consider a single cell downlink scenario for multiple scalable video streaming over MIMO-OFDM systems under QoE criterion. The multiuser MIMO-OFDM system is the same as in Section 5.1.2. The downlink multiuser MIMO-OFDM system adopted at the transmitter side is depicted in Fig. 6.13. In the system, scalable videos are streamed to K receivers simultaneously from a base station (BS) which is equipped with M antennas. The BS determines how resources, e.g., bit rate and power, are to be allocated to each user.

We recall some discussion on parameter selection in Section 5.1.2 when considering the QoE maximization problem. In the multiuser MIMO-OFDM system, time, frequency, space,

Figure 6.13: The transmitter structure of SVC video transmission over multiuser MIMO-OFDM systems (based on Fig. 1, page 1735 of [66] and Fig. 1, page 109 of [67]).

and user diversities can be integrated together for intelligent resource allocation. Specifically, adaptive subcarrier assignment, beamforming vector adjustment, spatial multiplexing, power control and others can be optimized to satisfy QoS requirements. With the channel information, the beamforming weight vector can be configured such that users are orthogonal in the space domain as long as the user number is no more than the antenna number M . Time axis is divided into OFDM symbols where L symbols occupy one time slot, as shown in Fig. 5.3. As the coherent time of the wireless channel is much larger than that of an OFDM symbol, we let L OFDM symbols compose a system frame and resource allocation be carried on at the beginning of each system frame. On the other hand, subcarrier grouping can acquire higher multiuser diversity [66]. We follow the same configuration that b adjacent subcarriers are grouped together to form a subchannel based on frequency reuse as a minimum unit in resource allocation. Therefore, the resource allocation scheme is executed for each symbol-subchannel unit to maximize the overall system performance in terms of QoE by determining:

1. which users are admitted to transmit in a unit,
2. how much power should be distributed for transmission, and
3. what modulation scheme should be selected.

6.5.1 QoE Maximization for Multiple Scalable Video Streaming

The relevant parameters adopted for MIMO-OFDM system are the same as in Section 5.1.2. Let $c_{k,n,l}$ denote the allocation indicator where $c_{k,n,l} = 1$ means that the unit of the n th sub-channel and l th OFDM symbol is allocated to user k for transmission; otherwise, $c_{k,n,l} = 0$. Hence, the transmission during an OFDM symbol yields a total power consumption of $\sum_{k=1}^K \sum_{n=1}^N c_{k,n,l} p_{k,n,l}$.

We characterize the QoE of client k by a utility function $\mathcal{Q}_k$, i.e., Eq. (6.8) where the QoE is expressed as function of allocated rate and the maximal rate of an SVC video. The resources are allocated among users to maximize the summation of QoE for all users where the power resource P_T is limited in each symbol unit:

$$\max_{\substack{c_{k,n,l} \\ p_{k,n,l} \\ m_{k,n,l}}} \sum_{k=1}^K \mathcal{Q}_k \quad \text{s.t.} \quad \begin{cases} \sum_{k=1}^K \sum_{n=1}^N c_{k,n,l} p_{k,n,l} \leq P_T \\ 0 \leq \sum_{k=1}^K c_{k,n,l} \leq M \\ p_{k,n,l} \geq 0 \\ c_{k,n,l} \in \{0, 1\} \\ m_{k,n,l} \in \mathcal{M} \end{cases} \quad (6.9)$$

The related variables include: 1) $c_{k,n,l} \in \{0, 1\}$, deciding user assignment in a time-frequency unit, 2) $m_{k,n,l} \in \mathcal{M}$ where $\mathcal{M}$ is the set containing the number of bits for different modulation schemes, and 3) $p_{k,n,l} \geq 0$, power for user k in subchannel n and symbol l . The power and modulation assignments depend on the channel gain and the target BER, i.e., to guarantee a target BER under the certain channel condition, suitable modulation and enough power should be allocated. The solution lies in solving a combinatorial, nonlinear optimization problem by numerical analysis. The following theorem shows the NP-hardness of the problem.

Theorem 6.5.1. *The optimization problem (6.9) is NP-hard even for $L = 1$ and $M = 1$.*

Proof: We show that a special instance of the optimization problem (6.9) is an FDMA frequency allocation problem in [177] even if $L = 1$ and $M = 1$. The FDMA frequency allocation problem for two user cases has been proved to be NP-hard in [177] by reducing the known NP-complete equipartition problem. Only in this proof, we ignore the subscript l . If we simplify the utility definition as the rate acquired at current symbol, the optimization problem can be reduced to a rate maximization problem as

$$\max \sum_{k=1}^K \sum_{n=1}^N c_{k,n} m_{k,n} \quad \text{s.t.} \quad \sum_{k=1}^K \sum_{n=1}^N c_{k,n} p_{k,n} \leq P_T \quad (6.10)$$

$$e_{k,n,l} = \begin{cases} \infty & \text{if } m_{k,n,l} = \lceil \mathcal{M} \rceil \\ \infty & \text{if } c_{k,n,l} = 0 \text{ \& } \sum_{z=1}^K c_{z,n,l} = M \\ p(\mathcal{K}_{n,l}, m_{k,n,l}^+) - p(\mathcal{K}_{n,l}, m_{k,n,l}) & \text{if } c_{k,n,l} = 1, m_{k,n,l} \neq 0 \\ p(\mathcal{K}_{n,l}^+, m_{k,n,l}^+) + \sum_{k' \in \mathcal{K}_{n,l}} \left(p(\mathcal{K}_{n,l}^+, m_{k',n,l}) - p(\mathcal{K}_{n,l}, m_{k',n,l}) \right) & \text{if } c_{k,n,l} = 0 \text{ \& } \sum_{z=1}^K c_{z,n,l} < M \end{cases} \quad (6.11)$$

The problem (6.10) is known to be NP-hard in [177] therefore the original problem (6.9) with more general QoE function must be NP-hard.

The Theorem 6.5.1 suggests that finding a globally optimal solution for the resource allocation problem (6.9) is computationally intractable even if the number of users is small. The computational complexity will increase exponentially with the number of variables and it cannot be solved in polynomial time. In a practical wireless system, seeking the optimal solution to such a problem is costly. Generally, adaptive resource allocation schemes are proposed to solve such a problem [62, 66]. In this work, we propose a joint resource allocation scheme based on QoE modeling at the APP layer and radio resource allocation at the PHY layer. The joint efforts by combining QoE modeling, channel status and accumulative rate allocations will be proved to be effective solution to carry on the resource allocation in such a system with abundant diversities.

6.5.2 QoE-aware Resource Allocation

We solve the resource allocation problem by joint efforts of considering both the user channel status and its QoE model. The consideration of combined factors makes better use of multiple diversities and provides a feasible way to solve such an NP-complete problem. The core idea of the adaptive radio resource allocation scheme is summarized as follows: increase one order of modulation for a user in a subchannel where the target video QoE can be maximized.

We increase one order of modulation for a user in a subchannel where more increment of QoE could be achieved by one unit of the power consumption. To accomplish such a filling based bit loading scheme, we define a power cost function for each user in a unit. It is the required power for increasing one order of the modulation. The power cost $e_{k,n,l}$ for user k in unit (n, l) is expressed as in Eq. (6.11), which is similar to the expression in Eq. (5.7). We briefly introduce the idea here for the completeness of this chapter. A set $\mathcal{K}_{n,l}$ is constructed to record all users that have been admitted into unit (n, l) and have chosen the modulation order from $\mathcal{M}$. The power cost is calculated by cases: 1) If a user reaches the highest modulation

6. QOE-AWARE SCALABLE VIDEO ADAPTATION AND STREAMING

order in a unit or it cannot be admitted because the number of admitted users reaches M , the power cost is set to infinity. 2) If a user has previously selected a modulation order from the set $\mathcal{M}$, the power cost is the difference in power consumption between a higher order modulation and the current one. 3) If the number of admitted users is less than M and the user has not been selected before, the inclusion a new user requires the recalculation of beamforming vectors. All cost is included in calculating $e_{k,n,l}$. At the beginning of each system frame, every selected user is initialized by an initial cost. This is derived on the assumption that only a single user is admitted to the unit (n, l) to transmit and the user is entitled to the first order of modulation. The initial costs are only calculated at the first symbol and used across the current system frame.

The solution to the proposed problem (6.9) summarized in **Algorithm 4** is based on a joint power allocation and bit loading strategy. First the power allocation $p_{k,n,l}$ and modulation order $m_{k,n,l}$ are set to 0. After every increment of the modulation order, the needed power $p_{k,n,l}$ to guarantee the target BER is calculated by referring to (5.4) and (5.5).

The bit loading scheme runs in an iterative procedure for symbol l until power resource is used up. Then it proceeds to the next symbol to find power and modulation vectors. At a certain symbol l^* , we examine all users admitted into subchannels to select the user in a subchannel which could achieve the most increment of utility by one unit of power. The selected user-unit pair increases its modulation order by one. The QoE increment by increasing one order of modulation is given by

$$\Delta Q_k = \ln \left(a_0 + a_1 \frac{R_k + b}{\mathcal{R}_k} + a_2 \left(\frac{R_k + b}{\mathcal{R}_k} \right)^2 \right) - \ln \left(a_0 + a_1 \frac{R_k}{\mathcal{R}_k} + a_2 \left(\frac{R_k}{\mathcal{R}_k} \right)^2 \right) \quad (6.12)$$

where b is determined by the difference of modulation schemes and the number of subcarriers in a subchannel. For all active users, the user-unit pair to increase one order of modulation is selected by

$$\{k^*, n^*\} = \arg \max_{\{k,n|l^*\}} \frac{\Delta Q_k}{e_{k,n,l^*}} \quad (6.13)$$

The accumulated rate R_k is updated after each selection and this procedure is carried out iteratively until the power in current symbol l^* is used up then the bit loading proceeds to the next symbol. At the end of a system frame, the rate R_k for user k is updated and is recorded for reference in the next system frame. As the user selected in each step leads to the maximal QoE increment, the bit loading by filling strategy improves the summation of QoE.

