Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN
SCIENZE DELL'INGEGNERIA

CICLO XXIV

COORDINATORE Prof. Stefano Trillo

QUALITY OF EXPERIENCE AND ADAPTATION TECHNIQUES FOR MULTIMEDIA COMMUNICATIONS

Settore Scientifico Disciplinare ING-INF/03

Dottorando
Dott. Haseeb Abdul
A. Haseeb
(firma)

Tutore
Prof. Tralli Velio
Tralli
(firma)

Co-tutore
Prof. Mazzini Gianluca
Gianluca Mazzini
(firma)

Anni 2009/2011

Abstract

The widespread use of multimedia services on the World Wide Web and the advances in end-user portable devices have recently increased the user demands for better quality. Moreover, providing these services seamlessly and ubiquitously on wireless networks and with user mobility poses hard challenges. To meet these challenges and fulfill the end-user requirements, suitable strategies need to be adopted at both application level and network level. At the application level rate and quality have to be adapted to time-varying bandwidth limitations, whereas on the network side a mechanism for efficient use of the network resources has to be implemented, to provide a better end-user Quality of Experience (QoE) through better Quality of Service (QoS). The work in this thesis addresses these issues by first investigating multi-stream rate adaptation techniques for Scalable Video Coding (SVC) applications aimed at a fair provision of QoE to end-users. Rate Distortion (R-D) models for real-time and non real-time video streaming have been proposed and a rate adaptation technique is also developed to minimize with fairness the distortion of multiple videos with difference complexities. To provide resiliency against errors, the effect of Unequal Error protection (UXP) based on Reed Solomon (RS) encoding with erasure correction has been also included in the proposed R-D modelling. Moreover, to improve the support of QoE at the network level for multimedia applications sensitive to delays, jitters and packet drops, a technique to prioritise different traffic flows using specific QoS classes within an intermediate DiffServ network integrated with a WiMAX access system is investigated. Simulations were performed to test the network under different congestion scenarios.

Acknowledgements

First and above all, I praise God, the almighty for providing me this opportunity and granting me the capability to proceed successfully. This thesis appears in its current form due to the assistance and guidance and prayers of several people. I would therefore like to offer my sincere thanks to all of them.

I am thankful to my adviser Velio Tralli for his support, guidance and ideas that he gave during my Ph.D which made this thesis possible. I appreciate all his contributions of time and ideas to make my Ph.D experience productive and stimulating. The joy and enthusiasm he has for his research was contagious and motivational for me, even during tough times in the Ph.D pursuit. I am grateful to Gianluca Mazzini and Andrea Conti for their active support whenever I needed.

I am also thankful to Maria Martini for her supervision, guidance and funding me in *Royal Society Project* for the months that I spent for my research in *Kingston University, London.*

I cannot forget the forever and everlasting support of my parents and siblings. I am thankful to my elder brother Asif for his financial support before starting my Ph.D and my parents and siblings for theirs prayers. In short I cannot express my gratitude to them in words.

I am thankful to my colleagues in TLC lab. They were always helpful and cooperative to me in academic and non academic acttivites, to name few in particular are Danilo, Honorine and Sergio.

My time in Ferrara is the best I had so far and it was made enjoyable in large part due to the many friends and groups that became a part of my life. In particular I am thankful to all my friends of residential college “Cenacolo” where I found myself as to be in my home.

I will also extend my gratitude to Lena Fabbri as she helped me a lot, for official work both in and out of university and during my stay in Ferrara.

Last but not the least I am thankful to University of Ferrara for giving me the opportunity and funding me during these three years of my Ph.D.

Abdul Haseeb

University of Ferrara, *Italy*

March 2012

DEDICATION

To my Loving Parents

Ma and *Pa*

Contents

Abstract iii

Acknowledgements v

Contents viii

List of Tables x

List of Figures xi

1 Introduction 1

1.1 Scope of the Thesis 2

1.2 Outline of Thesis 3

2 Overview of Scalable Video Coding 5

2.1 Introduction 5

2.2 Concept of H.264/AVC Extension to H.264/SVC 6

2.3 Types of Scalabilities 8

2.3.1 Spatial scalability 8

2.3.2 Temporal scalability 9

2.3.3 SNR scalability/Quality scalability 11

2.3.3.1 Coarse Grain Scalability (CGS) 12

2.3.3.2 Medium Grain Scalability (MGS) 12
2.3.3.3 Fine Grain Scalability (FGS) 13
2.4 Backward Compatibility 14

3 Rate Adaptation Using MGS for SVC **17**
3.1 Introduction 17
3.2 General Problem Formulation for Multi-Stream Rate Adaptation 20
3.3 Rate Distortion Model for MGS with Quality Layer 23
3.4 GOP-Based Multi-Stream Rate Adaptation Framework 27
3.4.1 Problem Solution 28
3.5 Numerical Results 32
3.6 Conclusions 37

4 Rate Adaptation for Error Prone Channels in SVC **39**
4.1 Introduction 39
4.1.1 Related Works 41
4.2 Unequal Erasure Protection 44
4.2.1 Frame error probability and expected distortion 46
4.2.2 Proposed UXP profiler 47
4.2.2.1 A case study for the design of EEP 49
4.3 Rate-distortion modeling with Packet Losses 49
4.4 Packet-erasure channel 52
4.5 Conclusions 54

5 Rate Distortion Modeling for Real-time MGS Coding **55**
5.1 Introduction 55
5.2 Overview of Rate Distortion modeling 57
5.3 Proposed Model 58

5.3.1 Validation of the proposed models 66
5.4 Simulation and Model Verification 66
5.5 Conclusions 69

6 QoS for VoIP Traffic in Heterogeneous Networks 71
6.1 Introduction 71
6.2 Mechanism for IP QoS 73
6.3 QoS Mechanism in WiMAX Network 74
6.4 Inter-Working Model and Simulation 77
6.4.1 Priority Queuing (PQ) 78
6.4.2 Simulation Results 79
6.5 Conclusions 84

Bibliography 89

List of Tables

3.1 Comparison between the two semi-analytical model in (3.9) and (3.10) with respect to the minimum and maximum RMSE and the coefficient of determination R2 evaluated for each GOP (GOP size equal to 16) of five video sequence with CIF resolution and frame rate of 30 fps. The video are encoded with one base layer (QP equal to 38) and two enhancement layers (QP equal to 32 and 26), both with 5 MGS layers and a weights vector equal to [3 2 4 2 5]. 26

3.2 Average MSE of each video sequence with equal-rate (ER) assignment and rate adaptation with the proposed algorithm (OPT). Total bandwidth is equal to 3000 kbps. 34
3.3 Average modified MSE difference Δ_{av}, average MSE difference δ_{av} and MSE variance in each GOP interval. Comparison between the proposed algorithm (OPT) and equal-rate (ER) assignment with bandwidth equal to 3000 kbps. . . 36

4.1 Percentage of the overhead and expected distortion $d_{GQ,loss}$ in term of MSE with respect to the full quality video streams ($Q = 10$ and $G = 8$), for different values of RTP packet error probability and α parameter in the EEP profile. . . . 51
4.2 Average received distortion,$D^*_{rec,av}$, expected distortion, D^*_{av}, and encoding distortion, $D^*_{enc,av}$, in term of the MSE for different video sequences, GOP size G, and packet-erasure rate values $P_{e,rtp}$, resulting from the proposed rate-adaptation algorithm. Available bandwidth is R_c =7000 kbps. 53

5.1 Average MSE over 26 GOPs obtained with the model (5.1) and proposed model in the transmission of the training set of 6 videos. 67
5.2 Average MSE over 26 GOPs obtained with the model (5.1) and proposed model in the transmission of 4 videos not included in the training set. 67

6.1 WiMAX and DiffServ traffic class mapping 78

List of Figures

2.1 Principle of encoding. 7

2.2 Principle of decoding. 7
2.3 Multi-layer structure with additional inter-layer prediction for enabling spatial scalable coding. 9
2.4 Enhancement temporal (a) and quality (b) layer prediction for a GOP of 8 frames. 10
2.5 Fine granular scalability. 13

3.1 R-D Model (straight line), according to eq. (3.10) fitting the empirical R-D relationship for the GOP with the worst RMSE with reference to Table 3.1. . . 26
3.2 Rate assigned by our adaptation algorithm in each GOP, with bandwidth equal to 3000 kbps. 36
3.3 Variance of the MSE averaged over 15 GOPs, with different bandwidth values. Comparison between the proposed algorithm (OPT) and equal-rate (ER) assignment. 37
3.4 Average number of iterations required by our adaptation algorithm (OPT) and golden search algorithm (GSA) to converge. 37

4.1 System architecture. Each sequence is encoded to fully support temporal and quality scalability and a priority level is assigned to the NALUs. The UXP profiler evaluates the overhead required according to a certain protection policy and RTP packet failure rate, and provides R-D information to the Adaptation module. The Adaptation module extracts sub-streams according to the estimated bandwidth and sends the data bytes to the RS encoder. The resulting codewords are then encapsulated in a transmission block, interleaved in RTP packets and forwarded to the lower layers. The receiver performs the inverse operations (RS decoding and deinterleaving) in order to extract the NALUs which are sent to the SVC decoder. 42

4.2 Transmission Sub-Block (TSB) structure. Following the priority level, the NALUs of one GOP are placed into one TSB according to a given UXP profile (protection class) from upper left to lower right. The columns of one ore more TSB are then encapsulated into RTP packets 45
4.3 Resulting logarithmic FEP for the first I frame of *Football* (byte size equal to 11519) mapped to RS codewords (128,*m*) at different RTP packet error probability. 48
4.4 R-D Model (straight line), according to eq. (3.10) fitting the empirical R-D relationship for one GOP (size G equal to 8) of the *Football* test-sequence with different error probabilities and α=30. The lower curve refers to the R-D relationship of the encoder. 50

5.1 Proposed model for α with R^2= 0.987 and $RMSE$ = 1598 61
5.2 Proposed model for β with R^2= 0.973 and $RMSE$ = 21.2 62
5.3 Proposed model for (BL) with R^2= 0.979 and $RMSE$ = 22.98 62
5.4 Proposed model for (EL) with R^2= 0.985 and $RMSE$ = 79.36 63
5.5 R-D comparison among model in eq (5.1), proposed model and actual values for two sample GOPs. 64
5.6 BL and EL rates over 26 GOPs for two sequences in the training set (*Football* and *City*) and two sequences outside the training set (*Mobile* and *Foreman*). The marker points refer to the original BL and EL rates, whereas the solid lines refer to rates estimated from (5.4) and (5.5), respectively. 65
5.7 Averaged MSE for each GOP of two sample videos in the transmission over bandwidth constrained channel with rate adaptation. Figures *Football* and *City* refer to the transmission of the 6 videos of the training set ($R_c = 3500$ kbps), wheres figures *Mobile* and *Foreman* refer to the transmission of 4 videos not included in the training set ($R_c = 3000$ kbps). 68

6.1 DiffServ Code Point field. 74
6.2 IEEE 802.16 QoS Architecture 76
6.3 WiMAX and DiffServ Network Simulation Scenario 77
6.4 Priority Queuing Implemented in Edge Router 79
6.5 Packets dropped without DiffServ support. 80
6.6 Delay without DiffServ support and with 100% load. 81
6.7 Delay with DiffServ support and 100% load. 81
6.9 Delay with DiffServ support and 112.5% load 82
6.8 Delay without DiffServ support and 112.5% load. 82
6.10 Delay without DiffServ support and 125% load 83
6.11 Delay with DiffServ support and 125% load 83
6.12 VoIP service jitters in networks with and without DiffServ. 84

Chapter 1

Introduction

The popularity of multimedia applications is rapidly increasing. Multimedia applications include video on demand, IP-TV, sport broadcasting, VoIP as well as real-time streaming. They have become reality now thanks to the achievements in the compression and storage technologies and the advances in transmission systems. The penetration of end user devices such as 3G mobile devices, portable multimedia players (PMP), HDTV flat-panel displays, and the availability of wired and wireless broadband internet access provides different ways to deliver these multimedia services. Nevertheless, in such environment providing contents everywhere while achieving efficiency is a challenge. Scalable Video Coding (SVC) which is the extension of Advance Video Coding standard H.264/AVC provides an attractive solution to support video transmission in modern communication systems. In SVC some parts of the encoded video can be removed so that the video stream can be adapted to the network conditions. Moreover, SVC can fulfill the requirements of the users with different terminal capabilities and varying network conditions by providing spatial, temporal and quality scalabilities. Multimedia contents like voice and video can tolerate only to some extent jitters, delays and packet drops but they need sufficiently wide bandwidth. WiMAX, which is considered an alternative to DSL, can provide wireless broadband connectivity with its rich set of QoS classes for different types of multimedia applications which can be further

translated to the intermediate wired networks like DiffServ.

1.1 Scope of the Thesis

The aim of this thesis was to study the issues related to the provision of better Quality of Experience (QoE) for multimedia applications like video and voice and in this context particular emphasis was given to Scalable Video Coding (SVC). This thesis proposes new continuous Rate-Distortion (R-D) models both for real-time and non real-time videos. New rate adaptation techniques based on fairness for multi-stream video communication are also developed and applied to both the real and non real-time R-D models. Moreover, to further enhance the model an Unequal Error Protection (UXP) mechanism is introduced to cope with errors during transmission. The non real-time R-D model takes advantage of the SVC encoder to get the original R-D points from the video sequence and find the best possible R-D couple by curve fitting technique. To develop R-D models for real-time video transmission, raw video sequences are exploited to get Spatial Indexes (SI) and Temporal Indexes (TI), which are also referred as spatial and temporal complexities, to be used to predict the parameters of the R-D model, as well as the rate prior to the encoding process of the videos. The multimedia information may flow through several networks before reaching to the final destination. This can be the cause of quality degradation for the application in use because of the bandwidth limitation of heterogeneous networks, the absence of prioritizing policies for delay sensitive traffic in intermediate networks, and network congestion, just to name few of them. To address this issue in limited context a solution for WiMAX and DiffServ interworking is provided in which VoIP traffic from WiMAX network entering DiffServ network is prioritized to get a preferential treatment in DiffServ networks and minimize delays, jitters and packets drops.

1.2 Outline of Thesis

This thesis is organized as follows: In Chapter 2 the basic concepts of SVC are explained. In Chapter 3 a semi-analytical R-D model is proposed for non real-time SVC and, also a multi-stream rate adaptation technique based on fairness among videos is developed. The rate adaptation technique is then applied to the proposed R-D model and compared the results with the Equal Rate (ER) scheme. In Chapter 4, the proposed R-D model and rate adaptation technique of Chapter 3 is investigated for error prone channels using Unequal Error Protection (UXP). The UXP is based on Reed-Solomon (RS) encoding with erasure correction. In Chapter 5 a R-D model for SVC real-time video streams is proposed. The proposed model exploits SI and TI values GOP by GOP from the raw videos. The SI and TI values are then used to predict the rate of the video before encoding. The Proposed model is then used for multi-stream video delivery using the rate adaptation technique adopted in Chapter 3. The results of the proposed R-D model are compared with those obtained with R-D model in Chapter 3 by applying the rate adaptation technique. In Chapter 6 the interworking of the WiMAX and DiffServ heterogeneous network is described. In this proposed research work the Unsolicited Grant Service (UGS) from WiMAX network is mapped to the Expedited Forwarding (EF) service of DiffServ network. Priority Queuing (PQ) is applied inside the EF to deliver the delay sensitive traffic i.e. VoIP. The network is then tested on several congested scenario to test its efficiency to delays, jitters and packets drops with and without DiffServ network.

A part of work included in this thesis is published in [30] [46] [47] and [48] during my PhD.

Chapter 2

Overview of Scalable Video Coding

2.1 Introduction

Advances in video coding techniques and standardization along with the rapid development and improvements of network infrastructures, storage capacity and computing powers are enabling an increasing number of video applications. Applications area, today, range from MMS, video telephony and video conferencing over mobile TV etc. For these applications, a variety of video transmission and storage systems may be employed.

Due to the rapidly growing number of portable and non-portable devices, there is a strong need of a video standard that can be scaled according to the user and network needs. Because of all theses consideration, scalability and flexibility are key points for the near future of video services, whether these are new services or evolution of existing services. Such scalability is need not only on the architecture and infrastructure levels, but also at the content level [1].

Scalable Video Coding provides the appropriate tools to efficiently implement content scalability and portability. It is the latest scalable video-coding solution, and has been standardized recently as an amendment to the now well-known and widespread H.264/AVC standard [2] by the Joint Video Team (JVT).

In general, a video bit stream is called scalable when parts of the stream can be removed in a way that the resulting substream forms another valid bit stream for some target decoder, and the substream represents the source content with a reconstruction quality that is less than that of the complete original bit stream but is high when considering the lower quantity of remaining data. Bit streams that do not provide this property are referred to as single-layer bit streams [1]. Another benefit of SVC is that a scalable bit stream usually contains parts with different importance in terms of decoded video quality. This property in conjunction with unequal error protection is especially useful in any transmission scenario with unpredictable throughput variations and/or relatively high packet loss rates. By using a stronger protection of the more important information, error resilience with graceful degradation can be achieved up to a certain degree of transmission errors.

2.2 Concept of H.264/AVC Extension to H.264/SVC

In SVC encoding is performed once while it can be decoded multiple times to get the required bit stream as shown in figure 2.1. It states that the encoder has to encode once the bit stream which has details about spatial temporal and quality scalabilities. This scalable stream is then sent to the user and the user decode the stream according to its own requirement.

The principle of decoding is show in figure 2.2. As in Advance Video Coding, the encoding of the input video is performed at the Macro block basis. As the codec is based on the layer approach to enable spatial scalability, the encoder provides a down sampling filter stages that generates the lower resolution signal for each spatial layer. Encoder algorithm (not mention here in this thesis) may select between inter and intra coding for block shaped regions of each picture. 12 The video sequence is temporally decomposed into texture and motion information. Motion information from the lower layer may be used for prediction of the higher layer. The application of this prediction is switchable on a macro block or

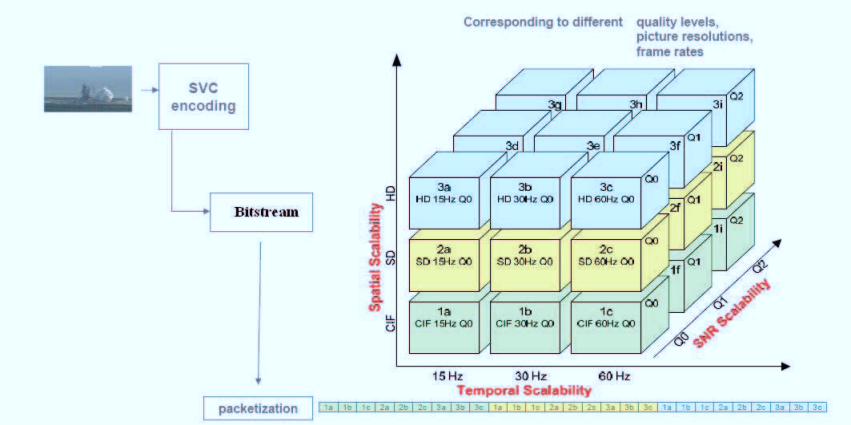


Figure 2.1: Principle of encoding.