Algorithm 4: Description of the QoE-aware Resource Allocation Framework

1 Initial Power Cost Calculation:

2 Calculate initial cost e_{k,n,l^*} when $l^* = 0$, keep the initial cost $e_{k,n,l^*} = e_{k,n,0}$ for all $1 < l^* < L$;

3 Power Allocation and Bit Loading:

4 **for** $l^* = 0; l^* < L; l^*++$ **do**

5 $m_{k,n,l^*} = 0, p_{k,n,l^*} = 0$;

6 **while** *True* **do**

7 Calculate the QoE increment ΔQ_k by Eq. (6.12);

8 Select the user-subchannel pair $\{k^*, n^*\}$ by Eq. (6.13);

9 Update the modulation scheme to the next level;

10 Update the Beamforming vector if k^* is newly added;

11 Update the power to satisfy a target bit error rate by referring to Eqs. (5.4) and (5.5);

12 If power is used up, $\sum_{k=1}^K \sum_{n=1}^N c_{k,n,l^*} p_{k,n,l^*} > P_T$, **break**;

13 Update the rate for user k^* by $R_{k^*} = R_{k^*} + b$;

14 Update the power cost e_{k,n,l^*} for every user-unit pair (k, n) ;

We analyze the performance of the proposed solution with reference to some existing solutions which address a simple version of our problem. The bit loading solution can optimally solve the rate maximization problem in single user systems [63] and the extension to multiuser system is proposed in [64]. The solutions cannot be directly used due to the focus on QoE maximization and cochannel users introduced in the multi-antenna system. Nevertheless, the optimality analysis in [64] verifies that our QoE-aware bit loading solution can achieve approachable performance to the optimal solution. When there is low interference among users, the multiuser bit loading algorithm is the optimal. We satisfy this case by restricting the user number in a unit no more than the transmitter antenna number. With limited users in a unit, the spatial precoding can achieve ideal interference suppression. The utility to cost measure stepwisely approaches to the optimal solution by counting in each filling decision. The proposed solution jointly explores both the user diversity (higher incremental video QoE) and the network diversity (lower incremental power costs). This promises that the proposed solution will only deviate slightly from the optimal solution but highly reduce the computational complexity.

We examine the computational complexity of the proposed scheme before concluding this

section. The complexity arises from two aspects: the cost estimation and the utility increment calculation before each filling step. We analyze the complexity of cost evaluation first. When a user has been previously added, the calculation of power cost for the user in the current unit needs only an operation of multiplication. Only when a user is newly added and the modulation is increased from zero to the first order, including the new user needs recalculation of beamforming vector. In any cases, the calculation of power cost needs only a handful of linear operations so the complexity is quite limited. On the other hand, the utility increment calculations are upper bounded by $\mathcal{O}(LN(M + 1))$ times within a system frame. To sum up, the computation involved in the proposed solution includes only a few linear operations, which is generally acceptable in most applications. Furthermore, our solution can always maintain a good system performance in terms of video QoE, as will be demonstrated in Section 5.3.

6.6 Simulation Results

6.6.1 Simulation Setting

We have conducted evaluation on the proposed QoE-aware scalability adaptation scheme by comparing it with references [160], [161] and [162] in Section 6.2. We have also discussed the rationality of the rate-QoE model in Section 6.5. In this part, we show the experimental results of the QoE-aware resource allocation scheme for SVC video transmission over multiuser MIMO-OFDM systems. A multiuser MIMO-OFDM downlink simulation platform has been developed where a base station transmitted multiple SVC videos simultaneously to multiple receivers. The platform is developed based on a C programming language complied MIMO-OFDM system [66] and the key parameters of the system were set according to the IEEE 802.16 standard [3] by referring to [156]. The receivers located and moved randomly in a cell so the channel status varies timely. The key parameters adopted in this simulation are summarized in Table 6.8. The eight videos, compressed by SVC using the same configuration as elaborated in Section 6.2, were simultaneously transmitted from the base station to receivers. Due to limited radio resources to accommodate all receivers with the original SVC bitstreams, some SVC video packets that cannot be transmitted timely are discarded due to the exceeding the delay tolerance. We engaged another group of 24 viewers to vote for the QoE of received videos. After a training session to let the viewers get familiar with the truncated and distorted videos, a total of 64 videos were shown to each viewer for voting. The QoE evaluation environment and method were the same as that in Section 6.3.

Table 6.8: List of System Parameters

Parameters	Values
Cell size	1.6 km
Modulation	QPSK, 16-QAM, 64-QAM
System bandwidth	5 MHz
Antenna number (M)	2
FFT size	512
Frame duration	2 ms
OFDM symbol duration	100.8 μsec
Subcarrier spacing	11.16 KHz
Subcarrier number	384
Subchannel number (N)	8
Subcarriers per Subchannel (b)	48
Symbols per frame (L)	8
Target bitrate BER_{target}	10^{-4}
Delay tolerance	100 MIMO-OFDM system frames

6.6.2 Evaluations and Discussions

To show the superiority of the QoE-aware adaptation in transmissions, we compare our proposed scheme to the basic bit stream extraction scheme in JSVM [170]. It should be noted that the bit extraction scheme based on Quality Layer information [138] only has a confined effective range as it cannot adapt the rate lower than the minimal rate for the highest resolution. Therefore it is not suitable for comparison in our evaluation. In the JSVM basic extraction, the packets are prioritized in a fixed pattern and bit rate adaptation is achieved by discarding lower priority packets first to conform to the rate constraint. The packet level adaptation is fine-granular and has the full-range adaptation. However, the JSVM adaptation only aims at retaining a good PSNR for the largest spatial layer video, thus is not QoE optimal.

To validate the performance of the proposed QoE-aware resource allocation (QoERA) scheme, we make comparison by invoking conventional resource allocation schemes. Three classical resource allocation schemes for multiantenna system are employed as reference. An adaptive radio resource allocation (ARRA) algorithm [66] for multimedia traffic transmission is included. The algorithm distributes resources base on a time-to-expiration index to provide high priority for stringent users but not take QoE into consideration. The well-known round-robin (RR) fashion of resource allocation is also adopted for comparison. The RR algorithm

6. QOE-AWARE SCALABLE VIDEO ADAPTATION AND STREAMING

Table 6.9: Average bit rate (Mbps) by different resource allocation algorithms

Video	City	Crew	Harbour	Soccer	Ducks	Ice	Oldtown	Park	Average
ARRA	0.59	0.06	0.05	0.67	3.47	0.01	0.34	0.18	0.67
RR	0.47	0.22	0.21	0.38	0.26	0.25	0.50	0.23	0.31
MMSR	5.31	0.19	0.15	2.16	0.50	0.38	2.34	0.22	1.41
QoERA	4.18	0.46	0.39	1.17	0.85	0.65	2.19	0.50	1.29

alternatively distributes transmission opportunity to users without any consideration on channel variation and video QoE differentiation. A multicarrier maximum sum rate (MMSR) scheme loads power and bits in order to maximize the system throughput [62]. However, the throughput maximization scheme does not coincide with QoE maximization so it cannot guarantee the optimal QoE for users. In contrast, the proposed QoERA scheme jointly considers the packet contribution on QoE and channel differences among users. It performs substantial allocation in each filling step and improves QoE for users, as will be justified below.

We first show the average bitrate by employing different resource allocation schemes in Table 6.9. We execute 30 runs of simulation. Because at each run users are placed and moved randomly, the resource allocation among users varies due to the channel variations. It is impossible to present the inefficiency by unbalanced reference schemes if we put the average rate for each user. Instead, we present one of the 30 results which has the least deviation from the average rate for four algorithms to make comparisons. It is clear that the RR scheme achieves the worst performance since it allocates the radio resource to the users without considering the users' network conditions or QoE factors. Regardless of network conditions, a user has to wait for another round of selection to transmit its packets in an arbitrary subchannel and OFDM symbol selected by the round-robin method. Hence, the allocated rates are close among users whereas the overall system performance is severely degraded. The other two references assign a unit to a user based on the channel status, so that the throughput is improved when compared to the RR method. However, these two schemes will lead to unbalanced allocation for overall QoE as they neglect the rate-QoE relationships, the corresponding QoE differences and rate requirements between different video users. Therefore, the system performance in terms of QoE is harmed.

We show the MOS ratings of received videos obtained by both the JSVM basic bit-extraction and our QoE-aware adaptation scheme in Table 6.10. An algorithm corresponds to two lines of results where the upper one is from the JSVM based extraction scheme and the lower one is

Table 6.10: Average MOS ratings by different rate adaptation and resource allocation algorithms.