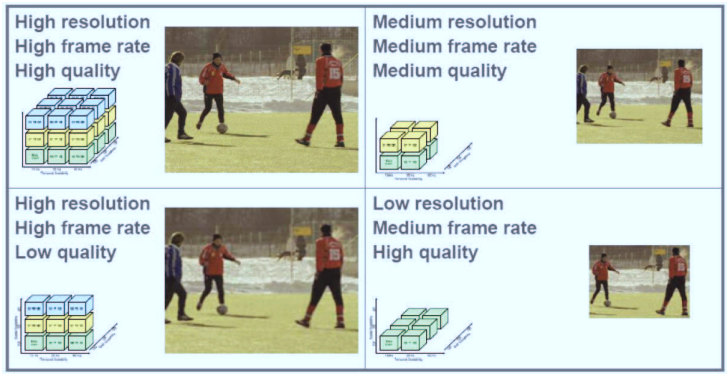


Figure 2.2: Principle of decoding.

block basis. In case of intra coding, a prediction from surrounding macro blocks or from co-located macro blocks of other layers is possible. These prediction techniques do not employ motion information and hence, are referred to as intra prediction techniques. Furthermore, residual data from lower layers can be employed for an efficient coding of the current layer. The redundancy between different layers is exploited by additional inter-layer prediction concepts that include prediction mechanisms for motion parameters as well as for texture data (intra and residual data).The residual signal resulting from intra or motion compensated inter prediction is transform coded using AVC features. Three kinds of prediction applied here are –Inter layer motion prediction, inter layer residual prediction

and Inter layer intra prediction. An important feature of the SVC design is that scalability is provided at a bit-stream level. Bit-streams for a reduced spatial and/or temporal resolution are simply obtained by discarding NAL units (or network packets) from a global SVC bit-stream that are not required for decoding the target resolution. NAL units of PR slices can additionally be truncated in order to further reduce the bit-rate and the associated reconstruction quality. Thus, one of the main design goals was that SVC should represent a straightforward extension of H.264/AVC. As much as possible, components of H.264/AVC are re-used, and new tools are only be added for efficiently supporting the required types of scalability. As for any other video coding standard, coding efficiency has always to be seen in connection with complexity in the design process.

2.3 Types of Scalabilities

Three scalability methods are possible in SVC, named temporal, spatial and SNR scalability, that allow to extract a sub-stream in order to meet a particular frame rate, resolution and quality, respectively. Each picture of a video sequence is coded and encapsulated into several Network Abstraction Layer Units (NALUs), which are packets with an integer number of bytes. Three key ID values, i.e. dependency id, temporal id, and quality id, are embedded in the header by means of the high level syntax elements, in order to identify spatial, temporal and quality layers.

2.3.1 Spatial scalability

For supporting spatial scalable coding, SVC follows the conventional approach of multiple-layer coding, which is also used in MPEG-2 Video / H.262, H.263, and MPEG-4 Visual. Each layer corresponds to a supported spatial resolution and is identified by a layer or dependency identifier D. The layer identifier D for the spatial base layer is equal to 0, and it is

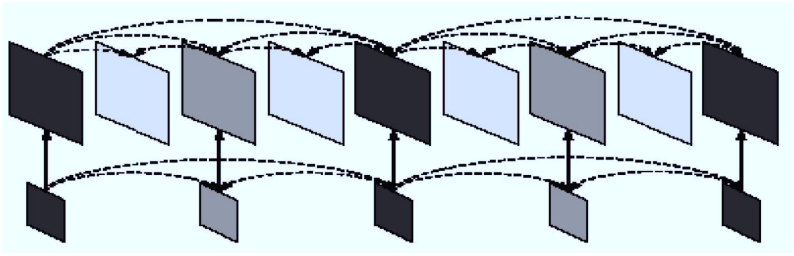
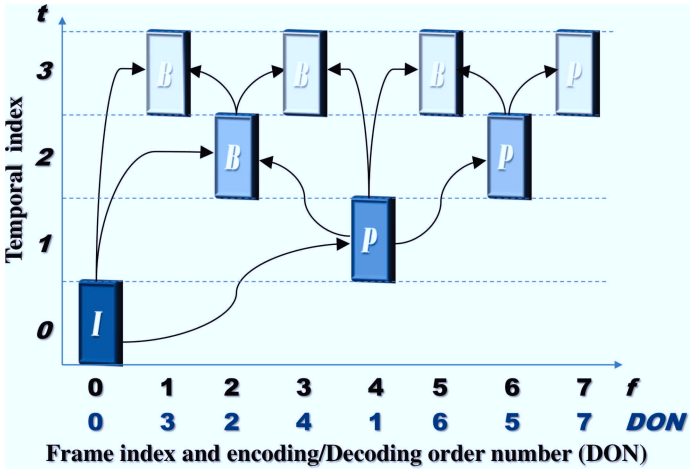


Figure 2.3: Multi-layer structure with additional inter-layer prediction for enabling spatial scalable coding.

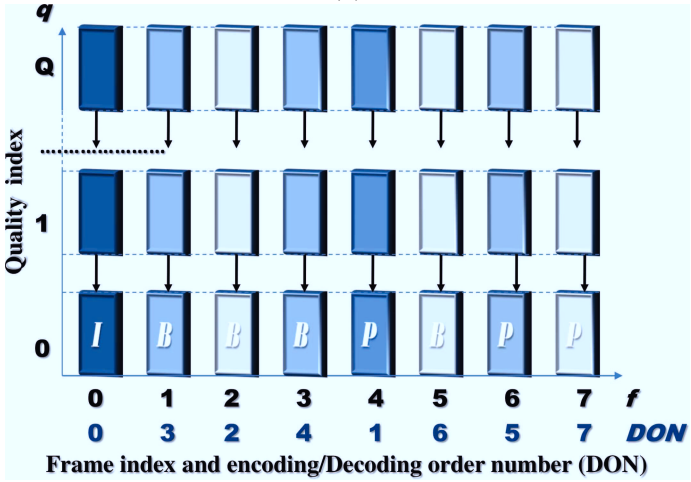
increased by 1 from one spatial layer to the next. In each layer, motion-compensated prediction and intra coding are employed as for single-layer coding. But in order to improve the coding efficiency in comparison to simulating different spatial resolutions, additional inter-layer prediction mechanisms are incorporated.. Although the basic concept for supporting spatial scalable coding is similar to that in prior video standards, SVC contains new tools that simultaneously improve the coding efficiency and reduced the decoder complexity overhead in relation to single-layer coding. In order to limit the memory requirements and decoder complexity, SVC requires that the coding order in base and enhancement layer is identical. All representations with different spatial resolutions for a time instant form an access unit and have to be transmitted successively in increasing order of their layer identifiers D. But lower layer pictures do not need to be present in all access units, which make it possible to combine temporal and spatial scalability as illustrated in Figure 2.6.

2.3.2 Temporal scalability

Temporal scalability can be achieved by means of the concept of hierarchical prediction. Each picture in one GOP is identified by a hierarchical temporal index or level $t \in \{0,1,\ldots,T\}$. The encoding/decoding process starts from the frame with the temporal index $t = 0$ that identifies a key-picture which must be intra-coded (I frame), in order to allow a GOP-based decoding. The remaining frames of one GOP are typically coded as



(a)



(b)

Figure 2.4: Enhancement temporal (a) and quality (b) layer prediction for a GOP of 8 frames.

P/B-pictures and predicted according to the hierarchical temporal index, thereby allowing to extract a particular frame rate. An implicit encoding/Decoding Order Number (DON) can be set up according to the temporal index and frame number of each frame.

In Figure 2.4(a) we show an example of the hierarchical prediction structure for a GOP with 8 pictures. The DON is obtained by ordering the pictures according to the temporal

index. If more than one frame have the same temporal level, the DON is assigned according to the picture index. Let us note that the last frame is encoded as P-frame in order to allow a GOP-based decoding, as mentioned before.

2.3.3 SNR scalability/Quality scalability

The SNR scalability allows to increase the quality of the video stream by introducing refinement layers. Two different possibilities are now available in SVC standard and implemented in the reference software [3], namely Coarse Grain Scalability (CGS) and Medium Grain Scalability (MGS). CGS can be achieved by coding quality refinements of a layer using a spatial ratio equal to 1 and inter-layer prediction. However, CGS scalability can only provide a small discrete set of extractable points equal to the number of coded layers. Here the focus is on MGS scalability which provides finer granularity with respect to CGS coding by dividing a quality enhancement layer into up to 16 MGS layers. MGS coding distributes the transform coefficients obtained from a macro-block by dividing them into multiple sets. The R-D relationship and its granularity depends on the number of MGS layers and the coefficient distribution, [4]. In [4] the authors analyzed the impact on performance of different numbers of MGS layers with different configurations used to distribute the transform coefficients. We also verified their results, by noting that more than five MGS layers reduce the R-D performance without giving a substantial increase in granularity. This is mainly due to the fragmentation overhead that increases with the number of MGS layers.

While extracting an MGS stream two possibilities are available in the reference software: a flat-quality extraction scheme, and a priority-based extraction scheme. The second scheme requires a post-encoding process, executed by an entity denoted as Priority Level Assigner, that computes a priority level for each NALU. It achieves higher granularity, as well as better R-D-performance [5]. The priority level ranges from 0 to 63, where 63 is intended for the base-layer, and is assigned to each NALU according to quality dependencies

and R-D improvement. Nevertheless, in order to exploit the temporal scalability at the decoder side, we re-assign different priority levels to the base-layer frames (those with $q = 0$), according to their temporal indexes, as specified afterwards. This feature is only exploited by the UXP profiler in subsection 4.2.2 and therefore does not change the 6-bit header of the packet which is necessary to perform the quality-based extraction. The R-D performance of the quality layers can be improved by using quality frames for motion compensation and introducing the concept of key-picture, which allows for a trade-off between drifting and coding efficiency. Nevertheless, this tool should not be applied in a rate-adaptation framework where all quality layers are often discarded by the rate adaptation module as exemplified in Figure 2.3(b).

2.3.3.1 Coarse Grain Scalability (CGS)

Coarse grain scalability (CGS) can be viewed as a special case of spatial scalability in H.264 SVC, in that similar encoding mechanisms are employed but the spatial resolution is kept constant. More specifically, similar to spatial scalability, CGS employs inter-layer prediction mechanisms, such as prediction of macroblock modes and associated motion parameters and prediction of the residue signal [1]. CGS differs from spatial scalability in that the up-sampling operations are not performed. In CGS, the residual texture signal in the enhancement layer is re-quantized with a quantization step size that is smaller than the quantization step size of the preceding CGS layer. SVC supports up to eight CGS layers, corresponding to eight quality extraction points [6], i.e., one base layer and up to seven enhancement layers.

2.3.3.2 Medium Grain Scalability (MGS)

While CGS provides quality scalability by dropping complete enhancement layers, MGS provides a finer granularity level of quality scalability by partitioning a given enhancement

layer into several MGS layers [1]. Individual MGS layers can then be dropped for quality (and bit rate) adaptation.

a) Splitting Transform Coefficients into MGS Layers: Medium grain scalability (MGS) splits a given enhancement layer of a given video frame into up to 16 MGS layers (also referred to as quality layers). In particular, MGS divides the transform coefficients, obtained through transform coding of a given macroblock, into multiple groups. Each group is assigned to a prescribed MGS layer.

b) Bit Rate Extraction: With MGS encoding, the video bit rate is adjusted by dropping enhancement layer NALUs, one at a time, until the target bit rate is achieved. No NALUs are dropped from the base layer.

2.3.3.3 Fine Grain Scalability (FGS)

In order to support fine-granular SNR scalability, so-called progressive refinement (PR) slices have been introduced. Each PR slice represents a refinement of the residual signal that corresponds to a bisection of the quantization step size (QP increase of 6). These signals are represented in a way that only a single inverse transform has to be performed for each transform block at the decoder side. The ordering of transform coefficient levels in PR slices allows the corresponding PR NAL units to be truncated at any arbitrary byte-aligned point, so that the quality of the SNR base layer can be refined in a fine-granular way. Figure 2.5 shows general concepts of Fine Granular Scalability in terms of layers.

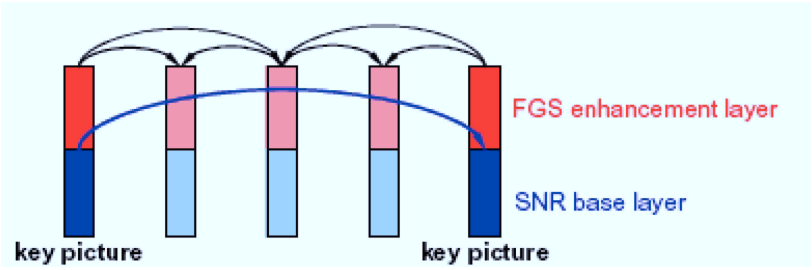


Figure 2.5: Fine granular scalability.

The main reason for the low performance of the FGS in MPEG-4 is that the motion compensated prediction (MCP) is always done in the SNR base layer. In the SVC design, the highest quality reference available is employed for the MCP of non-key pictures as depicted in Figure 2.5. Note that this difference significantly improves the coding efficiency without increasing the complexity when hierarchical prediction structures are used. The MCP for key pictures is done by only using the base layer representation of the reference pictures. Thus, the key pictures serve as resynchronization points, and the drift between encoder and decoder reconstruction is efficiently limited. In order to improve the FGS coding efficiency, especially for low-delay IPPP coding, leaky prediction concepts for the motion-compensated prediction of key pictures have been additionally incorporated in the SVC design.

2.4 Backward Compatibility

It is desirable in SVC scheme that a so called base layer be compatible with non Scalable video coding standards like AVC. It is also desired that additional scalable layers should be carried out in such a way that non-scalable video decoders, which have no knowledge of scalability, will ignore all scalable layers and only decode the base layer [7]. For these coded data that follow H.264/AVC and to ensure compatibility with existing H.264/AVC decoder, another new type of NAL (type 20) is used. This NAL carry the header information [8]. The base layer by design is compatible to H.264/AVC. During transmission, the associated prefix NAL units, which are introduced by SVC and when present are ignored by H.264/AVC decoders, may be encapsulated within the same RTP packet as the H.264/AVC VCL NAL units, or in a different RTP packet stream(when Multi session transmission mode is used) [9].

When using Multi session transmission mode-When a H.264/AVC compatible subset of the SVC base layer is transmitted in its own session in multi session transmission mode,

the packetization of RFC 3984 must be used, such that RFC 3984 of receivers can be part of multi transmission mode and receive only this session [10]. When using Single session transmission mode-When an H.264/AVC compatible subset of SVC base layer is transmitted using single session transmission, the packetization of RFC 3984 must be used, thus ensuring compatibility with RFC 3984 receivers [8].

Chapter 3

Rate Adaptation Using MGS for SVC

3.1 Introduction

H.264 Advanced Video coding (AVC) standard with scalable extension, also called Scalable Video Coding (SVC) [1], provides flexibility in rate adaptation by coding an original video sequence into a scalable stream. Three scalability methods are possible in SVC, named temporal, spatial and SNR scalability, that allow to extract a sub-stream in order to meet a particular frame rate, resolution and quality, respectively. Due to the dierent complexities of the scenes composing a video sequence, the relationships between the rate and the quality of a set of videos can be really different among them. If individual video streams are transmitted to different users in a broadcast dedicated channel, as for instance in the case of on-demand IP-TV services [11], an equal rate allocation can lead to unacceptable distortion of high-complexity videos with respect to low-complexity ones. Adaptive transmission strategies must be investigated to dynamically optimize the quality of experience (QoE) of each end-user.

In this chapter, we focus on rate adaptation, also called in literature statistical multiplexing, of SNR-scalable video streams, with a fixed temporal and spatial resolution. Many

contributions exist in the literature that provide rate adaptation exploiting the Fine Granularity Scalability (FGS) tool, e.g. [12],[13] and [14]. FGS coding allows to extract an arbitrary rate-distortion (R-D) point while maintaining the monotonic non-decreasing behavior of the R-D curves. Nevertheless FGS mode has been removed from SVC, due to its complexity.

Two different possibilities for the SNR scalability tool are now available in SVC standard and implemented in the reference software [3], namely Coarse Grain Scalability (CGS) and Medium Grain Scalability (MGS). CGS can be achieved by coding quality refinements of a layer using a spatial ratio equal to 1 and inter-layer prediction. However, CGS scalability can only provide a small discrete set of extractable points equal to the number of coded layers. MGS provides a ner granularity of quality scalability by dividing a CGS layer into up to 16 MGS layers. The granularity can be also improved if a post-processing quality layer (QL) insertion and a consequent quality-based extraction is performed with the aim to optimize the R-D performance [5]. With this tool MGS can be seen as alternative to the FGS coding.

The rest aim of this work is to analyze the performance of the MGS with QL and to provide a general R-D model. Other contributions exist in literature that estimate the R-D model for SNR-based scalable stream, with CGS and MGS, e.g. [15], [16], either analytical and semi-analytical. The analytical models are dependent on the probability distribution of discrete cosine transform (DCT) coefficients and often incur in a loss of accuracy. To achieve higher accuracy, semi-analytical R-D models are preferable. The semi-analytical models are based on parametrized functions that follow the shape of analytically derived functions, but are evaluated through curve-fitting from a subset of the rate-distortion empirical data points. In [16], the authors proposed an accurate semi-analytical square-root model for MGS coding and compared it with linear and semi-linear model. They concluded that the best performance is obtained by changing the model according to a parameter that

estimates the temporal complexity, evaluated before encoding the entire sequence. However, a general model, that is able to estimate the R-D relationship of a large range of video sequences, is necessary to perform analytical optimization of the rate-adaptation problem. Besides, they did not consider the post-processing QL insertion that produces a variation of the R-D performance.

In [17] the authors proposed a general semi-analytical rate-distortion model for video compression, also verified in [18] for SVC FGS layer, where the rate and the distortion have an inverse relationship. Three sequence-dependent parameters must be estimated through the knowledge of six empirical R-D points. We have also verified this model with reference to SNR scalability with MGS and QL. The high accuracy of the results led us to investigate a simplified model with lower complexity, where the number of R-D points can be reduced by eliminating one of the parameters to estimate. Thus, we propose and compare a simplified two-parameters semi-analytical rate-distortion model. This simplification has two main advantages: (i) only four empirical points are needed by the curve fitting algorithm to achieve good performance, (ii) it allows the derivation of a low-complexity optimal procedure to solve the multi-stream rate-adaptation problem, with a maximum number of iterations equal to the number of streams involved in the optimization.

This Chapter has the following main contributions: in section 3.2, a general optimization problem is formulated with the aim to provide the maximum quality to each video while minimizing their distortion difference, and by fulfilling the available bandwidth. In section 3.3 we analyze and verify two similar semi-analytical models for MGS with QL by comparing them with respect to complexity and the normally used goodness parameters: the root mean square error (RMSE) and the coefficient of determination R2 [19]. An optimum and computationally efficient procedure to solve the relaxed general problem is derived in section 3.4, with a discussion about complexity and optimality. Finally the numerical results, discussed in section 3.5, show (i) the goodness of our framework by looking at the error between the relaxed and discrete solutions, (ii) the performance improvement

with respect to a blind adaptation, and (iii) the complexity of the proposed algorithm with respect to a sub-optimal golden search algorithm proposed in literature.

3.2 General Problem Formulation for Multi-Stream Rate Adaptation

In general, the aim of multi-stream rate adaptation is to optimize a certain number of utility functions U_i with respect to a quality metric and according to rate constraints [20]. Before or after the encoding process the original high quality video must be adapted, to meet a particular QoE metric depending on spatial, temporal and SNR resolutions.

In this section we provide a general problem formulation for multi-stream rate adaptation. Let K be the number of streams involved in the optimization. Given a set of lossy compression techniques $1,...,N_k$, we can define in general $\mathscr{D}_k = d_{1,k},...,d_{N,k}$, $k = 1,...,K$ as the set of distortion values for the k-th stream. Let us note that its cardinality $|\mathscr{D}_k| = N_k$ is generally not equal for each video source, as in the case of high- flexibility SNR-based compression techniques.

The rate-distortion theory evaluates the minimum bit-rate R_k required to transmit the k-th stream with a given distortion $d_{n,k}$, by defining a function $\mathscr{F}_k$ that maps the distortion to the rate, i.e.

$$F_k : d_{n,k} \rightarrow \mathbb{R}^+$$

$$d_{n,k} \rightarrow R_k = \mathscr{F}_k(d_{n,k}) \tag{3.1}$$

One of the desirable properties of $\mathscr{F}_k$ is the strictly decreasing monotony, i.e.

$$\mathscr{F}_k(d_{i,k}) > \mathscr{F}_k(d_{j,k}), \quad d_{i,k}, d_{j,k} : d_{i,k} < d_{j,k} \tag{3.2}$$

When multiple streams have to be transmitted in a shared channel the rate adaptation algorithm must choose at each time slot and according to one optimization strategy, the best vector $D^* = [D_1, ..., D_K^*] \in \mathscr{D} = \mathscr{D}_1 \times ... \times \mathscr{D}_K$. $\mathscr{D}$ contains all the possible combinations of the elements of $\mathscr{D}_k$, $k = 1, ..., K$ and has cardinality $N = \prod_{k=1}^{K} N_k$.

The main purpose of multi-stream rate adaptation is to provide the minimum distortion, or equivalently the maximum rate according to assumption (2), to each video under a total bit-rate constraints R_c. However, the solution of such problem can generally lead to large distortion variations among different streams, due to the different complexity of video sources. Quality fairness is an important issue that must be addressed when multiple videos from different sources are transmitted in a shared channel. In [13] the authors have shown that, given a continuous decreasing exponential R-D relationship with a constant exponent equal for each source, the solution to the problem of minimizing the distortion variations is also the solution to the problem of minimizing the total average distortion. However, an exponential R-D relationship is not an accurate model for all the different video compression techniques, particularly for the SVC SNR scalable stream [13]. Thus, a general multi-objective problem has to be formulated and a continuous relaxation of the problem leads to some particular simplification under certain assumptions. The general objective of our proposed framework is to minimize the differences among the distortions provided to each video stream while maximizing the sum of the rates until a maximum bit-rate is met. As mentioned above, these two objectives alone can generally lead to different solutions.

Thus, we formulate the general problem as a multi-objective problem:

$$\min_{\boldsymbol{D} \in \mathscr{D}} \sum_i \sum_{j<i} \Delta(D_i, D_j) \tag{3.3}$$

$$\max_{\boldsymbol{D}\in\mathscr{D}} \sum_{k=1}^{K} \mathscr{F}_k(D_k) \tag{3.4}$$

$$s.t. \sum_{k=1}^{K} \mathscr{F}_k(D_k) \leq R_c \tag{3.5}$$

where

$$\Delta(D_i, D_j) = \begin{cases} 0 & \text{if (i,j)} \in \mathbb{X}_D \vee \text{(j, i)} \in \mathbb{X}_{\mathscr{D}} \\ |D_i - D_j| & \text{otherwise} \end{cases} \tag{3.6}$$

with

$$\mathbb{X}_{\mathscr{D}} = \{(i,j) \in \mathbb{Z}^2 : (D_i = D_{max,i} \wedge D_j > D_i) \vee (D_i = D_{min,i} \wedge D_j < D_i)\} \tag{3.7}$$

and $D_{min,i} = \min_n d_{n,i}, D_{max,i} = \max_n d_{n,i}$. The operators $\wedge$ and $\vee$ are the logic "AND" and "OR", respectively.

Ideal fairness among the distortion values assigned to the multiple video streams, i.e. $D_i = D_j, \quad \forall i \neq j$, is hard to be achieved. This fact is due to (i) the discretization of the R-D relationship and (ii) the presence of the minimum and the maximum distortion values for each source that are related to the complexity of each video and which can be very different. The definition of the fairness metric takes this fact into account. In fact, the difference among video distortions $\Delta(D_i, D_j)$ is slightly modified to take into account the minimum and the maximum constraints. It is worth noting that, under the assumption (3.2), this problem admits a feasible solution only if at least the sum of the minimum rates of the video sequences is supported by the transmission bandwidth R_c, i.e

$$\sum_{k=1}^{K} \mathscr{F}_k(D_{max,k}) \leq R_c \tag{3.8}$$

otherwise a certain number of videos are not admitted in the transmission until this constraint is not satisfied. The solution of the problem in (3.3)-(3.5) requires in general an exhaustive search in the space $\mathscr{D}$ of all possible vectors. If N becomes large the required complexity can be not suitable for real-time adaptation. On the other hand if N is small, i.e there are few video sources as well as few related R-D points, the problem solution can lead to a waste of the available bandwidth and a large distortion differences among multiple videos.

In the next section we will propose a semi-analytical R-D model with reference to the SNR scalability tool of SVC with MGS and QL layers [5]. This continuous model will allow us to apply a continuous relaxation to the optimization problem leading to a simplification in a single-objective problem formulation.

3.3 Rate Distortion Model for MGS with Quality Layer

We consider here SNR scalability obtained through the MGS coding and QL post-processing insertion, with a fixed temporal and spatial resolution. In this case the components of $\mathscr{D}_k$ are the distortion values of the extractable sub-streams from the high quality original encoded stream.

MGS coding allows to distribute the transform coefficients obtained from a macro-block by dividing them into multiple sets. The number of sets identifies the number of weights, often named MGS layers, in the MGS vector. Thus, the elements of the MGS vector correspond to the cardinality of each set.

The R-D relationship and its granularity depend on the number of MGS layers and the coefficient distribution [21], [4]. In [4] the authors analyzed the impact on performance of different numbers of MGS layers with different configurations used to distribute the transform coefficients. We also verified their results, by noting that more than five MGS layers reduce the R-D performance without giving a substantial increase in granularity.

This is mainly due to the fragmentation overhead that increases with the number of MGS layers.

While extracting an MGS stream two possibilities are available in the reference software: a flat-quality extraction scheme, and a QL-based extraction scheme. The second scheme requires a post-encoding process that computes a priority index for each NAL unit, but achieves higher granularity, as well as better R-D-performance [5]. However, differently to flat-quality extraction scheme, the quality-based extraction process does not give substantial variations in granularity and R-D performance when varying the distribution of the coefficients, as also shown in [15]. In our extensive simulation campaign the best results in terms of granularity and R-D performance are obtained with a MGS vector equal to [3 2 4 2 5].

When the SVC video has to be adaptively transmitted it is common practice to analyze the R-D model with respect to a xed set of frames identified by one group of pictures (GOP). In this way, the adaptation module can follow the complexity variations of the different scenes. Therefore, throughout this paper we assume that the reference time interval used to analyze the R-D relationship as well as to optimize the distortion of multiple streams is the GOP interval.

In [17] the authors propose a general continuous semi-analytical R-D model for video compression, also verified in [18] for SVC FGS layers, with the following relationship :

$$\mathscr{R}_k(D) = \frac{\eta_k}{D+\theta_k} + \phi_k \tag{3.9}$$

The distortion D is evaluated as the average mean square error (MSE) of the decoded video. The drawback of this approach is the need to estimate the three sequence/encoder dependents parameters, η_k, θ_k *and* ϕ_k, by using curve-fitting from a subset of the rate-distortion data points. The curve-fitting algorithm requires a relevant number of iterations and function evaluations and six empirical R-D points. To reduce the complexity, we can simplify this parametrized model by eliminating one parameter, i.e.

$$\mathscr{R}_k(D) = \frac{\alpha_k}{D} + \beta_k \tag{3.10}$$

In this case, only four R-D points need to be evaluated to estimate the two sequence-dependent parameters α_k *and* β_k, and as a result the number of iterations and function evaluations decreases. Beside the complexity reduction, this model allows a simple derivation of the solution of the problem (3.3)-(3.5), as we will show later.

Table 3.1 compares the goodness of the two models with respect the coefficient of determination R^2, the RMSE, the number of iterations and function evaluations required by a non-linear Least Square Trust-Region (LSTR) algorithm to converge. It can be noted how the number of function evaluations as well as the number of iterations decrease while a minimum loss occurs in the goodness parameter. In Figure 3.1, we plot the empirical R-D relationship for the five sequences, used to obtain numerical results, as well as their related R-D curves based on model (3.10). All of them are referred to the GOP with the worst RMSE value (the minimum in Table 3.1). We can also appreciate in this figure the achievable granularity of the quality-based extraction.

In the next section we will apply a continuous relaxation to the problem (3.3)-(3.5) by exploiting the model (3.10) and we will provide a low-complexity optimal procedure to solve it.

Video	Model	R2[min,max]	RMSE [min,max]	Av. No. iteration	Av. No. Function Evaluation
Coastguard	Model(10)	[0.9842 , 0.9934]	[37.895 , 79.992]	30.23	89.6
	Model (9)	[0.9956 , 0.9982]	[22.261 , 36.724]	34.7	155.9
Crew	Model(10)	[0.9752 , 0.9944]	[23.038 , 89.130]	30.9	94.2
	Model (9)	[0.9914 , 0.9972]	[20.019 , 52.489]	35.6	159.9
Football	Model(10)	[0.9662 , 0.9891]	[53.403 , 205.572]	29.0	89.5
	Model (9)	[0.9809 , 0.9993]	[12.940 , 99.810]	38.0	169.3
Foreman	Model(10)	[0.9669 , 0.9955]	[19.710 , 53.371]	25.7	73.2
	Model (9)	[0.9906 , 0.9980]	[13.516 , 33.745]	34.1	154.3
Harbour	Model(10)	[0.9854 , 0.9907]	[51.860 , 73.344]	37.5	129.8
	Model (9)	[0.9952 , 0.9991]	[18.883 , 44.822]	45.3	164.3

Table 3.1: Comparison between the two semi-analytical model in (3.9) and (3.10) with respect to the minimum and maximum RMSE and the coefficient of determination R2 evaluated for each GOP (GOP size equal to 16) of five video sequence with CIF resolution and frame rate of 30 fps. The video are encoded with one base layer (QP equal to 38) and two enhancement layers (QP equal to 32 and 26), both with 5 MGS layers and a weights vector equal to [3 2 4 2 5].

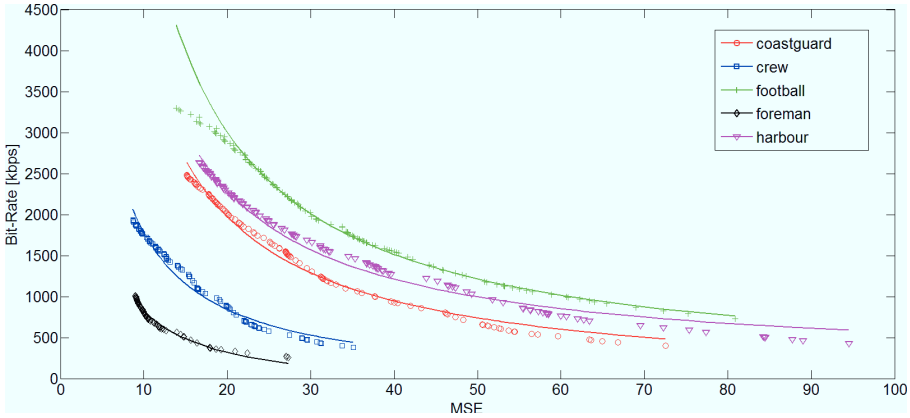


Figure 3.1: R-D Model (straight line), according to eq. (3.10) fitting the empirical R-D relationship for the GOP with the worst RMSE with reference to Table 3.1.

3.4 GOP-Based Multi-Stream Rate Adaptation Framework

Without loosing generality we assume that each video is coded with the same GOP size and the rate allocation is performed at the GOP boundaries. Thus, from now on we focus on one GOP interval. Considering all the discussions in the previous sections, we apply a continuous relaxation to the optimization problem based on the model (10). Therefore we assume that the discrete variable D_k becomes continuous (denoted by $\tilde{D}_k$), but limited by the minimum and maximum distortion, i.e.

$$\tilde{D}_k \in [D_{min,k}, D_{max,k}] \tag{3.11}$$

With reference to the SNR scalability, the points $\{D_{max,k}, \mathscr{F}_k(D_{k,max})\}$ and $\{D_{min,k}, \mathscr{F}_k(D_{k,min})\}$ are the base layer and the highest enhancement layer points, respectively. Those values are two of the four R-D points required by the curve-fitting algorithm.

It is worth noting that a trivial solution can be derived if the sum of the full quality encoded stream rates is less then or equal to the available bandwidth, that corresponds to transmit the entire encoded streams without adaptation. Thus, we analyze the non-trivial case where the following constraint holds :

$$\sum_{k=1}^{K} \mathscr{F}_k(D_{k,min}) > R_c \tag{3.12}$$

According to the continuous relaxation (3.11) and the assumptions (3.8) and (3.12), a feasible solution is obtained when the constraint on the overall channel bandwidth is active with equality. A single-objective problem, where the second objective, i.e (3.4) in the problem formulation, is eliminated and replaced by an equality constraints can be then formulated. Nevertheless, as a result of the relaxation of the problem, the two constraints referred to the maximum and minimum available rates of each stream must be added. They

imply that each video sequence has to obtain at least the base layer and not more than the maximum available bit-rate must be allocated to each video source to save bandwidth.

Thus, the relaxed problem can be formulated as

$$\min_{\tilde{\boldsymbol{D}}\in\mathbb{R}^K} \sum_i \sum_{j<i} \Delta(\tilde{D}_i, \tilde{D}_j) \tag{3.13}$$

$$s.t. \sum_{k=1}^{K} \mathscr{R}_k(\tilde{D}_k) = R_c \tag{3.14}$$

$$\mathscr{R}_k(\tilde{D}_k) \geq \mathscr{F}_k(D_{k,max}) \quad \forall k \tag{3.15}$$

$$\mathscr{R}_k(\tilde{D}_k) \leq \mathscr{F}_k(D_{k,min}) \quad \forall k \tag{3.16}$$

Note that the model $\mathscr{R}_k(\tilde{D}_k)$ replaces the actual R-D relationship $\mathscr{F}_k(D_k)$. In the next subsection we will derive an optimal procedure to solve this relaxed problem using methods that are computationally efficient and without the use of heuristics or brute-force search.

3.4.1 Problem Solution

A solution to the relaxed problem (3.13)-(3.16) can be derived by using sub-optimal procedures as the golden search algorithm proposed in [12] for a piecewise linear model. Nevertheless, the continuous formulation of model (3.10) allows us to derive a low-complexity optimal procedure, by noting that the solutions to the problem without the constraints (3.15) and (3.16) can be easily derived as follows:

$$\tilde{D}^* = \tilde{D}_k^* = \frac{\sum_{k=1}^{K} \alpha_k}{R_c - \sum_{k=1}^{K} \beta_k}, \quad \forall k \tag{3.17}$$

Since those constraints imply that a minimum (maximum) or a maximum (minimum) rate (distortion) has to be allocated to each video stream, these solutions can be improved successively through a simple iterative procedure.

Let $x_k, y_k \in \{0,1\}, k = 1,\ldots,K$, be binary variables that indicate whether or not the two constraints are active for the video stream k and will be updated during the procedure. We can then define:

$$A_{\boldsymbol{x},\boldsymbol{y}} = \sum_{k=1}^{K} x_k y_k \alpha_k \tag{3.18}$$

$$B_{\boldsymbol{x},\boldsymbol{y}} = \sum_{k=1}^{K} x_k y_k \beta_k \tag{3.19}$$

$$R^{av}_{\boldsymbol{x},\boldsymbol{y}} = R_c - \sum_{k=1}^{K} (1-x_k)\mathscr{F}_k(D_{k,max}) - \sum_{k=1}^{K} (1-y_k)\mathscr{F}_k(D_{k,min}) \tag{3.20}$$

where $R^{av}_{\boldsymbol{x},\boldsymbol{y}}$ is the available rate for the videos which have not active constraints. The iterative procedure works as follows:

1. Initialize : $x_k = 1$ *and* $y_k = 1 \quad \forall k = 1,\ldots,K$

2. For each $k : x_k \cdot y_k = 1$ *Compute* :

$\tilde{D}^*_k = \frac{A_{\boldsymbol{x},\boldsymbol{y}}}{R^{av}_{\boldsymbol{x},\boldsymbol{y}} - B_{\boldsymbol{x},\boldsymbol{y}}}$

$\tilde{R}^*_k = \mathscr{R}_k(\tilde{D}^*_k)$ based on model (3.10)

$condition = 0$

2a. **If** $\tilde{R}^*_k > \mathscr{F}_k(D_{k,min})$ then

$\tilde{R}^*_k = \mathscr{F}_k(D_{k,min})$

$\tilde{D}^*_k = D_{k,min}$

$y_k = 0$

$condition = 1$

2b. elseif $\tilde{R}^*_k < \mathscr{F}_k(D_{k,max})$

$\tilde{R}^*_k = \mathscr{R}_k(D_{k,max})$

$\tilde{D}^*_k = D_{k,max}$

$x_k = 0$

$condition = 1$

3. If $condition = 1$

Go to step 2

4. else break

The final relaxed solutions, given x_k and y_k, $k = 1, \ldots, K$, are then given by:

$$\tilde{R}_k^* = \begin{cases} \frac{\alpha_k}{\tilde{D}_k^*} + \beta_k & \text{if } x_k \cdot y_k = 1 \\ \mathscr{F}_k(D_{k,max}), & \text{if } x_k = 0 \\ \mathscr{F}_k(D_{k,min}), & \text{if } y_k = 0 \end{cases} \tag{3.21}$$

with

$$\tilde{D}_k^* = \begin{cases} \frac{A_{\boldsymbol{x},\boldsymbol{y}}}{R_{\boldsymbol{x},\boldsymbol{y}}^{av} - B_{\boldsymbol{x},\boldsymbol{y}}} & \text{if } x_k \cdot y_k = 1 \\ D_{k,max}, & \text{if } x_k = 0 \\ D_{k,min}, & \text{if } y_k = 0 \end{cases} \tag{3.22}$$

The algorithm requires in the worst case, a maximum of K iterations with $(K-1)/2$ rate and distortion evaluations. At the first iteration, due to the initialization, $\tilde{D}_k^*$ is computed as in (\ref{primal-solution}). At each iteration the algorithm checks if the related rate solutions violate one of the constraints (3.15), (3.16). If it happens for one video, the algorithm assigns the relative minimum or maximum rate to this particular video and re-evaluates the distortion for the other video streams.

The optimality of the solutions (3.21) and (3.22) can be easily proved, by noting that the sum of the difference functions in (3.13) is always kept to zero, i.e. $\sum_i \sum_{j<i} \Delta(\tilde{D}_i^*, \tilde{D}_j^*) = 0$ and the sum of the rates is always equal to the available bandwidth. In fact, if at the n-th iteration a maximum rate constraint (condition of step 2a) is violated for the i-th video, the distortion of the other videos at the next iteration, $\tilde{D}_k^*[n+1]$, will decrease, i.e.

$$\tilde{D}_k^*[n+1] < \tilde{D}_k^*[n] < D_{i,min}, \quad \forall k \neq i : x_k[n+1] \cdot y_k[n+1] = 1, y_i[n] = 0 \tag{3.23}$$

Vice versa, when the second constraint (condition of step 2b) is violated for the j-th video the distortion $\tilde{D}_k^*[n+1]$ of the other video will increase, i.e.

$$\tilde{D}_k^*[n+1] > \tilde{D}_k^*[n] > D_{j,max}, \quad \forall k \neq j : x_k[n+1] \cdot y_k[n+1] = 1, x_j[n] = 0 \tag{3.24}$$

For all other videos with $x_k \cdot y_k = 1$ the solutions are left untouched, as shown in (3.22). The inequalities (3.23) and (3.24) follow from the monotony property of the R-D function.