Video		City	Crew	Harbour	Soccer	Ducks	Ice	Oldtown	Park	Average
ARRA	JSVM	2.29	1.63	1.46	2.46	4.54	1.00	1.46	1.00	1.98
	QoE	3.36	1.46	1.50	3.92	4.00	1.71	2.58	1.71	2.53
RR	JSVM	1.67	2.04	2.42	1.00	1.00	1.00	2.00	2.25	1.67
	QoE	3.38	2.50	2.04	2.50	2.46	2.58	2.58	1.71	2.47
MMSR	JSVM	5.00	2.29	2.00	4.88	1.00	2.29	5.00	2.00	3.06
	QoE	4.88	2.17	2.04	3.92	2.08	3.67	5.00	1.71	3.18
QoERA	JSVM	4.88	1.87	2.63	4.58	1.00	3.29	4.75	2.42	3.18
	QoE	4.54	3.17	2.83	3.92	2.71	4.29	4.71	2.38	3.57

from our QoE-aware adaptation. As discussed, the JSVM basic extraction scheme only prioritizes packets in a fixed pattern. The ultimate goal for the extraction is to maintain a good PSNR for the largest resolution video (4CIF in our case). When the adapted rate decreases to a lower level, the quality degradation for the 4CIF video is large and in this situation, a better way is to discard some spatial enhancement layers instead of quality enhancement layers, as observed in our QoE assessment database. The severely distorted video at the highest spatial resolution can sometimes hardly present any meaningful information. On the other hand, users may prefer a lower resolution video with better fidelity from the QoE point of view. This explains why in our scalability adaptation track, spatial downsizing is preferred at a certain lower frame rate and quality. The outperformance of the QoE-aware adaptation scheme over JSVM can be envisaged by comparing these two adaptation algorithms applied in the same resource allocation scheme in Table 6.10. We can see that the QoE-aware adaptation outperforms conventional methods because it fully explores the QoE preference on SVC multidimensional adaptation. It achieves an average gain of 0.47 in MOS when compared to the JSVM basic scheme.

It is clear that both the QoE-aware adaptation and resource allocation schemes achieve notable improvements. On the other hand, the proposed QoE-aware resource allocation scheme outperforms MMSR, which is the best among other references, as high as 0.4 in MOS. This is a significant improvement which comes from the efficient QoE-aware resource distribution, and the exploitation of time, frequency and space diversities of multiuser MIMO-OFDM systems. To show the performance comparisons from another angle, we highlight a group of sample pictures from the received videos in Fig. 6.14. We show the rightmost picture from full rate sequence without adaptation for comparison. It is clear that the picture from the proposed scheme shows outstanding visual performance. It is worth noting that, the frame rate of some adapted

Figure 6.14: Sample pictures from video Crew. The upper row is from the JSVM based extraction and the bottom row is from the QoE-aware adaptation. From left to the right: ARRA, RR, MMS, QoERA and the full rate.

videos has been reduced by temporal scalability adaptation in QoE based adaptation or frame loss in the JSVM based adaptation. Although it cannot be presented from the static sample pictures, the frame loss or frame rate reduction also leads to detrimental visual experience.

To further verify the rationality of the rate-QoE model, we compare in Table 6.11 the average scores from the new subjective tests and those calculated directly from the rate-QoE model in Eq. (6.8). The approximation coefficients used in the model are: $a_0 = 0$, $a_1 = 2.02$, and $a_2 = -1.05$. The first column is derived from the rate-QoE model and the second column is from users' subjective ratings in this round of evaluations. The last column summarizes the deviation between the two schemes. The ratings from the new subjective assessments are close to those derived by the rate-QoE model and the mean absolute difference is only 0.14. It verifies that the approximated function in Eq. (6.8) is robust in modeling the rate-MOS relationship. The consistency of the results between the rate-QoE model and those from the new subjective assessment on the streaming videos promises that in practical systems, we can employ the rate-QoE model to predict the QoE accurately for the proposed common scalability adaptation track in SVC adaptation. This will save a great deal of time by compared to conducting subjective assessments.

6.7 Concluding Remarks

In this chapter, we present a QoE-aware resource allocation solution for SVC video streaming over multiuser MIMO-OFDM systems. We accomplish it by constructing a QoE database for SVC video rate adaptation and investigate various scalability adaptation tracks. Based on

Table 6.11: Validation of the rate-QoE model

Schemes	Model	Subjective Rating	Deviation
ARRA	2.16	2.53	0.37
RR	2.59	2.47	0.12
MMSR	3.18	3.18	0.00
QoERA	3.63	3.57	0.06
Average	–	–	0.14

the analysis and evaluation of the scalability adaptation, a rate-QoE model is developed accordingly and applied to the QoE-aware resource allocation in the video streaming systems. By exploiting the high throughput and abundant diversities of the multiuser MIMO-OFDM system, the time-frequency unit assignment, power allocation, and modulation selection are jointly optimized to maximize users' QoE. The superior performance of the proposed solution is demonstrated by comparisons with conventional rate adaptation schemes and resource allocation strategies.

Conclusions and Future Work

Scalable video coding, first appeared in video coding standards at the early of 1990s, has come a long way. It has been widely supported by various standards for its diversity and adaptability. The in-depth investigations undergone in the past decades on both the encoding and its applications put forward its maturity and consolidate its status. Only when it comes to the year of 2007, the finalization of the scalable extension of the H.264/AVC, which is the dominant video coding standard in nearly all common applications, summarizes the key components and concepts in the latest development. The standardization of Scalable Video Coding (SVC) inspires a great deal of research work on the codec analysis and its application in communications. Among one of them, this thesis investigates on some unexploited areas and with special attention paid to the wireless streaming services. With the aim to enhance the system performance for scalable video streaming, the efforts have been geared towards investigating both the video coding standards and wireless network protocols. Resource allocation and scheduling are key issues addressed by this thesis. This chapter serves as a summary of the work contributed by the thesis and explores some unsolved problems for future study.

7.1 Main Contributions and Results

In this thesis, the fundamental concept and underlying challenges in wireless video streaming was first introduced and explained in Chapter 1. These potential challenges make the selection of scalable video coding a ready choice in wireless streaming. The reviews on both SVC coding standard and two widely deployed networks were provided in Chapter 2. Besides, related technologies in wireless video streaming and a great deal of existing work were also

7. CONCLUSIONS AND FUTURE WORK

reviewed and discussed in this chapter. Specifically, this chapter surveyed the existing work with a focus on the resource allocation and scheduling issues. The work on other video coding standards and wireless networks has also been reviewed, and the finding has been summarized for setting up scalable video streaming systems in later chapters.

To exploit the benefits of scalable video, we proceeded to investigate and evaluate the resource allocation and scheduling issues in several typical wireless networks. In Chapter 3, we considered the packet scheduling issue for the SVC streaming over WLANs. Specifically, we selected the multimedia-favored IEEE 802.11e standard as the network structure. To take advantage of the differentiated scalable video packets and the prioritized packets protection in the network's MAC layer, we considered an effective packet mapping scheme to utilize constrained resources. The packet departure rate was fed back by the link layer. With an efficient packet prioritization and scheduling strategy, the design achieved enhanced system performance in terms of video quality. Packets are differentially protected and the network congestion was reduced to the minimum level by the proposed scheme. The collaborative efforts of application, MAC, and link layers contributed to an efficient and effective solution for scalable video streaming over the IEEE 802.11e networks, as verified by drawing comparison with existing solutions. This part of work is based on our published papers [108, 109, 178, 179, 180].

Chapter 4 proposed an enhanced packet prioritization scheme for fast rate adaptation. It is well known that the rate-distortion analysis is vital and useful for any video coding standard. When it comes to the scalable video coding, packets are encoded with differentiated priority in the realization of layer coding. The enhanced packet prioritization was achieved by an in-depth investigation on the packet prediction relationship in SVC coding. More importantly, the investigation on prediction relationship refrained from extensive calculations and hence provided an easy approach to achieving rate adaptation. The analysis covered various scalability combinations and therefore, the proposal was applicable to different coding configurations. By employing the fast rate adaptation scheme, the resource coordinator of wireless networks could design efficient resource allocation and scheduling schemes by considering both the network specifications and video characteristics. The next chapter served as an example to take on this advantage. This part of work has been published in our paper [141].

Chapter 5 demonstrated the advantage of streaming scalable video in different wireless networks. Beside the widely deployed WLANs, WMANs have been well studied in both academia and industry. The evolvments of both the WLANs and WMANs have been rooted from an

advanced physical layer technology, the MIMO-OFDM technique. Therefore, the consideration of scalable video streaming over MIMO-OFDM systems is important for both research investigation and practical deployment. Chapter 5 investigated the effective resource allocation schemes in the MIMO-OFDM systems to enhance the overall video utility, which has been defined based on the fast rate adaptation scheme in Chapter 4. Specially, we presented in Chapter 5 a scalable resource allocation framework to address the controversial issues for fairness and efficiency exploitations in resource allocation problems. To utilize the inherent advantage provided by the scalable video, fairness and efficiency schemes were designed respectively for the base and enhancement layers. This design severed as a tradeoff for the overall system performance and user fairness. To provide every admitted user a fair opportunity to obtain a basic viewing experience, the framework first distributed resources via a MAXMIN fair scheme. Subsequently, resources were allocated by maximizing the overall system utility. The scalable framework was effective as it provided a fair opportunity for heterogeneous video terminals to obtain a basic view of video programs, and network efficiency could be exploited if there are surplus resources. The novel idea of integrating a scalable resource allocation framework and making use of the inherent superiority of scalable video is the main contribution of this chapter, which is based on our published papers [181, 182, 183, 184, 185].