Let us finally note that the conditions of steps 2a and 2b are auto-exclusive for each video source if

$$D_{s,max} > D_{p,min}, \quad \forall s \neq p, \quad s, p = 1, \ldots, K \tag{3.25}$$

When two or more video streams have a very different scene complexity in the same GOP, the inequality (3.25) may not be verified and the evaluated distortion $\tilde{D}_k^*$ may fall inside the interval $[D_{s,max}, D_{p,min}]$. In this particular case, to assure the best fairness, the algorithm would require some temporary additional steps to evaluate which constraints has to be applied first, which leads to a small increase in the complexity. In order to keep the complexity low we propose for this case to prioritize the distortion minimization. Thus, we first apply the constraints on the maximum rate (step 2a) by assigning the minimum distortion $D_{p,min}$ to the p-th video. At the next iteration, the distortion will decrease, due to the convexity of the R-D functions. If the distortion decreases in such way that the evaluated rate of the$s$-th video do not violate its maximum distortion constraint, the algorithm will be able to assign a lower distortion to it. Let us note that this choice does not compromise the optimality of the solution of the problem according to eq. (3.6).

From a mathematical perspective the optimal discrete solution $\boldsymbol{D}^*$, starting from the relaxed one $\tilde{\boldsymbol{D}}^*$, should be derived by applying optimization techniques, e.g. branch & bound search. Nevertheless, such techniques require the knowledge of all the empirical discrete R-D points or a subset of R-D points close to the relaxed optimum solutions, with an increase in complexity. To keep the complexity low, it is common practice to extract the higher discrete bit-rate under the optimal relaxed solution, by paying a minimum waste of bandwidth due to the granularity of the empirical R-D relationship.

3.5 Numerical Results

In this section we evaluate the performance of the proposed rate adaptation framework by using the JSVM reference software [3]. We encode five video sequences with different scene complexity, i.e. *coastguard*, *crew*, *football*, *foreman*, *harbour* in CIF resolution with a frame rate of 30 fps. The SNR-scalability is obtained through 2 enhancement layers, each one split in 5 MGS layers with vector distribution [3 2 4 2 5]. The quantization parameter (QP) of the base and enhancement layers are equally spaced and set to 38, 32 and 26, respectively. Each sequence is coded GOP-by-GOP with a GOP size equal to 16, and the post-processing quality-based process is then applied, as mentioned throughout the paper. We first provide the performance metrics for a particular case of bandwidth, i.e. $R_c = 3000$ kbps, then we study the impact of different R_c values. The fairness is evaluated through two metrics: the average MSE difference $\delta_{av} = (1/S)\sum_i \sum_{j<i} |D_i^* - D_j^*|$, where the average is computed with respect to the number $S = K(K-1)/2$ of terms in the sum, and the most used MSE variance for each GOP. We first compare the solution of our algorithm (OPT) with an equal-rate (ER) scheme where no adaptation is performed, i.e. the same proportion of the available bandwidth is assigned to each video. To have a fair comparison we apply to ER scheme the constraints (3.15) and (3.16) in order to guarantee the base-layer to each video and to fulfill the available bandwidth. Therefore, after sorting the

streams in two vectors into decreasing order according to base-layer bit-rate and into increasing order according to highest layer bit-rate, respectively, we iteratively check if the bit-rate $R_k = R_c/K$ required by each ordered stream violates one of those constraints. If it happens, we assign the corresponding bit-rate and equally re-distribute the remaining bandwidth to the other streams. Table 3.2 shows the average MSE resulting from the rate assigned to each video sequences for the first 15 GOPs. As expected, the ER scheme is able to provide less distortion to the low-complexity video, i.e. *crew*, *foreman*, by compromising the distortion of the video sequences with more complexity. Our algorithm, while providing fairness, is able to improve the performance of the complex videos, by allocating more bits to video with more complex scenes. This is more clear in figure 3.2 where we plot the rate assigned to each video sequence GOP-by-GOP. More bit-rate is assigned to *coastguard*, *football and harbour* video sequences, allowing them to achieve more quality. In Table 3.3, we show the improvements of our proposed schemes with respect to ER. The average MSE difference is significantly reduced and equivalently the variance is decreased up to ten times. However, in this particular case of bandwidth, the MSE difference (variance) is still quite high, due to the minimum rate constraints. The average modified MSE difference $\Delta_{av} = (1/S)\sum_i \sum_{j<i} \Delta(D_i^*, D_j^*)$ according to definition in (3.6), is also evaluated in Table 3.3 . Let us note that this metric also give us the information of the error generated when the discrete solution replaces the continuous solution of the relaxed problem, whose Δ_{av} is zero. This error includes two contributions: the estimation error of the model and the integrality gap. As expected the average error is not small due to mainly the low granularity of the low-rate points.

In figure 3.3, the MSE variance averaged over 15 GOPs is evaluated for different bandwidths. In the bandwidth interval considered, the assumptions (3.8) and (3.12) hold for each GOP. When the bandwidth is very low the two schemes provide approximately the same MSE because the optimization range is limited by the minimum rate constraints. When the bandwidth increases, our procedure improves the fairness leading the variance close to 0.

GOP Index	Coastguard		Crew		Football		Foreman		Harbour	
	ER	OPT	ER	OPT	ER	OPT	ER	OPT	ER	OPT
1	53.71	53.71	18.59	34.64	80.86	55.87	18.40	31.66	74.28	55.52
2	57.35	54.57	19.79	37.85	74.65	59.56	18.24	29.96	81.23	56.98
3	69.45	54.63	23.52	38.67	64.02	54.06	24.63	29.99	94.54	58.27
4	81.35	59.02	39.87	39.87	63.69	56.29	17.75	33.34	75.92	57.75
5	53.71	47.36	24.89	41.67	49.53	43.55	17.73	31.58	71.93	50.97
6	55.16	41.70	28.22	38.26	16.85	24.55	19.51	34.00	73.82	46.48
7	49.11	42.22	39.87	44.31	20.23	31.36	12.40	27.35	82.14	49.31
8	49.38	42.64	33.87	38.57	31.49	39.12	14.21	28.35	73.47	48.10
9	45.79	44.11	37.47	41.71	43.89	44.20	19.20	36.12	73.51	50.37
10	42.02	46.06	42.85	43.02	47.94	45.19	19.51	32.24	69.64	52.51
11	44.49	49.17	34.40	45.68	59.81	48.88	17.77	31.33	67.82	53.78
12	42.07	40.36	25.56	39.42	41.44	41.17	18.73	30.32	71.87	46.47
13	40.17	43.18	27.09	41.48	50.24	43.84	16.55	27.87	72.23	50.91
14	42.11	56.76	23.86	35.08	82.50	56.45	25.39	45.48	68.08	57.95
15	38.29	60.28	24.81	38.76	84.63	56.84	25.92	57.12	69.48 2	55.8
Av.	50.95	49.05	29.64	39.93	54.12	46.73	19.06	33.78	74.66	52.74

Table 3.2: Average MSE of each video sequence with equal-rate (ER) assignment and rate adaptation with the proposed algorithm (OPT). Total bandwidth is equal to 3000 kbps.

A slight variance increase occurs at large bandwidths when the maximum rate constraints limit the achievable distortion. On the other hand the ER scheme generally increases the MSE variance until the base-layer constraints are active for most of the streams. This behavior can be partially reduced by controlling the base-layer bit-rate [22] to each video according to their complexity as performed for instance in [12].

To further assess our proposed scheme, we compared it to the golden search algorithm (GSA) proposed in [12], to solve the problem (3.13)-(3.16). This algorithm can be seen as a suboptimal version of our procedure. The initial solution is computed as function of the golden-section value and the difference between the lower and higher bounds, i.e. $a = \min_k D_{k,min}$ and $b = \max_k D_{k,max}$, identified by the minimum and the maximum distortion among the videos. At each iteration the solution is updated by applying the per-video constraints and by compressing the search interval consequently. The GSA terminates when the difference between the sum of the assigned rates and the available bandwidth is less of a chosen value ε. Nevertheless, an additional termination condition must be introduced to assure the convergence of the algorithm, that is usually indicated by the tolerance τ, i.e. $|a-b| \leq \tau$. In order to provide a fair comparison we set $\varepsilon = 0.0002R_c$, and $\tau = 0.01$, leading to a sub-optimality error under 0.5% over all the investigated cases. In figure 3.4 the plot shows average number of iterations required by the two algorithms for different bandwidths. The number of iterations of our algorithm is limited by the number of video sequences, as mentioned in sub-section 3.4.1, and decreases away from the minimum and the maximum bandwidths obtained as the sum of minimum and maximum rates of each video. The GSA algorithm requires generally more iterations due to the sub-optimal choice of the starting-point. This result does not change by increasing the number of videos involved in the optimization, as we also verified.

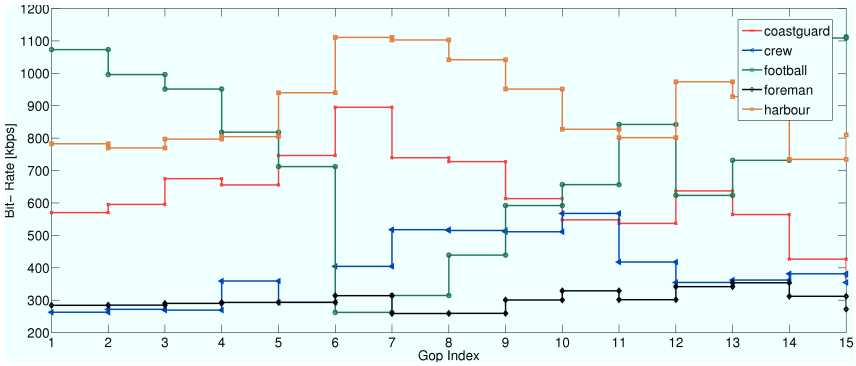


Figure 3.2: Rate assigned by our adaptation algorithm in each GOP, with bandwidth equal to 3000 kbps.

GOP Index	Δ_{av}		δ_{av}		Variance	
	ER	OPT	ER	OPT	ER	OPT
1	36.12	0.43	36.12	13.86	884.40	145.41
2	35.51	1.00	36.17	15.67	889.50	171.43
3	33.78	0.84	37.37	14.50	941.76	148.35
4	19.53	0.55	32.65	13.85	705.43	139.62
5	24.79	1.48	27.44	8.89	489.84	53.75
6	29.92	1.64	29.92	10.31	614.97	69.38
7	33.67	1.42	33.67	11.37	752.18	84.72
8	27.28	2.21	27.28	8.72	495.50	52.27
9	23.39	2.01	23.39	6.20	382.93	26.39
10	21.24	1.93	21.24	8.72	319.33	54.28
11	24.30	1.46	25.10	9.68	398.50	73.46
12	24.56	1.22	24.56	6.81	420.64	34.09
13	26.90	1.54	26.90	9.69	463.11	70.69
14	32.00	0.30	32.00	11.40	680.44	98.23
15	32.64	1.05	32.64	8.87	730.21	73.16
Av.	28.37	1.21	29.76	10.57	611.25	86.35

Table 3.3: Average modified MSE difference Δ_{av}, average MSE difference δ_{av} and MSE variance in each GOP interval. Comparison between the proposed algorithm (OPT) and equal-rate (ER) assignment with bandwidth equal to 3000 kbps.

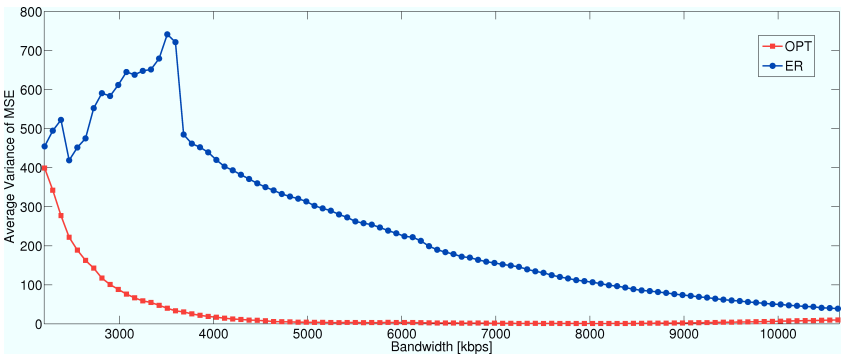


Figure 3.3: Variance of the MSE averaged over 15 GOPs, with different bandwidth values. Comparison between the proposed algorithm (OPT) and equal-rate (ER) assignment.

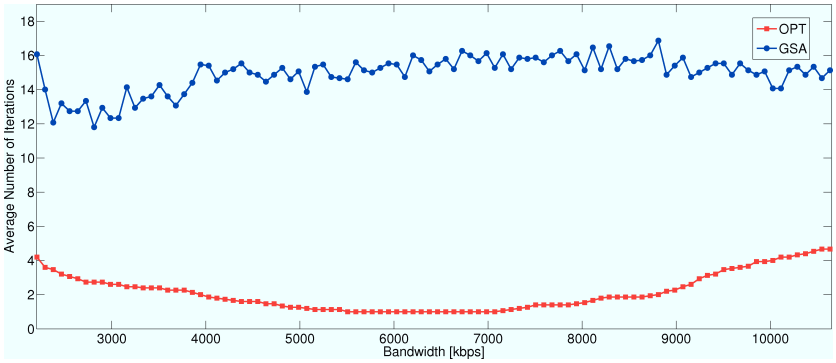


Figure 3.4: Average number of iterations required by our adaptation algorithm (OPT) and golden search algorithm (GSA) to converge.

3.6 Conclusions

In this work we proposed a multi-stream rate adaptation framework with reference to SNR-scalability of SVC with MGS and QL. We formulated a general discrete problem with the aim to minimize the average distortion while providing fairness to different video sources. Two similar semi-analytical model that estimate the R-D relationship of each video source GOP-by-GOP are evaluated and compared with respect to goodness parameters and complexity.

The general discrete problem was then relaxed and an optimal procedure was derived

based on a low-complexity model. In the numerical results we showed the feasibility of our framework by analyzing the gap between the relaxed and discrete solution according to fairness metrics, the improvements with respect to an equal-rate scheme and the lower complexity of the proposed procedure with respect to an existing algorithm in the literature.

Chapter 4

Rate Adaptation for Error Prone Channels in SVC

4.1 Introduction

The high data-rate resulting from the actual and the next generation systems is enabling the providers to support several video services, as, for instance, video-on demand, IP-TV and real-time streaming. A high degree of flexibility and adaptivity is required from the video delivery system to meet different levels of quality requirements depending on the different characteristics of end-user devices and access networks. This is made possible by encoding video sequences with encoders that support multiple layers or bit-streams that can be sequentially dropped providing a graceful degradation. H.264 Advanced Video coding (AVC) standard with scalable extension, also called Scalable Video Coding (SVC) [1], allows flexibility in rate adaptation by encoding an original video sequence into a scalable stream.

Due to the different complexities of the scenes composing a video sequence, the relationships between the rate and the quality can be really different within a set of videos. If

individual video streams are transmitted to different users in a broadcast dedicated channel, an equal rate allocation could lead to unacceptable distortion of high-complexity videos with respect to low-complexity ones. Adaptive transmission strategies have to be investigated to dynamically optimize the quality of experience (QoE) of each end-user.

Beside the distortion due to lossy encoding process, the quality of each video can be heavily reduced due to the transmission errors and the consequent loss of part of the video stream. The automatic repeat-request (ARQ) schemes have the main drawback to increase the delay and can not be suitable for many application where the playback time is a stringent constraint. Within the framework of video delivery schemes based on SVC, Forward Error Correction (FEC) has been proposed to recover channel errors and many contributions in the literature have proved its effectiveness[23], [24], [25].

In this chapter we analyze a scenario that can cover different video applications. The unique assumption is that the multimedia provider is able to perform o-line some computation-expensive processes, such as encoding and quality-computation for each video. In this framework, applications like video on-demand [26], IP-TV [11], sport broadcasting, where an initial transmission delay in the order of seconds can be tolerated by the end-users, as well as real-time streaming [27], are well suited to the low-complexity transmission scheme proposed. Each one of these applications requires a multimedia provider that has to serve several end-users which request different video sources. Thus, we suppose that the lower-layers dedicate a shared constant bandwidth to a particular set of users, and inform the application layer about channel conditions, in terms of packet losses. In this scenario quality fairness is an important issue that must be addressed. In fact, the end-user expectation is to receive the best feasible quality independently of the particular video complexity. In this light, the adaptation module of the media provider is required to extract from the original video sequences a set of scaled streams with a fair assignment of expected end-user quality, even in presence of packet losses.

In this work, we focus on rate adaptation of temporal and quality scalable video streams

transmitted with a fixed spatial resolution over an error-prone channel. Nevertheless, the entire framework can be extended to spatially scalable streams. In Figure 4.1 shows the architecture of the video delivery system. Each video sequence is encoded by the SVC encoder to fully support temporal and quality scalability. The resulting streams are encapsulated into Network Abstraction layer Units (NALUs), which are packets of an integer number of bytes, and stored in a media server. The NALUs have different importance according to a certain coding paradigm. To support the features of both Adaptation module and Unequal Erasure Protection (UXP) profiler, the video streams are also processed with the aim of extracting the information on the quality of each stream. After the encoder, the priority level assigner evaluates a priority index for each NALU, by considering the Rate-Distortion (R-D) relationship and the dependency on the other NALUs. Such information is encapsulated in the NALU header and then exploited by both the UXP profiler and the Adaptation module. These two processes are executed off-line.

The UXP profiler aims at determining for each NALU the level of protection against transmission losses, which is obtained by adding parity bytes according to a specified UXP strategy. We assume, in the case investigated here, that the UXP is based on the use of Reed-Solomon (RS) encoding with erasure correction. This task is executed by taking into account the estimated packet-loss rate of the lower layers which can be supplied at regular intervals. The protection profile is then sent to the Adaptation module which first estimates the expected R-D relationship, then extracts a suitable bit-stream from each video stream to meet fairness and bandwidth constraints. Each outcoming bit-stream is then encoded by the RS encoder. Finally, the resulting codewords are encapsulated in a transmission block and interleaved over RTP packets which are forwarded to the lower layers.

4.1.1 Related Works

One of the aims of this paper is to analyze the performance of the SVC encoder and to provide a general R-D model. Other contributions exist in literature that estimate the R-D

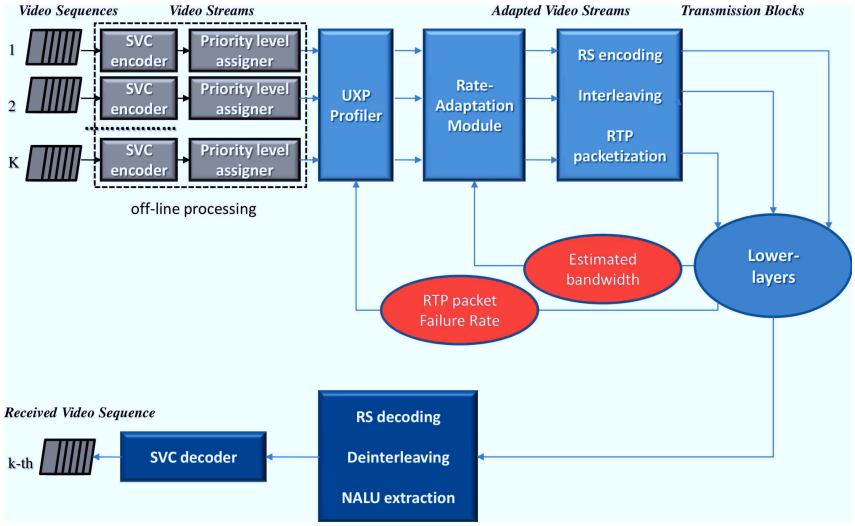


Figure 4.1: System architecture. Each sequence is encoded to fully support temporal and quality scalability and a priority level is assigned to the NALUs. The UXP profiler evaluates the overhead required according to a certain protection policy and RTP packet failure rate, and provides R-D information to the Adaptation module. The Adaptation module extracts sub-streams according to the estimated bandwidth and sends the data bytes to the RS encoder. The resulting codewords are then encapsulated in a transmission block, interleaved in RTP packets and forwarded to the lower layers. The receiver performs the inverse operations (RS decoding and deinterleaving) in order to extract the NALUs which are sent to the SVC decoder.

model for SNR based scalable stream, e.g. [15], [16], either analytical or semi-analytical. The analytical models are dependent on the probability distribution of discrete cosine transform (DCT) coefficients and often incur in a loss of accuracy. To achieve higher accuracy, semi-analytical R-D models are preferable. The semi-analytical models are based on parametrized functions that follow the shape of analytically derived functions, but are evaluated through curve-fitting from a subset of the RD empirical data points. In [16], the authors proposed an accurate semi-analytical square-root model for MGS coding and compared it with linear and semi-linear models. They concluded that the best performance is obtained by changing the model according to a parameter that estimates the temporal complexity, evaluated before encoding the entire sequence. However, a general model for the

estimation of the R-D relationship for a large set of video sequences, is necessary to derive analytical solutions for the rate-adaptation problem. In [17] the authors proposed a general semi-analytical R-D model for video compression, also verified in [24] for SVC FGS layer, where the relationship between rate and distortion depends on three sequence-dependent parameters which must be estimated through the evaluation of six empirical R-D points. We have verified this model with reference to SNR scalability with MGS and the high accuracy of the results led us to investigate a simplified two-parameters model with lower complexity, where the number of R-D points needed to estimate the parameters is reduced. Many contributions exist in the literature that consider fairness-oriented rate adaptation, but they exploit the Fine Granularity Scalability (FGS) tool, e.g. [12], [13], [14]. Nevertheless, FGS mode has been removed from SVC, due to its complexity, and these works do not take into account the effects of transmission losses.