All previous chapters focused on the conventional wireless video streaming issues which are devoted to providing services under some objective video quality metrics. The adoption of Quality of Experience (QoE) has extended the research on video communication into a new realm. The adaptation and streaming of scalable video based on a QoE criterion are still new topics that have not been well examined. Chapter 6 served as one of the pioneering studies to investigate on this topic. The main results were summarized based on our papers [168, 169, 186, 187]. Chapter 4 introduced an optimized scalable video adaptation scheme with the aim to maintain the best objective quality, i.e., PSNR, for video of the largest spatial size in an SVC bitstream. The proposed scheme is feasible for fast rate adaptation but is QoE uncertain. This chapter made improvement to adapt scalable videos at all scalability directions. Specifically, a QoE evaluation database was constructed for SVC video adaptation. The database, based on normative specifications, led to a list of new findings in the scalable video adaptation. It helped to derive an optimal video adaptation track for each individual video and summarize common adaptation tracks for group videos with similar content characteristics. Based on the analysis and evaluation of the scalability adaptation, a QoE model has been developed accordingly and applied to the QoE-aware resource allocation in video streaming system. By

7. CONCLUSIONS AND FUTURE WORK

exploiting the high throughput and abundant diversities of the multiuser MIMO-OFDM system, the time-frequency unit assignment, power allocation, and modulation selection were jointly optimized to maximize users' QoE. The superior performance of the proposed solution has been demonstrated by comparing it with conventional rate adaptation and resource allocation schemes.

In exploring the above problems discussed in each chapter of this thesis, the proposed schemes have been extensively and fairly compared with a multitude of conventional strategies.

7.2 Future Work

We have investigated the SVC delivery over WLANs and WMANs. The problem formulations and proposed solutions can be applied to other types of wireless networks. However, because applications vary from one to another, there are still other traits that should be exploited from SVC such that various requirements and scalabilities could be supported. On the other hand, some wireless mediums exhibit distinctive characteristics and are designed for specific service goals, so the algorithms designed should also be extended accordingly. While encouraging results have been attained, there are several interesting issues that need to be addressed in the future:

1. For rate adaptation of SVC, we have addressed problems for a single user, as well as multiusers, with similar video specifications or coding structures. It is interesting to extend them to more general cases where clients diverge significantly either in hardware specifications or user requirements. We would like to design more general and efficient solutions to meet any possible combinations of scalabilities so as to accommodate heterogeneous clients ranging from high definition TV receivers to handheld users. The packet analysis should be efficiently designed as multiple types of information can be utilized to increase the accuracy and also reduce computational complexity. Possible reference information includes layer IDs, motion vectors, rate control options, quantization parameters, and so on.
2. As the next-generation wireless systems are expected to provide services for heterogeneous traffics. The major challenges we are confronted with including very diverse QoS requirements, harsh wireless channels, scarce radio resources, and stringent energy constraints. This motivates the development of smart resource allocation techniques to

provide QoS differentiation, user fairness, and efficient power and spectrum utilization over the wireless media. The Multiple Input Multiple Output (MIMO) and Orthogonal Frequency Division Multiplexing (OFDM) considered in this thesis, focuses only on the application and physical layers' resource management. It would be advantageous to combine cross-layer information for a high level of optimization in order for much better video performance.

3. Conventionally, the algorithms for wireless video transmission lie in a formulation of the optimization problem and the proposal of a solution to it. As SVC exhibits distinctive characteristics compared with other coding standards, we can expect different formulations and solutions based on new mathematical tools. On the other hand, as wireless multimedia applications become diversified and personalized, the formulation of the problem from other subjects could be helpful in understanding resource allocation and packet scheduling mechanisms. Some examples are game theory for fairness, operational theory for competition, mechanism design for system construction, and so on. Under all these possible extensions, our goal is to design a more generic scheme for scalable video delivery over wireless networks and meanwhile diverse QoS requirements could be satisfied.
4. The increasing focus on QoE for video, and especially for wireless video communication, attracts more attention from both the academy and the industry. The measurement of QoE can be carried out from either the frame pixel domain or the network side. Chapter 6 summarizes some useful guidelines for SVC video adaptation. However, the spatial size is only confined in the QCIF, CIF and 4CIF, which mainly belong to the handheld specification. A necessary extension is the coverage of various display sizes. The most popular frame sizes that are used in the mobile internet and home media services include 1080p, 720p, WVGA, VGA, etc. The inherent differences in the observation distance, viewing habits and content sharing ways lead to distinctive QoE interpretations. The scalable video adaptation on these sizes has not been well exploited. Besides, it is also important to exploit this issue on TV sizes. The arising of new wireless video markets also calls for diverse QoE requirements. Possible applications include home high definition television (HDTV), online video games, video sharing over social networks, and so on. These emerging applications inherently adopt different wireless network structures. The needs of video consumption also diverge significantly. The provision of QoE

7. CONCLUSIONS AND FUTURE WORK

for these kinds of services should not be defined by a single guideline or using a fixed resource allocation or packet scheduling strategy. We are always expecting that more efforts would be used to exploit the QoE issues for scalable video streaming over wireless networks. An important direction is to incorporate the analytical rate-quality modeling in [163, 166, 171, 172, 173, 174] and our intelligent resource allocation and scheduling schemes, to design the most comprehensive video streaming framework in terms of the best subjective quality.

Author's Publications

Book Chapter

- M. Li, S.-P. Chuah, Z. Chen, and Y.-P. Tan, “Scalable video delivery over Wireless LANs,” in Streaming Media Architectures, Techniques, and Applications: Recent Advances, C. Zhu, Y. Li, and X. Niu eds, *IGI Press*, 2010.

Journal Papers

- M. Li, Z. Chen, and Y.-P. Tan, “QoE-aware resource allocation for scalable video streaming over multiuser MIMO-OFDM systems,” *under review in IEEE Trans. Multimedia*
- M. Li, Z. Chen, and Y.-P. Tan, “On quality of experience of scalable video adaptation,” *J. Visual Commun. Image Represent.*, vol. 24, pp. 509-521, 2013
- M. Li, Z. Chen, and Y.-P. Tan, “A scalable framework for SVC video streaming over multiuser MIMO-OFDM networks,” *accepted by IEEE Trans. Multimedia*
- M. Li, Z. Chen, and Y.-P. Tan, “Cross-layer optimization for SVC video delivery over the IEEE 802.11e wireless networks,” *J. Visual Commun. Image Represent.*, vol. 22, pp. 284-296, 2011.

- Z. Chen, M. Li, and Y.-P. Tan, "Perception-aware multiple scalable video streaming over WLANs," *IEEE Signal Process. Letters*, vol. 17, pp. 675-678, July 2010.

Conference Papers

- M. Li, Z. Chen, S.-P. Chuah, and Y.-P. Tan, "A fast rate adaptation scheme for SVC based on the packet dependencies," *IEEE Int. Symp. Circuits and Syst.*, May 2013. **(Oral)**
- M. Li, Z. Chen, and Y.-P. Tan, "QoE analysis for scalable video adaptation," *IEEE Visual Commun. Image Process.*, Nov, 2012. **(Oral)**
- M. Li, Z. Chen, and Y.-P. Tan, "QoE-aware Resource Allocation for Scalable Video Transmission over Multiuser MIMO-OFDM Systems," *IEEE Visual Commun. Image Process.*, Nov, 2012. **(Oral)**
- S.-P. Chuah, Z. Chen, M. Li, and Y.-P. Tan, "Energy-minimized wireless video transmissions via sharing of retransmission limits," *IEEE Int. Conf. Image Process.*, Sep. 2012. **(Oral)**
- M. Li, Z. Chen, and Y.-P. Tan, "A scalable resource allocation framework for SVC video transmissions over downlink MIMO-OFDM networks," *IEEE Int. Symp. Circuits and Syst.*, May 2012. **(Oral)**
- M. Li, Z. Chen, and Y.-P. Tan, "Joint power allocation and bit loading for enhanced SVC video downlink transmissions over SDMA/OFDMA networks," *IEEE Conf. Multimedia and Expo*, pp. 1-6, July 2011.
- N. Liu, J. Lu, Y.-P. Tan, and M. Li, "Set-to-set gait recognition across varying views and walking conditions," *IEEE Conf. Multimedia and Expo*, pp. 1-6, July 2011. **(Oral)**
- M. Li, Z. Chen, and Y.-P. Tan, "A MAXMIN resource allocation approach for

scalable video delivery over multiuser MIMO-OFDM systems,” *IEEE Int. Symp. Circuits and Syst.*, pp. 2645-2648, May 2011. **(Oral)**

- M. Li, Z. Chen, and Y.-P. Tan, “Scalable video transmission over multiuser MIMO-OFDM systems,” *5th Int. Conf. Commun. and Networking in China*, pp. 1-8, Aug. 2010. **(Invited Paper)**

- M. Li, Z. Chen, S.-P. Chuah, and Y.-P. Tan, “Efficient packet scheduling for scalable video delivery to mobile clients,” *IEEE Int. Symp. Circuits and Syst.*, pp. 2251-2254, May 2010. **(Oral)**

- M. Li, Z. Chen, and Y.-P. Tan, “Joint packet prioritization and QoS mapping for SVC over WLANs,” *IEEE Int. Conf. Acoustics, Speech, and Signal Process.* 2298-2301, Mar. 2010. **(Oral)**

Bibliography

- [1] Institute of Electrical and Electronics Engineers, “IEEE 802.15 Working Group for wireless personal area networks at <http://www.ieee802.org/15/>,”
- [2] Institute of Electrical and Electronics Engineers, “IEEE Standard 802.11-1999, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications,” IEEE Standard 802.11-1999.
- [3] Institute of Electrical and Electronics Engineers, “Local and metropolitan area networks-part 16: air interface for fixed broadband wireless access systems,” IEEE Standard 802.16-2004.
- [4] “Coding of audio-visual objects - part 2: Visual,” ISO/IEC 14492-2 (MPEG-2 Visual), 1994.
- [5] “Coding of audio-visual objects - part 2: Visual,” ISO/IEC 14496-2 (MPEG-4 Visual), 1999.
- [6] ITU-T, “Advanced video coding for generic audiovisual services,” ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), 2003.
- [7] T. Stockhammer, M. Hannuksela, and T. Wiegand, “H.264/AVC in wireless environments,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 13, pp. 657–673, July 2003.
- [8] J. G. Apostolopoulos, W.-T. Tan, and S. J. Wee, “Video streaming: Concepts, algorithms, and systems,” tech. rep., HP Laboratories, 2002.