Cross-layer optimization of video streaming over packet-erasure channel is also highly investigated, within the framework of SVC [24], [25], [28]. In [24] and in earlier works the authors proposed a complete framework to analyze and model the video streaming system over packet erasure channel, also in presence of play-out deadline. They derived an analytical model to estimate the the R-D in case of base-layer packet losses, while using a semi analytical model for the quality layers. An UXP profiler, based on the same priority level assigner used in our work, solves a rate-minimizing cost functions. However, the rate adaptation aims at minimizing the distortion of each video without taking into account fairness issues. Maani et al. [25] proposed a model to solve the problem of joint bit extraction and channel rate allocation over packet erasure channels.

This Chapter is organized as follows. Section 4.2 we discuss the transmission of SVC streams over erasure-packet channel. In Section 4.3 we analyze semi-analytical R-D model for erasure-packet channel cases. Performance assessment in the case of transmission over packet erasure channel is illustrated in Section 4.4. Finally we present our conclusions in Section 4.5.

4.2 Unequal Erasure Protection

Due to the different importance and the temporal/quality dependency of the different frames, Unequal Erasure Protection (UXP) schemes can generally overcome schemes based on equal protection. In our work, we follow the guidelines presented and discussed in [23] for SVC transmission over packet-erasure channel, by focusing our attention on a GOP-based transmission. Each GOP is mapped into one Transmission Sub-Block (TSB) that carries either data and parity bytes, as exemplified in Figure 4.2. Each row of the TSB identifies a RS (n,m) codeword where m is number of data bytes and n is the total bytes of the codeword. If a packet-erasure detection is available at the lower-layers, the RS codes are able to correct up to $n-m$ bytes, equal to the number of parity bytes. The aim of the UXP profiler is to assign a different protection to each frame according to its dependencies and R-D improvements.

A first step is to order the NALUs according to their protection class. As mentioned before, a priority-index greater than 62 is re-assigned to the different temporal base layer frames ($q = 0$), to have lower priority indexes for high temporal indexes. Thus, all the frames are sorted according to the priority level p and sequentially inserted into one TSB, according a given UXP profile $\boldsymbol{M}^* = \{m^*_{f,q,p}\}$, where $m^*_{f,q,p}$ identify the protection class assigned to frame with frame index f, quality index q and priority level p.

Finally, one or more TSB are placed into a transmission block (TB) whose columns become the payload of RTP packets. In this way the RS codewords are interleaved over the different RTP packets. Therefore, RTP packet errors (or erasures) can be assumed as uniformly distributed inside the codewords. In order to reduce the overhead due to the need of padding for compensating the different NALU lengths, the part of the codeword left unused by a given NALU is filled with the data from the subsequent NALU. For simplicity of presentation and without loosing generality, we assume that the size $S_{f,q,p}$ of each NALU is always greater than or equal to the total size n of the RS code:

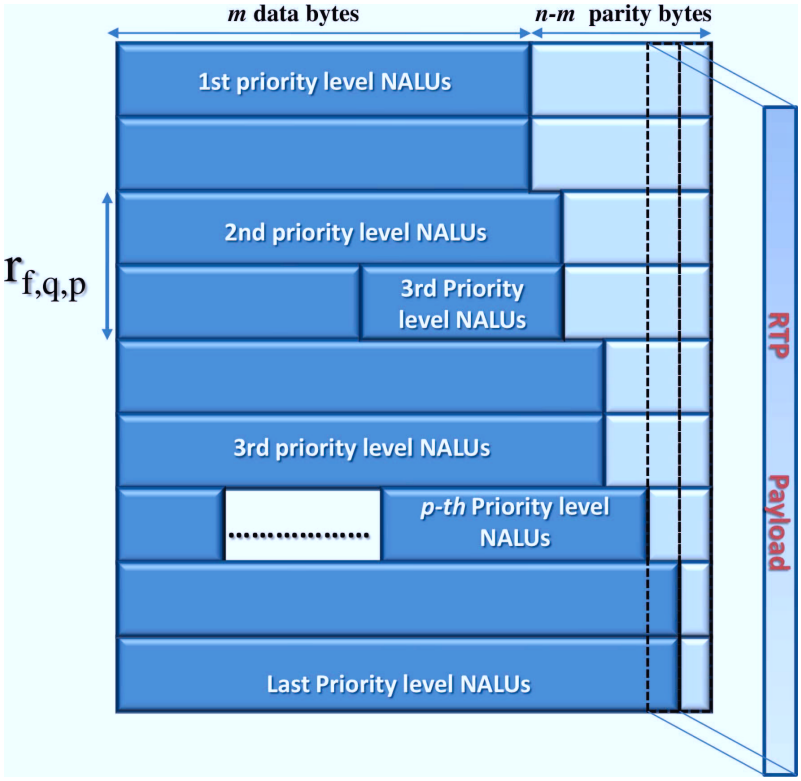


Figure 4.2: Transmission Sub-Block (TSB) structure. Following the priority level, the NALUs of one GOP are placed into one TSB according to a given UXP profile (protection class) from upper left to lower right. The columns of one ore more TSB are then encapsulated into RTP packets

$$S_{f,q,p} \geq n \tag{4.1}$$

This assumption ensure that each TSB row contains no more than two different frames. Let us finally note that a Multi Time Aggregation Packet (MTAP) header must be inserted before each priority level NALU in order to deliver the decoding order number (DON) and timing information assignment.

4.2.1 Frame error probability and expected distortion

Let assume that the RTP packet error rate information, $P_{e,rtp}$, is periodically collected from the lower-layers, as shown in Figure 4.1. According to the proposed UXP scheme a closed formulation of the expected error probability can be derived by using the failure probability of a single (n,m) RS codeword:

$$P(n,m) = \sum_{i=m-n+1}^{n} \binom{n}{i} P_{e,rtp}^{i} (1 - P_{e,rtp})^{n-i} \tag{4.2}$$

The individual frame error probability now depends on the number of TB rows associated to each frame $r_{f,q,p} = \left\lceil \frac{S_{f,q,p}}{m^*_{f,q,p}} \right\rceil$, and on whether or not some bytes of the frame are inserted in the row using the protection class of the preceding priority level. Let $z \in \{0,1\}$ be a boolean variable that indicates whether or not this last event occurs. The frame error probability FEP is then computed as one minus the probability that all codewords of the TB, associated to the frame, can be correctly decoded by the RS decoder:

$$FEP_{f,q} = 1 - \left[\Big(1 - P(m^*_{f,q,p}, n)\Big)^{r_{f,q,p}} \Big(1 - P(m^*_{f,q,p-1}, n)\Big)^{z} \right] \tag{4.3}$$

According to the derived FEP, a closed formula for the expected distortion can be now computed. Let $\Upsilon D_{f,q} = |d_{f,q} - d_{f,q-1}|$ be the quality improvement resulting from the correct decoding of the f-th frame with quality id q, which is computed by the priority level assigner. In order to compute the quality improvement $\Upsilon D_{f,0}$ due to the enhancement (temporal) frames of the base layer we assume an error concealment (EC) method based on the picture copy (PC). Therefore the distortion increment due to the loss of an enhancement picture is computed by considering the difference between the enhancement frame and the copy of the previous one. The expected distortion due to the loss of frames with quality index $q \leq Q$ can be computed as:

$$d_{f,q,loss} = \sum_{r=0}^{q} \Upsilon D_{f,r} \left[FEP_{f,0} u_{f-1} + \sum_{j=1}^{q} FEP_{f,j} \prod_{s}^{j-1} \Big(1 - FEP_{f,s}\Big) \right] \tag{4.4}$$

where u_x is the Heaviside function. The first term of the sum takes into account the distortion due to the loss of a temporal enhancement layer. Since a loss of the I-frame will results in a infinite distortion we assume here that the associated NALUs will receive enough protection to have $FEP_{0,0}$ close to zero.

The second sum, on the other hand, takes into account the cumulative probability that the $j-1$ quality layers have been successfully received but the j-th quality frame is lost, where $j \leq q$. Finally, the total expected distortion of the entire GOP is the sum of the individual frame loss distortions:

$$d_{s,loss} = \sum_{f=0}^{G-1} d_{f,q,loss} \tag{4.5}$$

Let us note that the number of quality layers of each frame in one GOP can be different after the rate adaptation. Thus, the index s maps the vector whose elements are the resulting number of the quality-layer of each frame f: its range is from 0 to GQ. The values of the expected distortion can be finally used, together with the required rate, to reshape the R-D relationship according to the values of the FEP.

4.2.2 Proposed UXP profiler

The derivation of an optimal UXP profile is hard to achieve. It should be computed according to the solutions of an optimization problem aimed at balancing the trade-off between protection and overhead. This it is a discrete problem since the FEP, as well as the overhead resulting from the RS encoding, strictly depends on the discrete variable m, as shown in Figure 4.3. In order to guarantee a rate distortion relationship strictly decreasing, the FEP of each frame should increase as the quality and the temporal indexes increase. However, due to the granularity of the available values of m, sometimes this condition is not met. This problem could be partially solved by a joint optimization of the encoding process and

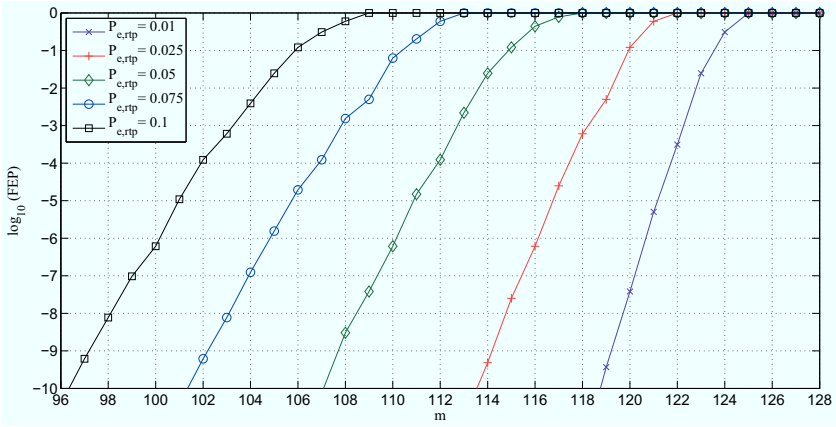


Figure 4.3: Resulting logarithmic FEP for the first I frame of *Football* (byte size equal to 11519) mapped to RS codewords (128,m) at different RTP packet error probability.

the UXP profiler. In this research work the UXP profiler simply drops this cases by slightly compromising the R-D granularity.

We propose a simple strategies by fixing an error probability profile (EPP) $\boldsymbol{\pi}_{f,q,p}$, for each frame f with quality id q and priority level p. Based on this approach, the UXP profile is derived by finding the minimum $m_{f,q,p} \in [\frac{n}{2}+1, n]$ such that

$$FEP_{f,q} \leq \pi_{f,q,p} \tag{4.6}$$

Differently to other solutions in literature, this approach has the main advantage that the expected distortion becomes quasi-independent from the RTP packet failure rate whereas a change of the $P_{e,rtp}$ will only results in a rate increment or decrement. This feature will be exploited while modeling the expected R-D curves, as we will see later.

As a case of study to provide numerical results and illustrate how rate adaptation works when UXP is implemented, we consider here the following choice for the EPP, by differentiating the base and the enhancement layer protections.

4.2.2.1 A case study for the design of EEP

Since the priority level of the quality layers carries both the information of the R-D improvements and the dependency of each frame, the values of the EPP for the quality frames, i.e. $q > 0$, can be derived according to the following formula

$$\pi_{f,q,p} = \begin{cases} \left(\frac{p}{\alpha}\right) 10^{-\frac{p}{\alpha}} & \text{if } p \geq \frac{\alpha}{ln(10)} \\ 1 + \left(\frac{1}{e} - \ln(10)\right) \frac{p}{\alpha} & \text{otherwise} \end{cases} \tag{4.7}$$

where α allows for a trade-off between protection and overhead.

The priority levels for the base layer frames are normally set equal to 63 by the quality processing tool. If the UXP profile used eq. (4.7), it would assign similar protection to the base layer and the first enhancement layers. A smaller frame error rate is ensured for the I-frame, since its loss will produce the drop of all the frames in the GOP. To avoid this we set then $\pi_{0,0,p} = 10^{-6} \quad \forall \alpha$. Moreover, in order to exploit the temporal scalability at the decoder we propose to re-assign to frames of the enhancement temporal layer, with $q = 0$, an higher priority level and to use again the eq. (4.7) to derive the relative EEP values. The choice of the priority level for the enhancement temporal layer depends on the particular frame rate that must be ensured to each user.

4.3 Rate-distortion modeling with Packet Losses

The model in (3.10) for the R-D relationship is still applicable in case of frame losses due to the transmission error in the channel. In this case the empirical points of the encoder are replaced by new points taking into account the effects of packet erasures and UXP. These new points are the result of the rate increase due to UXP , i.e. $\sum_{f=0}^{G-1} \frac{n-m^*_{f,q,p}}{m^*_{f,q,p}} r_{f,q,p}$, and the novel expected distortion $d_{s,loss}$ evaluated as in (4.5). In Figure 4.4 we plot the empirical R-D function resulting from the encoder, as the reference curve, and the related R-D functions outcoming from the UXP profiler at different packet error probabilities $P_{e,rtp} > 0$ for the

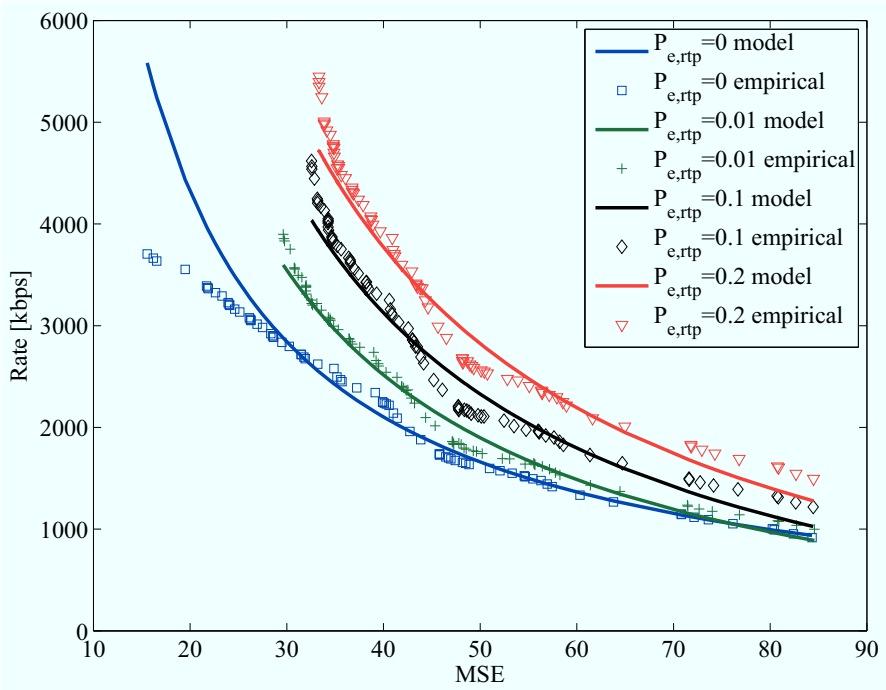


Figure 4.4: R-D Model (straight line), according to eq. (3.10) fitting the empirical R-D relationship for one GOP (size G equal to 8) of the *Football* test-sequence with different error probabilities and α=30. The lower curve refers to the R-D relationship of the encoder.

first GOP of the test-sequence *Football*. We can see that the distortion is almost unchanged for the lower points of the curve with respect to the reference case, since high protection is provided to the high priority levels which are the first to be extracted. At larger bit rates the gap with respect to the reference case increases due to insertion of quality frames with lower protection.

Generally a dynamic adaptation of the UXP to different $P_{e,RTP}$ would require the periodical application of the curve-fitting algorithm to derive the two parameters of the model, thereby increasing the complexity. This problem can be overcome when the UXP profiler adaptively tracks the FEP profile by changing the protection class assigned to the different NALUs. In this way only rate has significant changes while expected distortion practically does not change. While comparing the empirical points resulting from different error

Video	$P_{e,rtp}$	α = 15		α = 30	
		Overhead	$d_{GQ,loss}$[MSE]	Overhead	$d_{GQ,loss}$[MSE]
Foreman	0.01	8.4 %	1.82	5.3 %	4.54
	0.05	17.7 %	2.13	13.7 %	5.15
	0.1	28.0 %	2.17	23.1 %	5.31
Harbour	0.01	7.8 %	8.95	5.1 %	19.87
	0.05	17.1 %	9.86	13.3 %	20.32
	0.1	27.6 %	10.13	23.4 %	20.89

Table 4.1: Percentage of the overhead and expected distortion $d_{GQ,loss}$ in term of MSE with respect to the full quality video streams ($Q = 10$ and $G = 8$), for different values of RTP packet error probability and α parameter in the EEP profile.

probabilities ($P_{e,rtp} > 0$), we can note in the figure how the proposed UXP profile leads to similar distortion at different $P_{e,RTP}$ values. Therefore the adaptation module adapts the sequence-dependent parameters by simply adding a constant dependent on the value of $P_{e,rtp}$. According to extensive simulations the rate shifting is independent of the encoded sequence and can be determined by empirical evaluations.

This result can also be appreciated in Table 4.1 where the average expected distortion due to different $P_{e,rtp}$ and the resulting average overhead is evaluated for two video sequences with full quality.

The selection of a small value of α for the EEP results in a small FEP for the quality layers, thereby increasing the overhead. On the other hand, a loss in the expected quality is experienced by doubling α with a consequent rate gain in the order of 5%. As mentioned before, the overhead is approximately constant even for video sequences with high spatial and temporal complexity difference, such as *Foreman* and *Harbour*. On the other hand, the loss in the expected quality strictly depends on the range of the distortion values as normally increase with the complexity of the video raises if the same encoding paradigm is used for each sequence.

4.4 Packet-erasure channel

In this section we asses the performance in the case of transmission over packet-erasure channel, by evaluating only the proposed algorithm with two different GOP sizes. The number of bytes per RS codeword is set equal to $n = 128$ (as a shortened version of the code with natural length 255) by allowing the insertion of more than one GOP into a TB and then filling the payload of each RTP packet with a reasonable number of bytes. In order to limit the overhead to about 20% for the worst case considered, i.e. $P_{e,rtp} = 0.1$, the parameter α is set equal to 30 (see Table 4.1). According to extensive simulations we define the range of the EPP values for the enhancement temporal layers between 10^{-6}, which is intended to the I-frame, and $10^{-(6-T)}$. We also consider a value of bandwidth sufficiently high, i.e. R_c= 7000 kbps, to allow the insertion of the higher quality layers which have less protection.