BIBLIOGRAPHY

- [9] M. van der Schaar and S. Shankar N, “Cross-layer wireless multimedia transmission: challenges, principles, and new paradigms,” *IEEE Wireless Commun.*, vol. 12, pp. 50–58, Aug. 2005.
- [10] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable video coding extension of the H.264/AVC standard,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 17, pp. 1103–1120, Sep. 2007.
- [11] Available at: <http://www.streaminglearningcenter.com/articles/h264-scalable-video-coding—what-you-need-to-know.html>.
- [12] H. Schwarz and M. Wien, “The scalable video coding extension of the H.264/AVC standard [standards in a nutshell],” *IEEE Signal Process. Magazine*, vol. 25, pp. 135–141, Mar. 2008.
- [13] T. Schierl, T. Stockhammer, and T. Wiegand, “Mobile video transmission using scalable video coding,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 17, pp. 1204–1217, Sep. 2007.
- [14] ITU-T, “Video coding for low bit rate communication,” ITU-T Rec. H.263.
- [15] J.-R. Ohm, “Advances in scalable video coding,” *Proc. IEEE*, vol. 93, pp. 42–56, Jan. 2005.
- [16] H. Schwarz, D. Marpe, and T. Wiegand, “Analysis of hierarchical B pictures and MCTF,” pp. 1929–1932, July 2006.
- [17] J. Reichel, H. Schwarz, and M. Wien, “Technical description of the hhi proposal for svc ce1,” (ISO/IEC JTC 1/SC 29/WG 11, Doc. M11244), Oct. 2004.
- [18] M. Ghanbari, *Standard Codecs: Image Compression to Advanced Video Coding*. IET Press, 2011.
- [19] T. Wiegand, *Scalable video coding in H.264/AVC*. Available at: http://iphome.hhi.de/wiegand/assets/pdfs/DIC_SVC_07.pdf.
- [20] International Telecommunication Union, “Mean Opinion Score (MOS) terminology,” ITU-R Rec. P.800.1, 2006.

- [21] International Telecommunication Union, “Methodology for the subjective assessment of the quality of television pictures,” ITU-R Rec. BT.500-11, 2002.
- [22] International Telecommunication Union, “Subjective video quality assessment methods for multimedia applications,” ITU-T Rec. P.910, 1999.
- [23] S. Winkler and P. Mohandas, “The evolution of video quality measurement: From PSNR to hybrid metrics,” *IEEE Trans. Broadcasting*, vol. 54, pp. 660–668, Sep. 2008.
- [24] M. Venkataraman, M. Chatterjee, and S. Chattopadhyay, “Evaluating quality of experience for streaming video in real time,” in *IEEE Global Telecomm. Conf.*, pp. 1–6, Dec. 2009.
- [25] International Telecommunication Union, “User requirements for objective perceptual video quality measurements in digital cable television,” ITU-T Rec. J.1443M, 2000.
- [26] A. Takahashi, D. Hands, and V. Barriac, “Standardization activities in the ITU for a QoE assessment of IPTV,” *IEEE Commun. Magazine*, vol. 46, pp. 78–84, Feb. 2008.
- [27] S. Winkler, *Digital Video Quality – Vision Models and Metrics*. John Wiley & Sons, Jan. 2005.
- [28] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, “Objective video quality assessment methods: A classification, review, and performance comparison,” *IEEE Trans. Broadcasting*, vol. 57, pp. 165–182, Jun. 2011.
- [29] W. Lin and C.-C. J. Kuo, “Perceptual visual quality metrics: A survey,” *J. Visual Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, 2011.
- [30] Available at: http://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio.
- [31] M. H. Pinson and S. Wolf, “A new standardized method for objectively measuring video quality,” *IEEE Trans. Broadcasting*, vol. 50, pp. 312–322, Sep. 2004.
- [32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, pp. 600–612, Apr. 2004.

BIBLIOGRAPHY

- [33] Institute of Electrical and Electronics Engineers, “Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements,” IEEE Standard 802.11e-2005.
- [34] Institute of Electrical and Electronics Engineers, “Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: High-speed Physical Layer in the 5 GHz Band,” IEEE Standard 802.11a-1999.
- [35] C. H. Foh, Y. Zhang, Z. Ni, J. Cai, and K. N. Ngan, “Optimized cross-layer design for scalable video transmission over the IEEE 802.11e networks,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 17, pp. 1665–1678, Dec. 2007.
- [36] M. Jiang and L. Hanzo, “Multiuser MIMO-OFDM for next-generation wireless systems,” *Proc. IEEE*, vol. 95, pp. 1430–1469, July 2007.
- [37] D. Gesbert, M. Shafi, D. shan Shiu, P. Smith, and A. Naguib, “From theory to practice: an overview of MIMO space-time coded wireless systems,” *IEEE J. Sel. Areas Commun.*, vol. 21, pp. 281–302, Apr. 2003.
- [38] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, “An overview of MIMO communications - a key to gigabit wireless,” *Proc. IEEE*, vol. 92, pp. 198–218, Feb. 2004.
- [39] G. J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,” *Bell Labs Tech. J.*, vol. Autumn, pp. 41–59, Oct. 1996.
- [40] L. Hanzo, T. Keller, M. Muenster, and B.-J. Choi, *OFDM and MC-CDMA for Broadband Multi-User Communications, WLANs and Broadcasting*. New York, NY, USA: John Wiley & Sons, Inc., 2003.
- [41] S. M. Alamouti, “A simple transmit diversity technique for wireless communications,” *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.
- [42] V. Tarokh, N. Seshadri, and A. R. Calderbank, “Space-time codes for high data rate wireless communication: performance criterion and code construction,” *IEEE Trans. Info. Theory*, vol. 44, pp. 744–765, Mar 1998.

- [43] J. Blogh and L. Hanzo, *Third-Generation Systems and Intelligent Networking*. Press/Wiley, Available: <http://www-mobile.ecs.soton.ac.uk>, 2002.
- [44] I. Koffman and V. Roman, "Broadband wireless access solutions based on OFDM access in IEEE 802.16," *IEEE Trans. Commun. Magazine*, vol. 40, no. 4, pp. 96–103, 2002.
- [45] M. Etoh and T. Yoshimura, "Advances in wireless video delivery," *Proc. IEEE*, vol. 93, pp. 111–122, Jan. 2005.
- [46] D.-K. Kwon, M.-Y. Shen, and C.-C. Kuo, "Rate control for h.264 video with enhanced rate and distortion models," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 17, pp. 517–529, may 2007.
- [47] F. Zhai, Y. Eisenberg, T. Pappas, R. Berry, and A. Katsaggelos, "Rate-distortion optimized hybrid error control for real-time packetized video transmission," *IEEE Trans. Image Process.*, vol. 15, pp. 40–53, Jan. 2006.
- [48] M. Jiang and N. Ling, "On Lagrange multiplier and quantizer adjustment for H.264 frame-layer video rate control," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 16, pp. 663–669, May 2006.
- [49] H.-F. Chuan, Y. Zhang, Z. Ni, and J. Cai, "Scalable video transmission over the ieee 802.11e networks using cross-layer rate control," pp. 1760–1765, june 2007.
- [50] Y. Zhang, J. Xu, W. H., K. Li, and C. Xu, "Joint sender/receiver rate control algorithm for scalable video streaming," in *IEEE Global Telecommun. Conf.*, pp. 2048–2052, Nov. 2007.
- [51] H. Mansour, V. Krishnamurthy, and P. Nasiopoulos, "Delay-aware rate control for multi-user scalable video streaming over mobile wireless networks," in *Wireless Communications and Mobile Computing Conference*, pp. 476–481, Aug. 2008.
- [52] Y. P. Fallah, H. Mansour, S. Khan, P. Nasiopoulos, and H. Alnuweiri, "An optimized link adaptation scheme for efficient delivery of scalable h.264 video over ieee 802.11n," in *IEEE Int. Symp. Circuits and Syst.*, pp. 2058–2061, May 2008.
- [53] Y. Fallah, H. Mansour, S. Khan, P. Nasiopoulos, and H. Alnuweiri, "A link adaptation scheme for efficient transmission of h.264 scalable video over multirate wlans," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 18, pp. 875–887, July 2008.

BIBLIOGRAPHY

- [54] E. Maani and A. Katsaggelos, "Optimized bit extraction using distortion modeling in the scalable extension of H.264/AVC," *IEEE Trans. Image Processing*, vol. 18, pp. 2022–2029, Sep. 2009.
- [55] Q. Zhang, Z. Ji, W. Zhu, and Y.-Q. Zhang, "Power-minimized bit allocation for video communication over wireless channels," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 12, pp. 398–410, Jun. 2002.
- [56] X. Lu, E. Erkip, Y. Wang, and D. Goodman, "Power efficient multimedia communication over wireless channels," *IEEE J. Sel. Areas Commun.*, vol. 21, pp. 1738–1751, Dec. 2003.
- [57] Z. Ahmad, S. Worrall, and A. Kondo, "Unequal power allocation for scalable video transmission over WiMAX," in *IEEE Conf. Multimedia Expo*, pp. 517–520, Apr. 2008.
- [58] A. M. Kholaf, T. D. Todd, P. Koutsakis, and A. Lazaris, "Energy efficient h.263 video transmission in power saving Wireless LAN infrastructure," *IEEE Trans. Multimedia*, vol. 12, pp. 142–153, Feb. 2010.
- [59] S. Pollin, R. Mangharam, B. Bougard, L. V. D. Perre, I. Moerman, R. Rajkumar, and F. Catthoor, "Meera: cross-layer methodology for energy efficient resource allocation in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 617–628, Feb. 2007.
- [60] S. Pollin, R. Mangharam, B. Bougard, L. Van der Perre, I. Moerman, R. Rajkumar, and F. Catthoor, "Meera: Cross-layer methodology for energy efficient resource allocation in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 98–109, Jan. 2008.
- [61] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 21, pp. 171–178, Feb. 2003.
- [62] D. Bartolome and A. I. Perez-Neira, "Practical implementation of bit loading schemes for multiantenna multiuser wireless OFDM systems," *IEEE Trans. Commun.*, vol. 55, pp. 1577–1587, Aug. 2007.
- [63] A. Fasano, G. Di Blasio, E. Baccarelli, and M. Biagi, "Optimal discrete bit loading for DMT based constrained multicarrier systems," in *Proc. IEEE Int. Symp. Inf. Theory*, p. 243, 2002.

- [64] J. Lee, R. V. Sonalkar, and J. M. Cioffi, "Multiuser bit loading for multicarrier systems," *IEEE Trans. Commun.*, vol. 54, pp. 1170–1174, July 2006.
- [65] Z. Han and K. J. R. Liu, *Resource Allocation for Wireless Networks: Basics, Techniques, and Applications*. New York, USA: Cambridge University Press, 2008.
- [66] C.-F. Tsai, C.-J. Chang, F.-C. Ren, and C.-M. Yen, "Adaptive radio resource allocation for downlink OFDMA/SDMA systems with multimedia traffic," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 1734–1743, May 2008.
- [67] Y. J. Zhang and K. B. Letaief, "An efficient resource-allocation scheme for spatial multiuser access in MIMO/OFDM systems," *IEEE Trans. Commun.*, vol. 53, pp. 107–116, Jan. 2005.
- [68] D. Hu, S. Mao, Y. T. Hou, and J. H. Reed, "Scalable video multicast in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 28, pp. 334–344, Apr. 2010.
- [69] X. Huang and B. Bensaou, "On max-min fairness and scheduling in wireless ad-hoc networks: analytical framework and implementation," in *Proc. of the 2nd ACM international symposium on mobile ad hoc networking & computing*, pp. 221–231, 2001.
- [70] H. J. Kushner and P. A. Whiting, "Convergence of proportional-fair sharing algorithms under general conditions," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1250–1259, July 2004.
- [71] T. Taleb, N. Kato, and Y. Nemoto, "An explicit and fair window adjustment method to enhance TCP efficiency and fairness over multihops satellite networks," *IEEE J. Sel. Areas Commun.*, vol. 22, pp. 371–387, Feb. 2004.
- [72] M. Zukerman, L. Tan, H. Wang, and I. Ouveysi, "Efficiency-fairness tradeoff in telecommunications networks," *IEEE Commun. Letters*, vol. 9, pp. 643–645, July 2005.
- [73] 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. Circuits Syst. Video Techn.*, vol. 16, pp. 1217–1231, Oct. 2006.
- [74] H. Hu, Y. Guo, and Y. Liu, "Peer-to-peer streaming of layered video: Efficiency, fairness and incentive," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 21, pp. 1013–1026, Aug. 2011.

BIBLIOGRAPHY

- [75] E. B. Rodrigues and F. Casadevall, "Control of the trade-off between resource efficiency and user fairness in wireless networks using utility-based adaptive resource allocation," *IEEE Commun. Magazine*, vol. 49, pp. 90–98, Sep. 2011.
- [76] M. Kalman, E. Steinbach, and B. Girod, "R-D optimized media streaming enhanced with adaptive media playout," in *IEEE Conf. Multimedia Expo*, vol. 1, pp. 869–872, 2002.
- [77] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Trans. Multimedia*, vol. 8, pp. 390–404, Apr. 2006.
- [78] B. Girod and N. Farber, "Feedback-based error control for mobile video transmission," *Proc. IEEE*, vol. 87, pp. 1707–1723, Oct. 1999.
- [79] Y. Shan and A. Zakhor, "Cross layer techniques for adaptive video streaming over wireless networks," in *IEEE Conf. Multimedia Expo*, vol. 1, pp. 277–280, 2002.
- [80] J. Kim, R. Mersereau, and Y. Altunbasak, "Error-resilient image and video transmission over the Internet using unequal error protection," *IEEE Trans. Image Process.*, vol. 12, pp. 121–131, Feb. 2003.
- [81] M. van der Schaar and H. Radha, "Adaptive motion-compensation fine-granular-scalability (AMC-FGS) for wireless video," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 12, pp. 360–371, Jun. 2002.
- [82] A. Mohr, E. Riskin, and R. Ladner, "Unequal loss protection: graceful degradation of image quality over packet erasure channels through forward error correction," *IEEE J. Sel. Areas Commun.*, vol. 18, pp. 819–828, Jun. 2000.
- [83] International Telecommunication Union, "Quality of telecommunication services: concepts, models, objectives and dependability planning. Terms and definitions related to the quality of telecommunication services," ITU-R Rec. E.800, 1994.
- [84] X. Gu and K. Nahrstedt, "Distributed multimedia service composition with statistical QoS assurances," *IEEE Trans. Multimedia*, vol. 8, pp. 141–151, Feb. 2006.
- [85] Q. Zhang, W. Zhu, and Y. Zhang, "End-to-end qos for video delivery over wireless internet," *Proc. IEEE*, vol. 93, pp. 123–134, Jan. 2005.

- [86] S. Ci, H. Wang, and D. Wu, "A theoretical framework for quality-aware cross-layer optimized wireless multimedia communications," *Adv. MultiMedia*, vol. 2008, no. 2, pp. 1–10, 2008.
- [87] P. Zhu, W. Zeng, and C. Li, "Joint design of source rate control and qos-aware congestion control for video streaming over the internet," *IEEE Trans. Multimedia*, vol. 9, pp. 366–376, Feb. 2007.
- [88] W. Kumwilaisak, Y. T. Hou, Q. Zhang, W. Zhu, C.-C. J. Kuo, and Y.-Q. Zhang, "A cross-layer quality-of-service mapping architecture for video delivery in wireless network," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 10, pp. 1685–1698, 2003.
- [89] Z. He, J. Cai, and C. W. Chen, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 12, pp. 511–523, Jun. 2002.
- [90] K. B. Letaief and Y. J. Zhang, "Dynamic multiuser resource allocation and adaptation for wireless systems," *IEEE Wireless Commun.*, vol. 13, pp. 38–47, Aug. 2006.
- [91] H. Zhang, Y. Zheng, M. Khojastepour, and S. Rangarajan, "Cross-layer optimization for streaming scalable video over fading wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 28, pp. 344–353, Apr. 2010.
- [92] H. Luo, S. Ci, D. Wu, J. Wu, and H. Tang, "Quality-driven cross-layer optimized video delivery over LTE," *IEEE Commun. Magazine*, vol. 48, pp. 102–109, Feb. 2010.
- [93] "New Appendix I - Definition of quality of experience (QoE)," ITU-R Rec. P.10/G.100 Amend. 1, 2007.
- [94] R. Stankiewicz, P. Cholda, and A. Jajszczyk, "QoX: What is it really?," *IEEE Commun. Magazine*, vol. 49, pp. 148–158, Apr. 2011.
- [95] Available at: <http://www.dialogic.com/ /media/products/docs/whitepapers/11681-qoe-mobile-video-wp.pdf>.
- [96] International Telecommunication Union, "Quality assessment model for video-telephony services," ITU-T Rec. G.1070, 2007.