Table 4.2 shows the average distortion resulting at different $P_{e,rtp}$ for the different video sequences. The average is obtained by looping the first 240 frames of each sequences for 1000 times. Here, $D^*_{rec,av}$ is the average received MSE; D^*_{av} is the average expected distortion which is the discrete solution of the adaptation algorithm, and $D^*_{enc,av}$ is its related encoding distortion. We can note that the expected distortions as well as the received distortions at the same RTP packet failure rate $P_{e,rtp}$ are approximately equal, showing the goodness of the framework even in presence of packet erasures. The distortion values decrease for most of the video sequences, while the packet error rate increases, due to the effect of bandwidth constraint. At large values of $P_{e,rtp}$ the outcoming overhead from the UXP profiler increases and the Adaptation module reacts by reshaping the rate of each sequence, thereby increasing the distortion to provide fairness. This behavior is less marked in the case of GOP size equal to 8 for the *Foreman* sequence whose distortion does not

Video	$P_{e,rtp}$	$G=8$			$G=16$		
		$D^*_{rec,av}$	D^*_{av}	$D^*_{enc,av}$	$D^*_{rec,av}$	D^*_{av}	$D^*_{enc,av}$
Coastguard	0.01	33.9	37.4	29.6	27.3	29.4	19.8
	0.05	37.5	40.1	33.6	31.2	32.0	22.4
	0.1	40.8	42.3	37.8	36.1	37.7	27.0
Crew	0.01	36.5	36.6	36.2	28.4	28.4	28.2
	0.05	39.3	39.4	39.1	32.4	32.5	32.3
	0.1	41.4	41.5	41.3	36.6	37.0	36.0
Football	0.01	35.2	35.6	34.0	27.9	28.4	26.4
	0.05	38.4	38.9	37.1	30.8	31.6	29.2
	0.1	41.8	41.8	40.5	35.9	37.3	34.3
Foreman	0.01	35.7	35.6	34.2	28.1	28.7	27.9
	0.05	35.9	36.0	35.4	0.4	30.8	30.1
	0.1	36.2	37.1	36.1	33.8	34.9	33.2
Harbour	0.01	35.3	38.8	23.7	29.8	30.3	18.2
	0.05	40.6	42.2	26.5	32.0	32.3	20.3
	0.1	42.8	44.2	31.0	34.4	37.8	22.9

Table 4.2: Average received distortion,$D^*_{rec,av}$, expected distortion, D^*_{av}, and encoding distortion, $D^*_{enc,av}$, in term of the MSE for different video sequences, GOP size G, and packet-erasure rate values $P_{e,rtp}$, resulting from the proposed rate-adaptation algorithm. Available bandwidth is R_c =7000 kbps.

change significantly, since it receives in most cases only the base-layer with the highest protection. The slight increase of distortion with respect to the encoding MSE is due to the loss of certain enhancement temporal layers.

As expected, an higher GOP size decreases the distortion thanks to the higher coding efficiency, that allows to improve the R-D performance of the base layer. Nevertheless, such gain is reduced with respect to the case of error-free channel, since more quality layers with low protection are transmitted. This behavior can be improved with a more careful design of the EPP aimed at balancing overhead and degree of protection according to the available bandwidth.

4.5 Conclusions

In this work a multi-stream rate adaptation framework has been proposed with reference to temporal/SNR-scalability of SVC with MGS and by considering transmission over a packet-erasure channel. A simple UXP scheme has also been included with the aim to maintain high expected quality even in presence of high packet error rate. This framework is suitable for video applications such as video on-demand, IP-TV services and real-time streaming. A general discrete problem aimed at maximizing the average distortion while providing fairness to different video sources has been proposed. Then, a semi-analytical model that estimate the R-D relationship of each video source GOP-by-GOP has been developed and successively tested with respect to goodness parameters and complexity. The general discrete problem has then been relaxed and an optimal procedure has been derived based on the low-complexity R-D model. The numerical results have shown the feasibility of our framework through the investigation of the achieved fairness, the gap between the relaxed and the related discrete solution according to the fairness metrics adopted, and the improvements with respect to an equal-rate assignment scheme.

Chapter 5

Rate Distortion Modeling for Real-time MGS Coding

5.1 Introduction

Video streaming is one of the most popular applications of today's Internet. As the Internet is a best effort network, it poses several challenges specially for high quality video streams. The Advanced Video Coding (H.264/AVC) scalable extension, also called Scalable Video Coding (SVC), provides an attractive solution for the difficulties encountered when video source is transmitted over wireless transmission systems. Such challenges include error prone channels, heterogeneous networks and capacity limitations and fluctuations [1]. Scalable video coding provides three types of scalabilities, namely spatial, temporal and SNR scalability. These types of scalability allow a sub stream of a particular resolution, frame rate and quality to be extracted in order to be adapted to various network conditions and terminal capabilities.

Rate-Distortion (R-D) models are used to predict rate and distortion of video sequences prior to the encoding process. The rate of a video sequence is expressed in bytes/s, while the distortion is defined in terms of Mean Square Error (MSE). The Peak Signal to Noise

Ratio (PSNR) is more often used to express the quality of a video sequence.

Within SVC, each sequence is encoded with one base layer and several enhancement layers which can be sequentially dropped by providing a graceful degradation. SNR scalability is achieved by using Coarse Grained Scalability (CGS) or Medium Grained Scalability (MGS) [14]. In CGS a limited number of discrete points can be extracted which is equal to the number of coded layers, while MGS provides finer granularity of quality scalability by dividing each CGS layer into 16 MGS layers.

Different video sequences have different complexities, hence the relationship between rate and quality differs from one video sequence to another. Assuming the same physical resources are shared among different video sequences, an equal rate allocation scheme would divide the available rate equally among the sequences, which may lead to a high or even unacceptable level of distortion for more complex videos which require higher rates. To optimize transmission strategy based on the QoE of the end user, the rate should be allocated among the videos based on fairness criterion.

In the literature several R-D models have been proposed to predict rate and distortion prior to the completion of the encoding process. Enhanced R-D models for H.264/AVC were proposed for coded video sequences in [19]. However the parameter extraction is performed after transformation and quantization in the encoding process. The late extraction of the parameters can significantly affect real time applications such as video over wireless networks. An improved real time rate distortion model for medium grain scalable video coding is proposed in [15] which reduces significantly the dependency on the encoding process. In this model the delay is reduced by extracting the parameters before transformation.

In this chapter we propose a new rate-distortion model for real time MGS video streams. Our model only uses two parameters which are calculated taking into account the characteristics of the video sequences through a spatial and a temporal index extracted from the original raw video streams. Moreover we also use these complexity indexes to calculate

base layer and enhancement layer rates of the given video stream.

This Chapter is organized as follows: Section 5.2 reports a brief overview of rate-distortion modeling. Our proposed rate distortion model is illustrated in Section 5.3. Section 5.4 describes simulation and model verification of our algorithm, while conclusion is drawn in Section 5.5.

5.2 Overview of Rate Distortion modeling

In this section we give a brief overview of R-D models. R-D models describe the relationship between the bit rate and the expected distortion and vice versa in the reconstructed video stream. The trade-off between the goal of reducing the bit rate and the goal of keeping the distortion at acceptable levels can be afforded dynamically, in order to perform adaptation to different conditions. A R-D model enables to predict the minimum bit rate required to achieve a target quality. The performance of the streaming system is directly affected by the accuracy of the rate distortion model [29].

The time required to model the R-D curve for a given sequence may drive the decision on the methodology/algorithm to be adopted for the model.

For real time video streaming systems the computation of the model should be fast enough to deal with the timing constraints of the video stream. Many rate distortion models have been proposed in the literature for real time and non-real time video streaming. They are often categorized in analytic, semi analytic and empirical models. Empirical models require the computation of all the set the R-D points resulting in a high complexity. Semi-analytical models aim at reducing such complexity by deriving parametrized functions that follow the shape of analytically derived functions, but are evaluated through curve fitting from a subset of the rate-distortion empirical data points. In this preliminary work we investigate techniques to further reduces the complexity of semi-analytical models. This is made possible by introducing new functions dependent only on the uncoded video sequences.

The coefficients of this new functions can be estimated off-line through a prior knowledge of the parameters of a set of video sequence samples, and then used for any future video sequence.

5.3 Proposed Model

In this section we propose a parametric R-D model for MGS SVC which is simple enough to be used by rate-adaptation techniques in real-time video streaming. The models depends on the Spatial Indexes (SI) and the Temporal Indexes (TI) of the original raw video sequence.

After encoding, the GOP of the k-th generic video results in a finite discrete set of codes with rate r_k and distortion d_k. The rate-distortion function which represents this set of point is often modeled as a continuous function, because it can be more easily used to obtain simple rate adaptation algorithms. We consider as a reference R-D model $\mathscr{R}_k(D)$ the one introduced in [30] which is based on two parameters:

$$\begin{cases} \mathscr{R}_k(D) = \frac{\alpha_k}{D} + \beta_k \\ \mathscr{R}_k(D) \geq \mathscr{R}_{k,BL} \\ \mathscr{R}_k(D) \leq \mathscr{R}_{k,EL} \end{cases} \tag{5.1}$$

The parameters α_k and β_k are sequence dependent parameters of the k-th GOP while D is the distortion evaluated as a Mean Square Error (MSE). $\mathscr{R}_{k,BL}$ and $\mathscr{R}_{k,EL}$ are the Base Layer and highest Enhancement Layer rates obtained from the encoded video. The drawback of this model is the fact that its parameters can only be evaluated by looking for the best fitting of at least 4 R-D points after the encoding process of the video, making it of difficult use for real time applications.

The model proposed here replaces the parameters α_k and β_k with the spatial index SI_k and the temporal index TI_k, also called spatial and temporal complexities, in the following way:

$$\alpha_k = p_1 + p_2 SI_k + p_3 TI_k \tag{5.2}$$

$$\beta_k = q_1 + q_2 SI_k + q_3 TI_k \tag{5.3}$$

The same approach is used to replace base layer and enhancement layer rates, by modeling them as:

$$R_{k,BL} = r_1 + r_2 SI_k + r_3 TI_k \tag{5.4}$$

$$R_{k,EL} = s_1 + s_2 SI_k + s_3 TI_k \tag{5.5}$$

The sets $\{p_1, p_2, p_3\}$, $\{q_1, q_2, q_3\}$, $\{r_1, r_2, r_3\}$ and $\{s_1, s_2, s_3\}$ are the coefficients that are calculated using linear least square fitting method [31] with Least Absolute Residuals (LAR) [32] for robustness in a sufficiently large set of GOPs from different video sequences. As mentioned above, this process is executed offline only once, assuming the availability of a reasonable set of video sequences.

The spatial and temporal complexities are evaluated on the luminance component [33] of the video by means of Spatial Information and Temporal Information [34] of the k-th GOP respectively as follows:

$$SI_k = \max_n std_\sigma\{Sobel[F_n(\sigma)]\}$$

$$TI_k = \max_n std_\sigma\{M_n(\sigma)\}$$

where

$$M_n(\sigma) = F_n(\sigma) - F_{n-1}(\sigma)$$

$M_n(\sigma)$ is the motion difference and $F_n(\sigma)$ is the luminance component with n and σ temporal and spatial coordinates, respectively, of the frame sequence used to encode GOP k.

To summarize, the R-D model is obtained by substituting in (5.1) the parameters α_n and β_n from (5.2) and (5.3), and $\mathscr{R}_{k,BL}$ and $\mathscr{R}_{k,EL}$ from (5.4) and (5.5), respectively:

$$\begin{cases} \mathscr{R}_k(D) = \frac{p_1+p_2SI_k+p_3TI_k}{D} + q_1 + q_2SI_k + q_3TI_k \\ \mathscr{R}_k(D) \geq r_1 + r_2SI_k + r_3TI_k \\ \mathscr{R}_k(D) \leq s_1 + s_2SI_k + s_3TI_k \end{cases} \tag{5.6}$$

The proposed R-D model is verified by considering video sequences generated by the JSVM software [3] We encoded 6 video sequences i.e *Crew*, *Football*, *Coastguard*, *Soccer*, *City*, and *Mother and Daughter* (*MD*) having different scene complexities, in CIF resolution with the frame rate of 30 fps. We denote this set of 6 videos as the training set. Two enhancement layers are used to obtain SNR scalability where each layer is split into 5 MGS layers with vector distribution of [3 2 4 2 5]. All the videos are coded GOP by GOP with a GOP size of 8 to obtain sequences comprising 26 GOPs. The Quantization Parameter (QP) is set to 38, 32 and 26 to obtain the base layer and the two enhancement layers.

Figure 5.1, 5.2, 5.3 and 5.4 shows α, β, base layer (BL) and enhancement layer (EL) models as in (5.2), (5.3), (5.4) and (5.5), respectively, using spatial and temporal indexes.

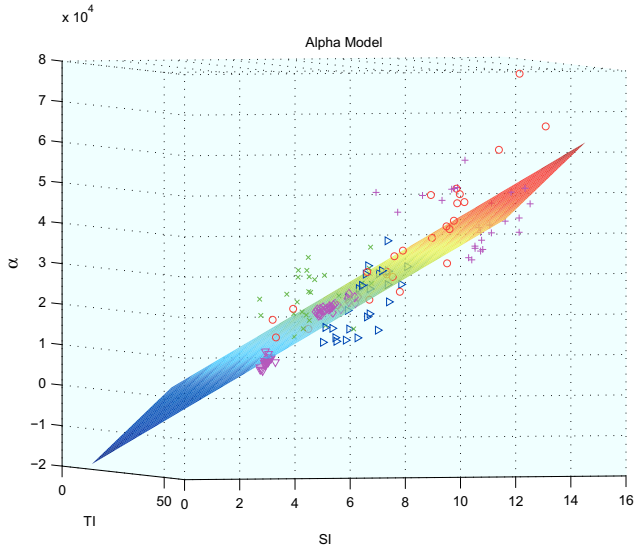


Figure 5.1: Proposed model for α with R^2= 0.987 and $RMSE$ = 1598

In figures 5.1 and 5.2 the markers are referred to the values of α_k and β_k derived according to model (5.1) and plotted for each GOP versus the corresponding value of SI_k and TI_k. In figures 5.3 and 5.4 the markers are referred to the BL and EL layer rates derived by encoding the sequences with JSVM [3].

It can be observed that the values of the parameters for all the models closely follow a linear behaviour. The metrics used to evaluate the goodness of the model in fitting the set of points are the coefficient of determination (R^2) and Root Mean Square Error ($RMSE$).

The sets of coefficients p, q, r, and s, appearing in (5.2), (5.3), (5.4) and (5.5) of the proposed model, result to be, for the training set, as follows:

$$p = \{-2.4\times10^4, 3975, 540.5\}$$

$$q = \{-246.1, 24.13, 3.328\}$$

$$r = \{41.27, 17.09, 9.12\}$$

$$s = \{-237, 145.6, 34.02\}$$

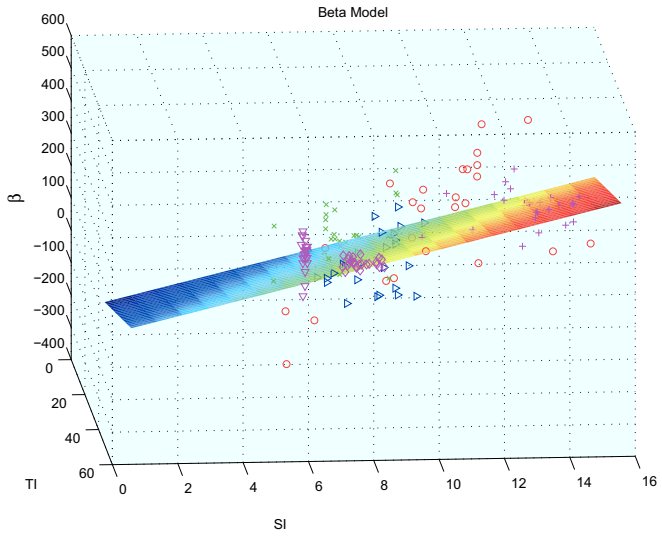


Figure 5.2: Proposed model for β with R^2= 0.973 and $RMSE$ = 21.2

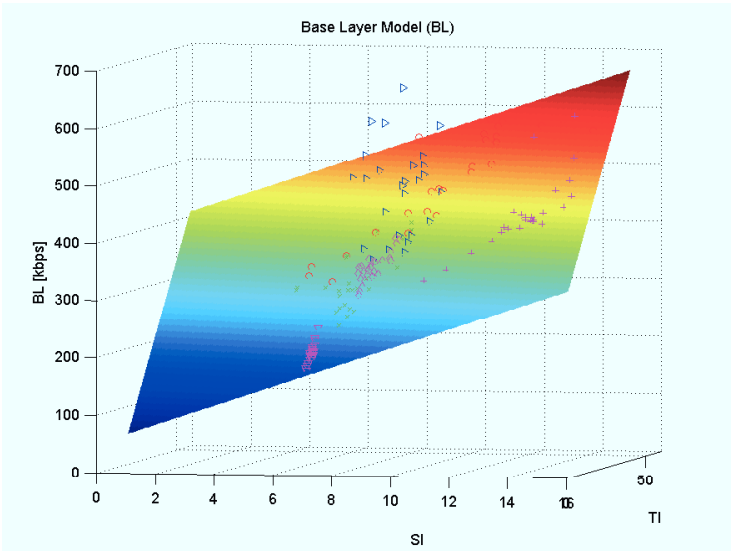


Figure 5.3: Proposed model for (BL) with R^2= 0.979 and $RMSE$ = 22.98

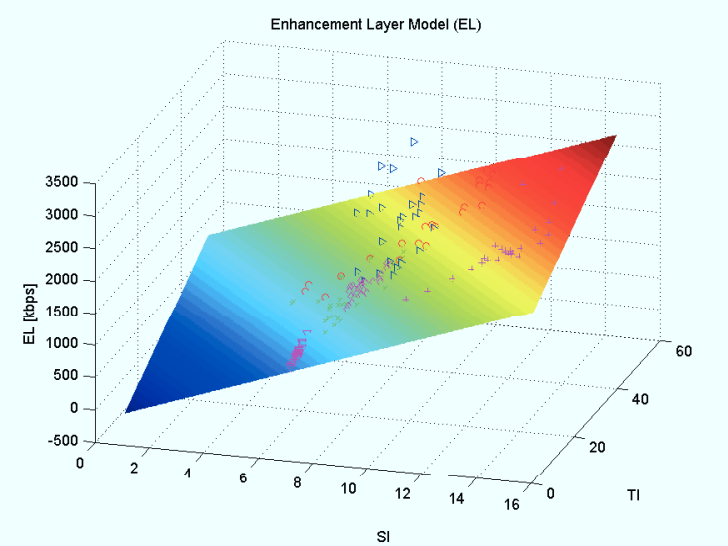


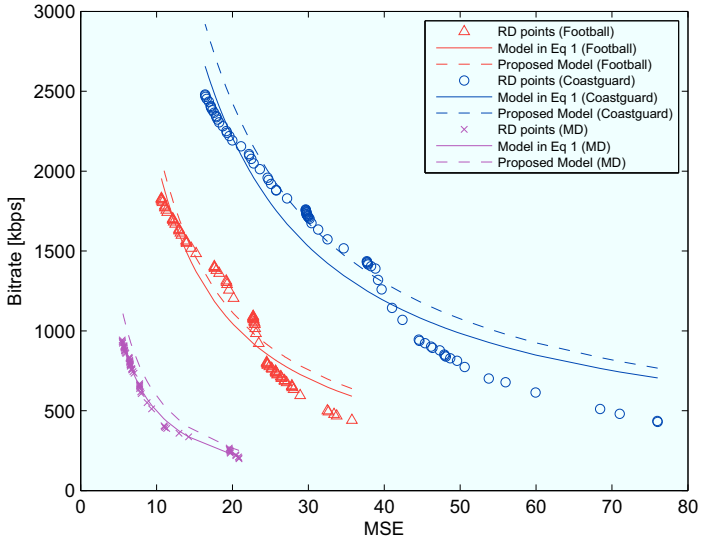
Figure 5.4: Proposed model for (EL) with R^2= 0.985 and $RMSE$ = 79.36

In Figure 5.5 the different R-D models are shown and compared for two sample GOPs of three video sequences. The accuracy changes GOP by GOP: figure 5.5(a) shows the result for a sample GOP with good matching between proposed model in (5.6) and model in (5.1), whereas the figure 5.5(b) shows a result with poor matching. As it will be shown in Section 5.4, the GOPs with less accurate model do not have significant impact on the behavior of rate adaptation strategies in real time multivideo transmission.