BIBLIOGRAPHY

- [97] International Telecommunication Union, “Perceptual visual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference,” ITU-T Rec. J.246, 2008.
- [98] International Telecommunication Union, “Objective perceptual multimedia video quality measurement in the presence of a full reference,” ITU-T Rec. J.247, 2008.
- [99] M. Venkataraman and M. Chatterjee, “Inferring video QoE in real time,” *IEEE Network*, vol. 25, pp. 4–13, Jan.-Feb. 2011.
- [100] M. Volk, J. Sterle, U. Sedlar, and A. Kos, “An approach to modeling and control of QoE in next generation networks,” *IEEE Commun. Magazine*, vol. 48, pp. 126–135, Aug. 2010.
- [101] A. ParandehGheibi, M. Médard, A. Ozdaglar, and S. Shakkottai, “Avoiding interruptions - a QoE reliability function for streaming media applications,” *IEEE J. Sel. Areas Commun.*, vol. 29, pp. 1064–1074, May 2011.
- [102] M. Fiedler, T. Hossfeld, and P. Tran-Gia, “A generic quantitative relationship between quality of experience and quality of service,” *IEEE Network*, vol. 24, pp. 36–41, Mar.-Apr. 2010.
- [103] Z. Chen, W. Lin, and K. N. Ngan, “Perceptual video coding: Challenges and approaches,” in *IEEE Conf. Multimedia Expo*, pp. 784–789, July 2010.
- [104] A. Vetro, T. Wiegand, and G. J. Sullivan, “Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard,” *Proceed. IEEE*, vol. 99, pp. 626–642, Apr. 2011.
- [105] A. Smolic and P. Kauff, “Interactive 3-D video representation and coding technologies,” *Proceed. IEEE*, vol. 93, pp. 98–110, Jan. 2005.
- [106] A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G. B. Akar, G. Triantafyllidis, and A. Koz, “Coding algorithms for 3DTV - A survey,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 17, pp. 1606–1621, Nov. 2007.
- [107] K. Ugur, K. Andersson, A. Fuldseth, G. Bjontegaard, L. P. Endresen, J. Lainema, A. Hallapuro, J. Ridge, D. Rusanovskyy, C. Zhang, A. Norkin, C. Priddle, T. Ruset,

- J. Samuelsson, R. Sjöberg, and Z. Wu, "High performance, low complexity video coding and the emerging HEVC standard," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 20, pp. 1688–1697, Dec. 2010.
- [108] M. Li, Z. Chen, and Y.-P. Tan, "Cross-layer optimization for SVC video delivery over the IEEE 802.11e wireless networks," *J. Visual Commun. Image Represent.*, vol. 22, pp. 284–296, Apr. 2011.
- [109] M. Li, Z. Chen, and Y.-P. Tan, "Joint packet prioritization and QoS mapping for SVC over WLANs," in *IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, pp. 2298–2301, Mar. 2010.
- [110] S. Aramvith, C.-W. Lin, S. Roy, and M.-T. Sun, "Wireless video transport using conditional retransmission and low-delay interleaving," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 12, pp. 558–565, June 2002.
- [111] C.-M. Chen, C.-W. Lin, and Y.-C. Chen, "Adaptive error-resilience transcoding using prioritized intra-refresh for video multicast over wireless networks," *Image Commun.*, vol. 22, no. 3, pp. 277–297, 2007.
- [112] C.-M. Chen, C.-M. Chen, C.-W. Lin, and Y.-C. Chen, "Error-resilient video streaming over wireless networks using combined scalable coding and multiple-description coding," *Image Commun.*, vol. 22, no. 4, pp. 403–420, 2007.
- [113] C.-M. Chen, C.-W. Lin, H.-C. Wei, and Y.-C. Chen, "Robust video streaming over wireless lans using multiple description transcoding and prioritized retransmission," *J. Visual Commun. Image Represent.*, vol. 18, no. 3, pp. 191–206, 2007.
- [114] Y. Zhu, W. Zeng, H. Liu, Y. Guo, and S. Mathur, "Supporting video streaming services in infrastructure wireless mesh networks: Architecture and protocols," in *IEEE Conf. Commun.*, pp. 1850–1855, May 2008.
- [115] Y. Liu, Q. Huang, S. Ma, D. Zhao, and W. Gao, "RD-optimized interactive streaming of multiview video with multiple encodings," *J. Visual Commun. Image Represent.*, vol. 21, no. 5-6, pp. 523–532, 2010.

BIBLIOGRAPHY

- [116] A. Fiandrotti, D. Gallucci, E. Masala, and E. Magli, “Traffic prioritization of H.264/SVC video over 802.11e ad hoc wireless networks,” in *Proc. 17th Int. Conf. Computer Commun. Networks*, pp. 1–5, Aug. 2008.
- [117] H.-L. Chen, P.-C. Lee, and S.-H. Hu, “Improving scalable video transmission over IEEE 802.11e through a cross-layer architecture,” in *Int. Conf. Wireless Mobile Commun.*, pp. 241–246, 2008.
- [118] A. Majumda, D. Sachs, I. Kozintsev, K. Ramchandran, and M. Yeung, “Multicast and unicast real-time video streaming over wireless lans,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 12, pp. 524–534, Jun. 2002.
- [119] H. Park and M. van der Schaar, “Fairness strategies for wireless resource allocation among autonomous multimedia users,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 20, pp. 297–309, Feb. 2010.
- [120] A. R. Fattahi, F. Fu, M. van der Schaar, and F. Paganini, “Mechanism-based resource allocation for multimedia transmission over spectrum agile wireless networks,” *IEEE J. Sel. Areas Commun.*, vol. 25, pp. 601–612, Apr. 2007.
- [121] X. Fan, W. Gao, Y. Lu, and D. Zhao, “Flicking reduction in all intra frame coding,” in *JVT-E070, Joing Video Team of ISO/IEC MPEG & ITU-T VCEG Meeting*, (Geneva, Switzerland), pp. 1–5, Oct. 2002.
- [122] F. C. Leone, L. S. Nelson, and R. B. Nottingham, “The folded normal distribution,” in *Technometrics*, pp. 543–550, Nov. 1961.
- [123] X. Yang, “Performance Analysis of Priority Schemes for IEEE 802.11 and IEEE 802.11e Wireless LANs,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1506–1515, 2005.
- [124] J. Hu, G. Min, M. E. Woodward, and W. Jia, “A comprehensive analytical model for IEEE 802.11e QoS differentiation schemes under unsaturated traffic loads,” in *IEEE Conf. Commun.*, pp. 241–245, May 2008.
- [125] L. Romdhani and C. Bonnet, “Performance analysis and optimization of the 802.11e EDCA transmission opportunity (TXOP) mechanism,” in *IEEE Int. Conf. Wireless and Mobile Comput. Netw. and Commun.*, pp. 68–73, Oct 2007.

- [126] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, 2000.
- [127] D. P. H. Kellerer, U. Pferschy, *Knapsack Problems*. Springer Verlag, 2005.
- [128] G. Marfia and M. Roccetti, "Dealing with wireless links in the ERA of bandwidth demanding wireless home entertainment," in *IEEE Conf. Multimedia Expo*, pp. 1376–1381, 2010.
- [129] J. Turner, "New directions in communications (or which way to the information age?)," *IEEE Commun. Magazine*, vol. 24, pp. 8–15, Oct. 1986.
- [130] P. Ferguson and G. Huston, *Quality of service: Delivering QoS on the Internet and in corporate networks*. New York, NY, USA: John Wiley & Sons, Inc., 1998.
- [131] The network simulator NS-2. Available: <http://www.isi.edu/nsnam/ns/>.
- [132] S. Wietholter, M. Emmelmann, C. Hoene, and A. Wolisz, "TKN EDCA model for NS-2," in *TKN Technical Report TKN-06-003*.
- [133] J. Reichel, H. Schwarz, and M. Wien, "JSVM 9.8 Software, Joint Video Team of ISO/IEC MPEG and ITU-T VCEG N9212," (Geneva, Switzerland), June 2007.
- [134] J. Sun, W. Gao, D. Zhao, and W. Li, "On Rate-Distortion modeling and extraction of H.264/SVC fine-granular scalable video," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 19, pp. 323–336, Mar. 2009.
- [135] J. Liu, Y. Cho, Z. Guo, and C.-C. J. Kuo, "Bit allocation for spatial scalability coding of H.264/SVC with dependent rate-distortion analysis," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 20, pp. 967–981, July 2010.
- [136] R. Li, J. Sun, and W. Gao, "Fast weighted algorithms for bitstream extraction of SVC Medium-Grain scalable video coding," in *IEEE Conf. Multimedia Expo*, pp. 249–254, July 2010.
- [137] Joint Video Team (JVT) of ISO-IEC MPEG & ITU-T VCEG, N9195, *Text of ISO/IEC 14496-4:2001/PDAM 19 Reference Software for SVC*, Sep. 2007.

BIBLIOGRAPHY

- [138] I. Amonou, N. Cammas, S. Kervadec, and S. Pateux, “Optimized rate-distortion extraction with quality layers in the scalable extension of H.264/AVC,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 17, pp. 1186–1193, Sep. 2007.
- [139] L. Lima, M. Mauro, and R. Leonardi, “Optimal rate adaptation in the scalable extension of H.264/AVC with combined scalability,” in *19th European Signal Process. Conf.*, Aug. 2011.
- [140] W.-H. Peng, J. K. Zao, H.-T. Huang, T.-W. Wang, and L.-S. Huang, “A rate-distortion optimization model for SVC inter-layer encoding and bitstream extraction,” *J. Visual Commun. Image Represent.*, vol. 19, no. 8, pp. 543–557, 2008.
- [141] M. Li, Z. Chen, S.-P. Chuah, and Y.-P. Tan, “A fast rate adaptation scheme for SVC based on the packet dependencies,” in *IEEE Int. Symp. Circuits Syst.*, May 2013.
- [142] M. Wien, H. Schwarz, and T. Oelbaum, “Performance analysis of SVC,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 17, pp. 1194–1203, Sep. 2007.
- [143] G. Bjontegaard, “Calculation of average PSNR differences between RD-curves,” Austin, USA: VCEG Meeting (ITU-T SG16 Q.6), Apr. 2001.
- [144] Executable programs with sample videos are available at: <http://www.ntu.edu.sg/home2007/e080002>.
- [145] G. L. Stuber, J. R. Barry, S. W. McLaughlin, Y. Li, M. A. Ingram, and T. G. Pratt, “Broadband MIMO-OFDM wireless communications,” *Proc. IEEE*, vol. 92, pp. 271–294, Feb. 2004.
- [146] Z. Cao and E. W. Zegura, “Utility max-min: an application-oriented bandwidth allocation scheme,” in *IEEE INFOCOM*, vol. 2, pp. 793–801, Mar. 1999.
- [147] T. Harks and T. Poschwatta, “Utility fair congestion control for real-time traffic,” in *IEEE INFOCOM*, pp. 1–6, Apr. 2006.
- [148] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, “Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels,” *IEEE Trans. Signal Process.*, vol. 52, pp. 461–471, Feb. 2004.

- [149] X. Ji, J. Huang, M. Chiang, G. Lafruit, and F. Catthoor, "Scheduling and resource allocation for SVC streaming over OFDM downlink systems," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 19, pp. 1549–1555, Oct. 2009.
- [150] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: a unified view," *IEEE Trans. Commun.*, vol. 49, pp. 1561–1571, Sep. 2001.
- [151] L. A. Wolsey, *Integer Programming*. Wiley, 1998.
- [152] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Athena Scientific, 2007.
- [153] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: complexity and duality," *IEEE J. Sel. Topics Signal Process.*, vol. 2, pp. 57–73, Feb. 2008.
- [154] G. M. Schuster, G. Melnikov, and A. K. Katsaggelos, "A review of the minimum maximum criterion for optimal bit allocation among dependent quantizers," *IEEE Trans. Multimedia*, vol. 1, pp. 3–17, 1999.
- [155] S. Martello and P. Toth, *Knapsack problems: algorithms and computer implementations*. New York, NY, USA: John Wiley & Sons, Inc., 1990.
- [156] H. Yaghoobi, "Scalable OFDMA physical layer in IEEE 802.16 WirelessMAN," *Intel Techn. Journal*, vol. 8, pp. 201–212, Aug. 2004.
- [157] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," *DEC Research Report*, Sep. 1984.
- [158] Z. Chen, "Quality of experience issues in media delivery," *IEEE MMTC E-Letter*, vol. 6, Aug. 2011.
- [159] P. Brooks and B. Hestnes, "User measures of quality of experience: why being objective and quantitative is important," *IEEE Network*, vol. 24, pp. 8–13, Mar.-Apr. 2010.
- [160] N. Staelens, S. Moens, W. Van den Broeck, I. Mariën, B. Vermeulen, P. Lambert, R. Van de Walle, and P. Demeester, "Assessing the perceptual influence of H.264/SVC Signal-to-Noise Ratio and temporal scalability on full length movies," in *QoMEX'09*, pp. 29–34, July 2009.

BIBLIOGRAPHY

- [161] N. Staelens, S. Moens, W. Van den Broeck, I. Mariën, B. Vermeulen, P. Lambert, R. Van de Walle, and P. Demeester, “Assessing quality of experience of IPTV and Video on Demand services in real-life environments,” *IEEE Trans. Broadcasting*, vol. 56, pp. 458–466, Dec. 2010.
- [162] T. Zinner, O. Hohlfeld, O. Abboud, and T. Hossfeld, “Impact of frame rate and resolution on objective QoE metrics,” in *QoMEX’10*, pp. 29–34, June 2010.
- [163] Y.-F. Ou, Z. Ma, T. Liu, and Y. Wang, “Perceptual quality assessment of video considering both frame rate and quantization artifacts,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 21, pp. 286–298, Mar. 2011.
- [164] G. Zhai, J. Cai, W. Lin, X. Yang, W. Zhang, and M. Etoh, “Cross-dimensional perceptual quality assessment for low bit-rate videos,” *IEEE Trans. Multimedia*, vol. 10, pp. 1316–1324, Nov. 2008.
- [165] J. Brandt and L. Wolf, “Impact of adaptation dimensions on video quality,” in *IEEE 14th International Symposium on Consumer Electronics*, pp. 1–6, June 2010.
- [166] Z. Ma, M. Xu, and Y. Wang, “Modeling video rate as a function of frame size, frame rate and quantization stepsize,” in *Proc. of IEEE MMSP*, Oct. 2011.
- [167] J.-S. Lee, F. De Simone, N. Ramzan, Z. Zhao, E. Kurutepe, T. Sikora, J. Ostermann, E. Izquierdo, and T. Ebrahimi, “Subjective evaluation of scalable video coding for content distribution,” in *ACM International Conference on Multimedia*, pp. 65–72, 2010.
- [168] M. Li, Z. Chen, and Y.-P. Tan, “On Quality of Experience of scalable video adaptation,” *J. Visual Commun. Image Represent.*, vol. 24, no. 5, pp. 509–521, 2013.
- [169] M. Li, Z. Chen, and Y.-P. Tan, “QoE-aware resource allocation for scalable video streaming over multiuser MIMO-OFDM systems,” *under review in IEEE Trans. Multimedia*.
- [170] “JSVM Software Manual, Ver. JSVM 9.19,” Mar. 2011.
- [171] Y.-F. Ou, Y. Xue, Z. Ma, and Y. Wang, “A perceptual video quality model for mobile platform considering impact of spatial, temporal, and amplitude resolutions,” in *IEEE IVMSWP Workshop*, pp. 117–122, Jun. 2011.

- [172] Y.-F. Ou, Y. Xue, and Y. Wang, “Q-STAR: A perceptual video quality model for mobile platforms considering impact of spatial, temporal, and amplitude resolutions,” *Under review, Available at arxiv.org*.
- [173] H. Hu, X. Zhu, Y. Wang, R. Pan, J. Zhu, and F. Bonomi, “Proxy-based multi-stream scalable video adaptation over wireless networks using subjective quality and rate models,” *IEEE Trans. Multimedia, accepted with minor revision*.
- [174] H. Hu, Z. Ma, and Y. Wang, “Optimization of spatial, temporal and amplitude resolution for rate-constrained video coding and scalable video adaptation,” in *IEEE Conf. Image Process.*, Sep. 2012.
- [175] Available at: <http://wftp3.itu.int/av-arch/jvt-site/>.
- [176] G. W. Snedecor and W. G. Cochran, *Statistical Methods, Eighth Edition*. Iowa State Univ. Press, 1989.
- [177] S. Hayashi and Z.-Q. Luo, “Spectrum management for interference-limited multiuser communication systems,” *IEEE Trans. Info. Theory*, vol. 55, pp. 1153–1175, Mar. 2009.
- [178] M. Li, Z. Chen, S.-P. Chuah, and Y.-P. Tan, “Efficient packet scheduling for scalable video delivery to mobile clients,” in *IEEE Int. Symp. Circuits Syst.*, pp. 2251–2254, May 2010.
- [179] Z. Chen, M. Li, and Y.-P. Tan, “Perception-aware multiple scalable video streaming over WLANs,” *IEEE Signal Process. Letters*, vol. 17, pp. 675–678, July 2010.
- [180] M. Li, S.-P. Chuah, Z. Chen, and Y.-P. Tan, *Streaming Media Architectures, Techniques, and Applications: Recent Advances*, ch. Scalable video delivery over Wireless LANs. IGI Press, 2010.
- [181] M. Li, Z. Chen, and Y.-P. Tan, “Scalable resource allocation for SVC video streaming over multiuser MIMO-OFDM networks,” *accepted by IEEE Trans. Multimedia*.
- [182] M. Li, Z. Chen, and Y.-P. Tan, “Joint power allocation and bit loading for enhanced SVC video downlink transmissions over SDMA/OFDMA networks,” in *IEEE Conf. Multimedia Expo*, pp. 1–6, July 2011.

BIBLIOGRAPHY

- [183] M. Li, Z. Chen, and Y.-P. Tan, “A scalable resource allocation framework for SVC video transmissions over downlink MIMO-OFDM networks,” in *IEEE Int. Symp. Circuits Syst.*, May 2012.
- [184] M. Li, Z. Chen, and Y.-P. Tan, “A MAXMIN resource allocation approach for scalable video delivery over multiuser MIMO-OFDM systems,” in *IEEE Int. Symp. Circuits Syst.*, pp. 2645–2648, May 2011.
- [185] M. Li, Z. Chen, and Y.-P. Tan, “Scalable video transmission over multiuser MIMO-OFDM systems,” in *5th Int. Conf. Commun. and Networking in China*, pp. 1–8, Aug. 2010.
- [186] M. Li, Z. Chen, and Y.-P. Tan, “QoE analysis for scalable video adaptation,” in *IEEE Visual Commun. Image Process.*, Nov. 2012.
- [187] M. Li, Z. Chen, and Y.-P. Tan, “QoE-aware resource allocation for scalable video transmission over multiuser MIMO-OFDM systems,” in *IEEE Visual Commun. Image Process.*, Nov. 2012.