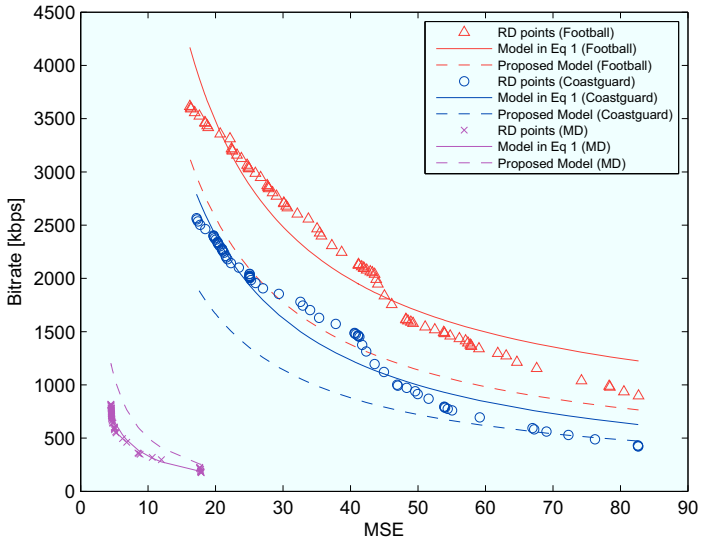
To evaluate the goodness of BL and EL rate estimation, we compare in Figure 5.6 the rates estimated with the model in (5.4) and (5.5) to the original rates obtained from the encoded sequences.

We consider not only the video sequences in the training set but also the sequences outside the training set. More emphasis is given to base layer rate as it is the minimum rate requirement of each video sequence when transmitted in bandwidth constrained channels.

It can be observed from Figure 5.6 that our model predicts the BL rate quite accurately, not only within the training set but also for the sequences outside this set, as shown for *Mobile* and *Foreman* in Figure 5.6. Moreover it can also be seen from Figure 5.6 that the



(a)



(b)

Figure 5.5: R-D comparison among model in eq (5.1), proposed model and actual values for two sample GOPs.

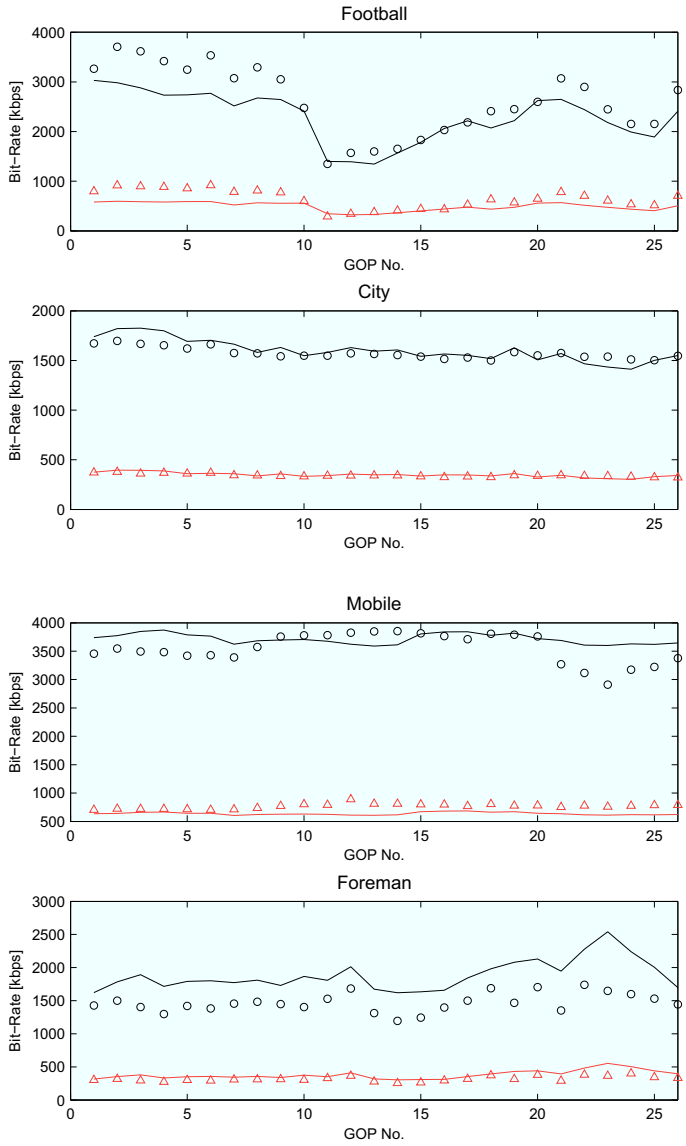


Figure 5.6: BL and EL rates over 26 GOPs for two sequences in the training set (*Football* and *City*) and two sequences outside the training set (*Mobile* and *Foreman*). The marker points refer to the original BL and EL rates, whereas the solid lines refer to rates estimated from (5.4) and (5.5), respectively.

estimation is good also for EL rate.

5.3.1 Validation of the proposed models

The proposed models are useful to build up rate-adaptation algorithms that adaptively set encoding parameters or scale the video to suitably optimize the transmission in a bandwidth-constrained, time-variant channel shared by multiple videos. The rate adaptation algorithm used to validate the proposed R-D model is explained in section 3.2 and 3.4 of Chapter 3.

5.4 Simulation and Model Verification

In this section we verify the proposed R-D model in the transmission of multiple videos over a bandwidth constrained channel by using the rate adaptation algorithms outlined in Section IV and described in detail in [30]. We propose results for both the videos in the training set and videos outside it. In the first case a bandwidth limited to $R_c = 3500$ kbps is considered. In the second case a set of 4 sequences, i.e *Foreman*, *Harbour*, *Container* and *Mobile*, and a bandwith limited to $R_c = 3000$ kbps is considered.

Tables 5.1 and 5.2 show the average MSE taken over the first 26 GOPs for the model (5.1) and our proposed model. It can be seen that ER algorithm assigns less distortion to the low complexity videos like *MD*, *City* in the training set and *Foreman* and *Container* outside the training set, thus compromising the quality of more complex videos like *Football*, *Coastguard* or *Harbour* and *Mobile*. This behavior is mitigated by the OPT algorithm, as expected. Moreover, it can also be observed from both tables that the average MSE values for the proposed model closely follow to the model (5.1) except for *Harbour* and *Container* in Table 5.2 with OPT algorithm. For the ER algorithm in both table 5.1 and 5.2, the results for model (5.1) and our proposed model show only slight differences mainly due to the fact that our estimated maximum and minimum rates which are the BL and EL rates are different from the original BL and EL rates.

Sequence	Model (1)		Proposed Model	
	ER	OPT	ER	OPT
Crew	36.99	38.00	36.24	44.09
Football	53.11	44.00	53.11	46.69
Coastguard	70.78	45.72	69.09	46.82
Soccer	39.93	42.15	38.21	34.72
City	37.61	51.05	35.41	49.10
MD	9.00	20.65	8.69	20.59

Table 5.1: Average MSE over 26 GOPs obtained with the model (5.1) and proposed model in the transmission of the training set of 6 videos.

Sequence	Model (1)		Proposed Model	
	ER	OPT	ER	OPT
Foreman	19.02	34.90	18.29	31.60
Harbour	79.78	57.86	79.18	81.11
Container	15.88	35.39	15.45	18.14
Mobile	103.84	65.44	103.84	72.76

Table 5.2: Average MSE over 26 GOPs obtained with the model (5.1) and proposed model in the transmission of 4 videos not included in the training set.

A more detailed observation can be done through Figure 5.7 which compares the MSE obtained after rate adaptation for the sequences *Football*, *City*, *Mobile* and *Foreman*, with our proposed model and the model (5.1). It can be observed that, with the exception of some large deviations experienced in few GOPs of *City* and *Mobile*, our model closely follow model (5.1). The exceptions suggest, in practical applications, that video servers determine off-line different models as in (5.6) for a limited number of video classes having homogeneous characteristics.

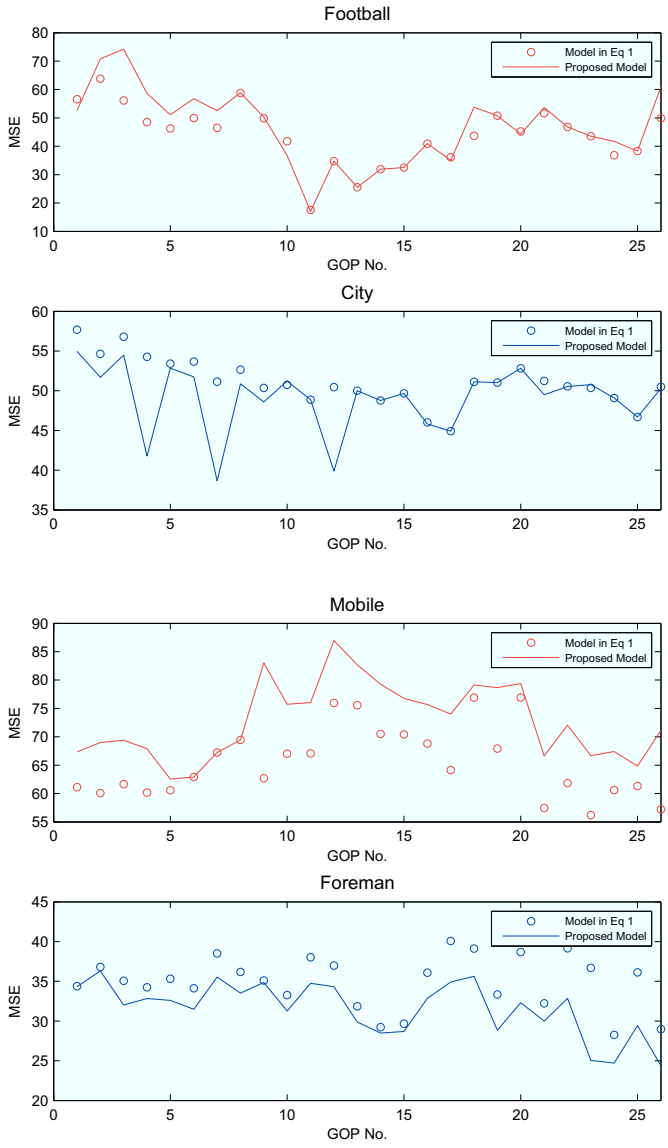


Figure 5.7: Averaged MSE for each GOP of two sample videos in the transmission over bandwidth constrained channel with rate adaptation. Figures *Football* and *City* refer to the transmission of the 6 videos of the training set ($R_c = 3500$ kbps), wheres figures *Mobile* and *Foreman* refer to the transmission of 4 videos not included in the training set ($R_c = 3000$ kbps).

5.5 Conclusions

In this work we proposed a new rate-distortion model using spatial and temporal indexes for MGS scalable video coded streams with reference to SNR scalability. The model has been developed with particular aim to real time video streaming over wireless channels, as it only needs spatial and temporal indexes from the original unencoded video to build the R-D relationship and to estimate base layer and enhancement layer rates. The model has been compared to the state of the art non real time model in applications where multiple video are transmitted over a bandwidth constrained channel with rate adaptation showing that the proposed model works well.

Chapter 6

QoS for VoIP Traffic in Heterogeneous Networks

6.1 Introduction

The popularity of wireless networks is widely recognized because of its strong support and ease of use in the end systems. Heterogeneous wireless networks are becoming of widespread use with Internet's real-time multimedia applications. Short range WLAN systems, as well as different cellular systems and WiMAX, provide some level of QoS and are needed to realize ubiquitous Internet services. But real time multimedia applications, in particular interactive and live streaming applications, set strict requirements for the QoS. Some applications need relatively wide bandwidth; the bandwidth should be available in both directions constantly. Applications like voice and video need short transmission delay and jitter but they still have ability to tolerate some packet loss [35]. WiMAX is capable of reaching remote areas with high data rate transfer, mobility support and a native Quality of Service management (even if just limited to the wireless IEEE802.16 links) [36]. By looking at the literature, a remapping mechanism is proposed in [37] to dynamically adjust the mapping rules for nrtPS and rtPS (for VBR traffic sources) classes of WiMAX to DiffServ.

An architecture for signaling and WiMAX resources management is proposed in [38] considering an end-to-end QoS enabled scenario. In this approach interoperability is provided between WiMAX and other networks which have different QoS schemes, like DiffServ. WiMAX and WLAN Integration design is proposed in [39] for link layer QoS. Here, a mapping scheme of DiffServ to the link layer services for both WiMAX and WLAN is shown. The end-to-end QoS mechanisms were developed to serve the users with the wired terminals. More research work on DiffServ approach applied to the wireless systems and mobile users in heterogeneous environment is needed in order to understand the benefits of the DiffServ networks. Current research is open regarding the mapping of QoS classes and the design of complete interworking models between WiMAX and DiffServ networks. In our work, a WiMAX DiffServ QoS test-bed scenario is implemented to test the interoperability and the different functionalities between domains with different QoS models. The aim of our research is to map and analyse the QoS class for CBR traffic types (VoIP without silence suppression, as an example) which need constant bandwidth in both wired and wireless networks.

To provide better Quality of Experience (QoE) to customers in efficient manner i.e with respect to cost as well as with QoS. From this point we understood that QoS & QoE are mutually dependent and to achieve QoE, QoS is the basic building block [40]. The requirement of QoE and QoS along with the QoS parameters with priority order can be helpful for both the operators and users to maximize the network performances and user satisfaction level with the limited resources. During the limited resource condition, QoS requirement can be optimized according to the service type, price, user requirement and priority of QoS parameters [41].

This chapter is organized as follows. Section 6.2 explains the mechanism for IP QoS. Section 6.3 depicts the QoS mechanism in WiMAX network. Section 6.4 illustrates our simulation scenario and results, whereas conclusion are drawn in section 6.5.

6.2 Mechanism for IP QoS

There are two main IP based QoS mechanisms, IntServ and DiffServ. IntServ provides end-to-end QoS in flow-based manner and uses the Resource Reservation Protocol (RSVP) for signalling, which follows the data path, performs the reservation and maintains per flow state in each router. DiffServ has more suitable mechanisms for providing end-to-end QoS by working with aggregate traffic classes [42]. Packets of a particular service class are marked with a QoS class and receive a specific Per Hop Behaviour (PHB) for forwarding. The PHB is an externally observable forwarding behaviour which is applied to a DiffServ compliant node, or it refers to queuing scheduling, shaping or policing behaviour of a node on any packet. There are several available standard PHBs, which include default PHB, Assured Forwarding (AF) PHB and Expedited Forwarding (EF) PHB. The packets scheduled by default PHB receive the traditional Best Effort (BE) service which has the lowest priority. The AF class is further categorized into four classes, namely AF1, AF2, AF3 and AF4, and each class has three drop precedences: Low, Medium and High. The purpose of the AF PHB is to allow the DiffSev network to provide different levels of QoS assurances. Generally AF class is used for the traffic which can tolerate more delay and packet loss, but requires better QoS than Best effort (BE) class. The main purpose of the EF PHB is to provide assured bandwidth equivalent to 'virtual leased line'. Asynchronous Transfer Mode (ATM) has also attempted the same assured service in its Constant Bit Rate (CBR) traffic mode. The characteristic of this type of service is to provide low delay and small packet loss ratio.

DiffServ uses IP header field (Type Of Service (TOS) in IPv4 and traffic class in IPv6) to denote the QoS class of a packet as shown in figure 6.1. Using DiffServ Code Point (DSCP) each router in the network can mark, shape or drop the incoming traffic. The DSCP field is made of eight bits out of which only six bits are currently in use while the last two bits are for future use. The first three bits of the class selector code points are used to specify the

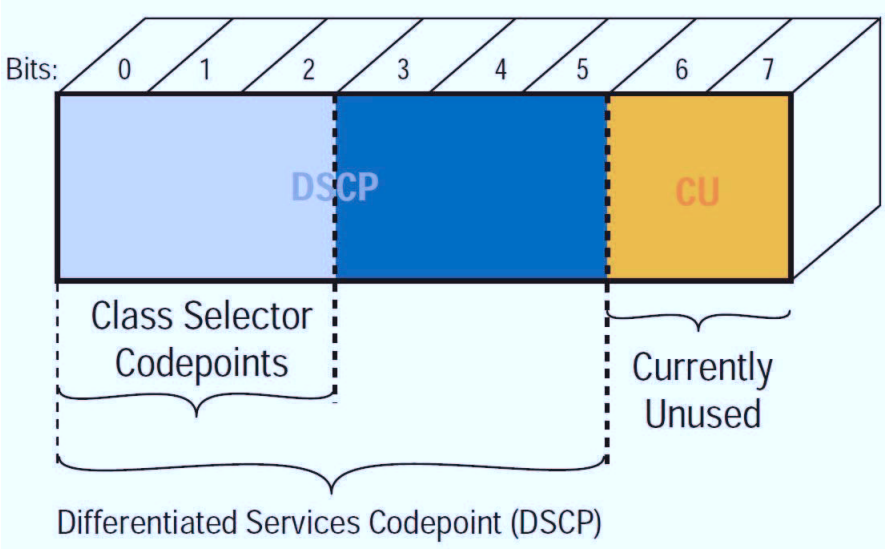


Figure 6.1: DiffServ Code Point field.

different classes with different priorities. The next three bits of the DSCP field are used to handle drop precedence of each of these classes.

6.3 QoS Mechanism in WiMAX Network

IEEE 802.16, referred to as WiMAX, provides specifications for air interface of Metropolitan Area Network. The standard specifies connection-oriented QoS support [43]. There are different types of services for different types of classes, which include: Unsolicited Grant Service (UGS) for real-time uplink Service Flows (SFs) of fixed packet size on periodic basis, Real-time Polling Service (rtPS) for real-time SFs having variable-size packets on periodic basis, Non real-time Polling Service (nrtPS) which supports delay tolerant data having variable-size packets for which minimum data rate is needed, and Best Effort (BE) for the data streams for which no minimum service is required. Service Flows are created and modified between MS and BS through MAC message exchange. The exchange of Dynamic Service Deletion (DSD), Dynamic Service Change (DSC) and Dynamic Service

Addition (DSA) messages are initiated by either BS or MS.

The distinguishing feature of WiMAX over its other competitors (i.e. 802.11 and 3G) is its QoS provisioning based on the association of each packet with a service flow. WiMAX is connection-oriented and each connection has a unique Connection ID (CID) and Service Flow ID (SFID) which is associated to that particular class. The data is mapped by the upper part of the MAC to QoS service classes. The external application can also request desired QoS parameters using the named service class. The traffic shaping engine is included in the MAC which is ultimately responsible for the transmission and reception of the 802.16 packets according to the applied QoS parameters. These parameters are different from one service flow to another.

WiMAX allocates traffic to a service flow and then maps it to a MAC connection using CID as shown in figure 6.2. In this way, IP and UDP protocols which are connectionless are transformed into connection-oriented service flows. An application or group of applications can be represented with a connection with same CID.

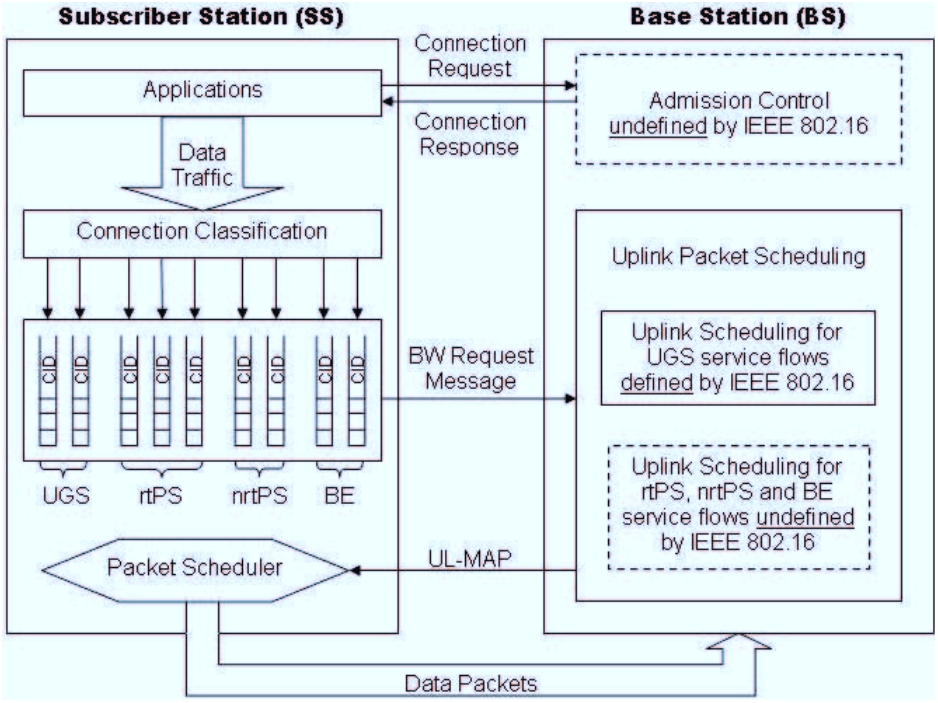


Figure 6.2: IEEE 802.16 QoS Architecture

The MAC layer of WiMAX is divided into two sub-layers: the common part sub-layer and the convergence sub-layer. The transport layer specific traffic is mapped by the convergence sub-layer to the core MAC common part sub-layer. The common part sub-layer is responsible for fragmentation and segmentation and is independent of the transport mechanism.

The incoming traffic type (e.g. web surfing, voice ATM CBR etc.) is classified by the convergence sub-layer and a 32- bit SFID is assigned to it. When a service flow is active or admitted, it is mapped to a 16 bit unique CID which handles its QoS requirements. Each service flow is defined by a QoS parameter set which describes its jitter, latency and throughput assurances.

After the service flow is assigned with a unique CID, it is then forwarded to the appropriate queue. Base Station (BS) performs the uplink packet scheduling by signaling to the

Subscriber Station (SS) [43]. The packet scheduler in the SS will extract the packets from the queues and transmits them to the network with an appropriate time slots sent by the BS in the Uplink MAP (UL-MAP) message.

6.4 Inter-Working Model and Simulation

We consider here the scenario of Figure 6.3, where the traffic from the WiMAX domain enters the DiffServ core network. In the core network the traffic will be mapped to the equivalent class of WiMAX according to Table 6.1 [35]. In this scenario the two nodes, Node1 and Node2, are used to generate traffic that competes with WiMAX traffic.

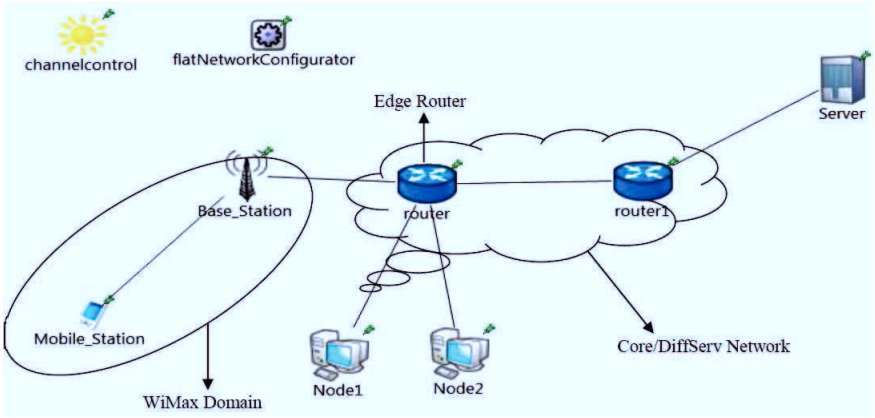


Figure 6.3: WiMAX and DiffServ Network Simulation Scenario

In our test network we have considered a VoIP application with Unsolicited Grant service (UGS) and mapped it to the Expedited Forwarding (EF) class of the DiffServ network. The aim of the simulation is to examine different QoS parameters when WiMAX traffic enters the DiffServ network and when DiffServ domain is not used for CBR traffic.

In the test-bed the UGS type of traffic coming from WiMAX is mapped to a high priority queue in the edge router of the DiffServ core network which provides the service of EF

WiMAX Scheduling Class	DiffServ PHB	Service Example
UGS	EF	VoIP without VAD
rtPS	AF4	Audio/Video streaming
nrtPS	AF3	Transactional Services
BE	BE	E-mail download

Table 6.1: WiMAX and DiffServ traffic class mapping

class. This queue has high scheduling priority compared to the other queues and the data in this queue will be scheduled first.

6.4.1 Priority Queuing (PQ)

In our simulation scenario we have implemented Priority Queuing in the edge router of the DiffServ core network. PQ realizes a simple way of class distinction. As shown in Figure 6.4, if N queues are created, then the priority goes from 1 to N. The scheduler will schedule higher priority queue first and when that queue is empty it will schedule the packets from the next high priority queue. The j-th queue packets are processed only if the higher priority queues, i.e. queue 1 to j-1, are empty. If the scheduler is at queue j a packet arrives in a higher priority queue, say j-3, the scheduler will go to the queue j-3. PQ is particularly suitable for high priority traffic and provides premium service to the traffic which is extremely critical and needs to be processed as soon as possible.

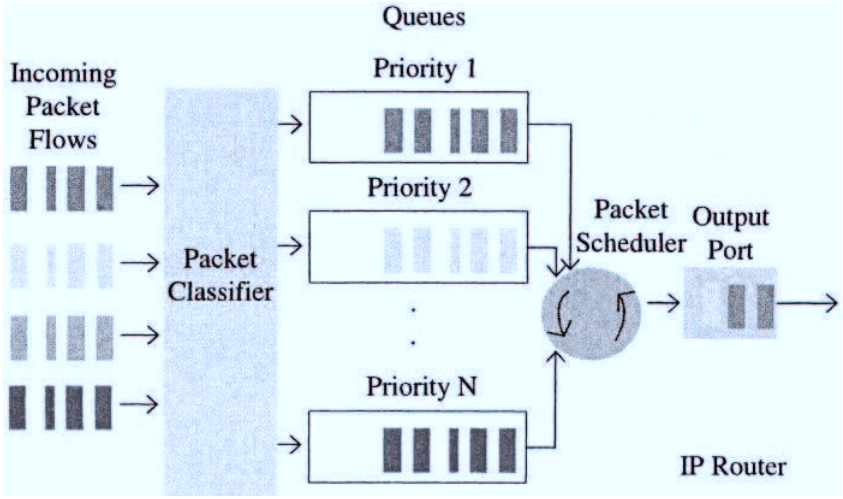


Figure 6.4: Priority Queuing Implemented in Edge Router

We selected priority queuing because it provides dedicated queues for real time traffic, e.g. video and voice over IP.

6.4.2 Simulation Results

The scenario is simulated in OMNeT++ 4.0 with INET framework [44]. The simulation is performed according to the parameters shown in table 6.2. The simulation is run under the traffic load of 100%, 112.5% and 125% with and without DiffServ enabled core network. Node 1 and Node 2 generate exponentially distributed traffic with different arrival rates mentioned in table 6.2. The packets from both the nodes are not marked with DiffServ code point so they will be treated as best effort traffic upon their arrival on the edge router of the DiffServ core network.

Figure 6.5 shows the number of dropped packets without DiffServ core network enabled with different loads. Packets are dropped when the bandwidth required to transmit them exceeds the allocated capacity. We found no dropped packets from VoIP stream with DiffServ enabled core network because of the preferential treatment over the other traffic

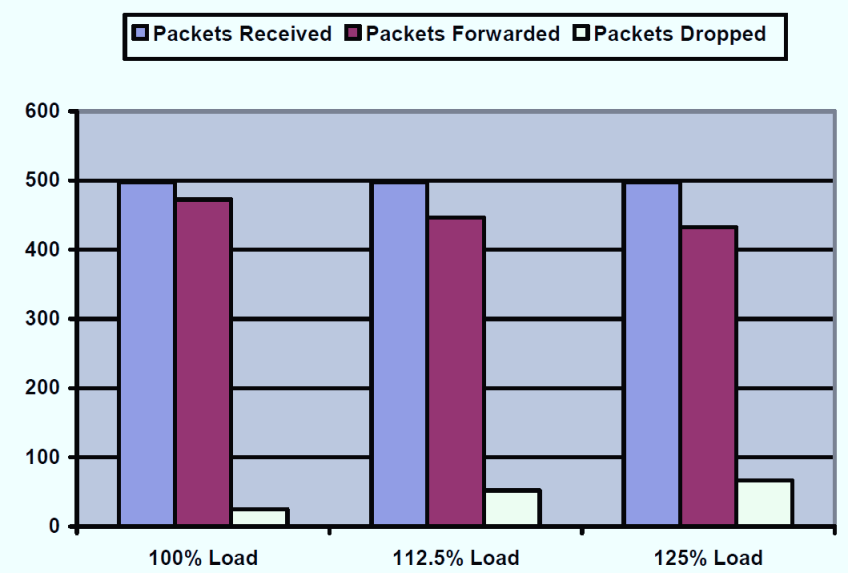


Figure 6.5: Packets dropped without DiffServ support.

types. When the core network is not DiffServ enabled, it cannot provide preferential treatment to the VoIP packets coming from WiMAX domain. All the VoIP packets are treated as best effort on the edge router along with the traffic from Node 1 and Node 2.

As our interest lies in VoIP traffic from the mobile station, in figure 6.5 we have not considered the dropped packets coming from the other nodes.

In Figures 6.6, 6.7 and 6.8 some simulation traces for packet delay in the network are shown under different network loads, with and without DiffServ enabled network. H1 and H2 represent Node1 and Node 2 while MS1 represents the Mobile Station in the simulation scenario of Figure 6.3.

Figure 6.6 shows the packet delay along the simulation time in the scenario without DiffServ core network and with 100% of network load. As can be seen from the figure, the packets from the mobile station are mixed with those from node 1 and node 2 and routed without QoS provision. The VoIP packets are experiencing different delays over the simulation time, though these packets need a fixed and constant bandwidth and are generated with fixed size on periodic intervals. These delays affect the time sensitive data.

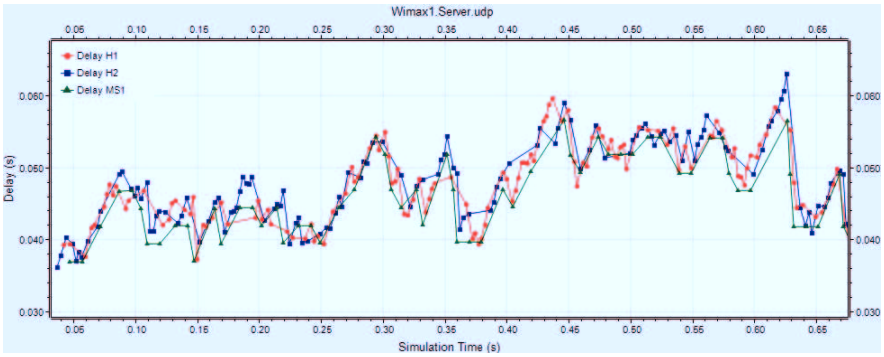


Figure 6.6: Delay without DiffServ support and with 100% load.

In Figure 6.7 the delay of the VoIP data from the mobile station varies between 37 milliseconds and 58 milliseconds in different intervals of simulation time. On the other hand, figure 6.7 refers to the VoIP stream passing through a DiffServ enabled network which shows a smooth data flow along the simulation time. It can be seen that almost all the packets experience the same fixed delay which is about 35 milliseconds and there is an average gain of 10 to 15 milliseconds achieved with DiffServ enabled network.

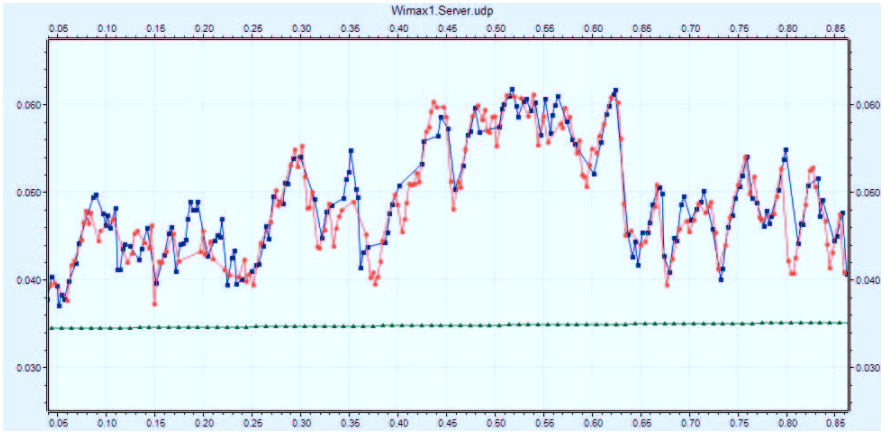


Figure 6.7: Delay with DiffServ support and 100% load.

Figure 6.8 shows the packet delays in a scenario with 112.5% traffic load condition, without DiffServ enabled network. The VoIP stream of the mobile station experiences

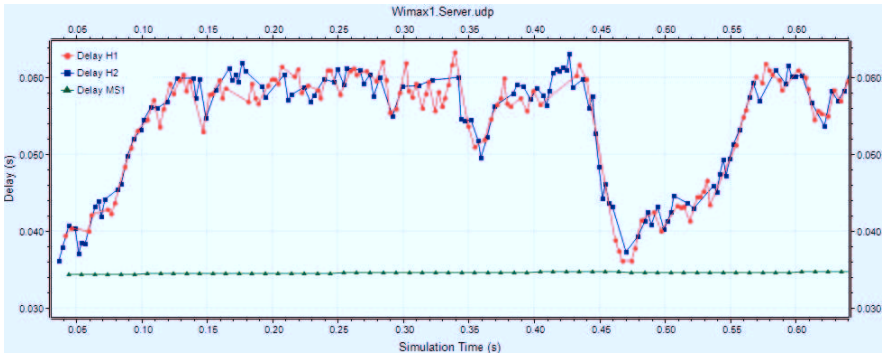


Figure 6.9: Delay with DiffServ support and 112.5% load

higher delays and most of the packets experience a delay between 46 milliseconds and 56 milliseconds. In Figure 6.9, referred to DiffServ enable network, the delay is stable at nearly 35 milliseconds.

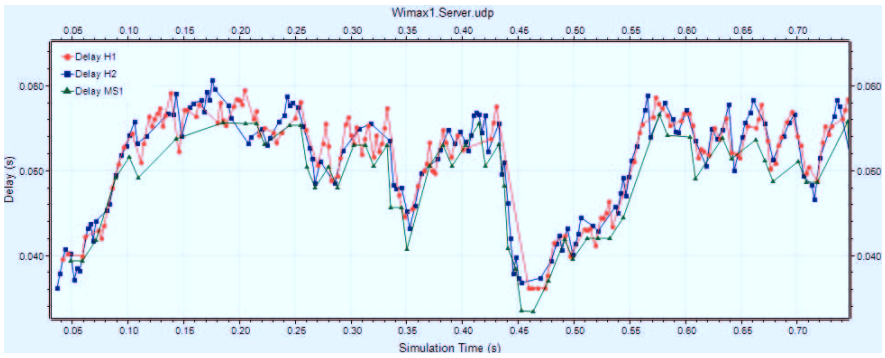


Figure 6.8: Delay without DiffServ support and 112.5% load.

Figures 6.10 and 6.11 show packet delays for VoIP traffic in scenarios with and without DiffServ and with 125% traffic load. Without DiffServ enabled network the majority of the packets have delays above 50 milliseconds whereas the delays in the DiffServ enabled network for VoIP is still constant around 35 milliseconds.

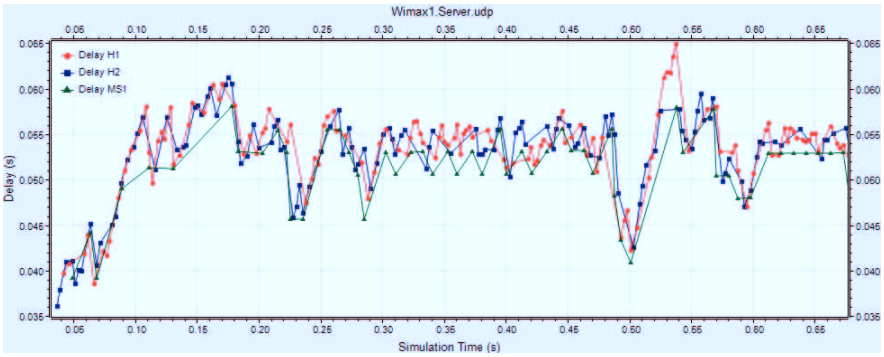


Figure 6.10: Delay without DiffServ support and 125% load

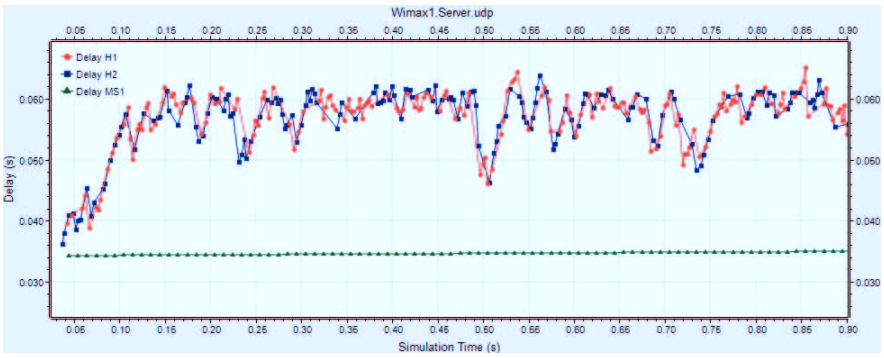


Figure 6.11: Delay with DiffServ support and 125% load

It is important to note from all the figures that the delays of the VoIP packets are constant for all the traffic loads when DiffServ is enabled, because, as mentioned before, PQ is used which provides premium service to the higher priority queue and all the packets will be scheduled first before providing the service to lower priority queue.

On the other hand, in the scenario with with DiffServ enabled network the traffic from node 1 and node 2 experiences delay larger than the one in non DiffServ network. Though in both the cases we have not marked the packets of nodes 1 and 2, so their traffic is treated as Best Effort. The increase of the delay is the cost of the preferential treatment provided to the VoIP packets.

Jitter and delays affect the QoS of the VoIP stream. From the point of view of perceived QoS, jitters and delays of few milliseconds are acceptable. According to ITU-T specification [45] one way delay should not be more than 150 ms while there is no precise limit for jitters. Figure 6.12 shows the values of jitters for different loads with and without DiffServ enabled network. It can be seen that as the network load increases, the jitter of the packets decreases: this is because when the network approaches saturation point the fluctuations of packet delays decrease.

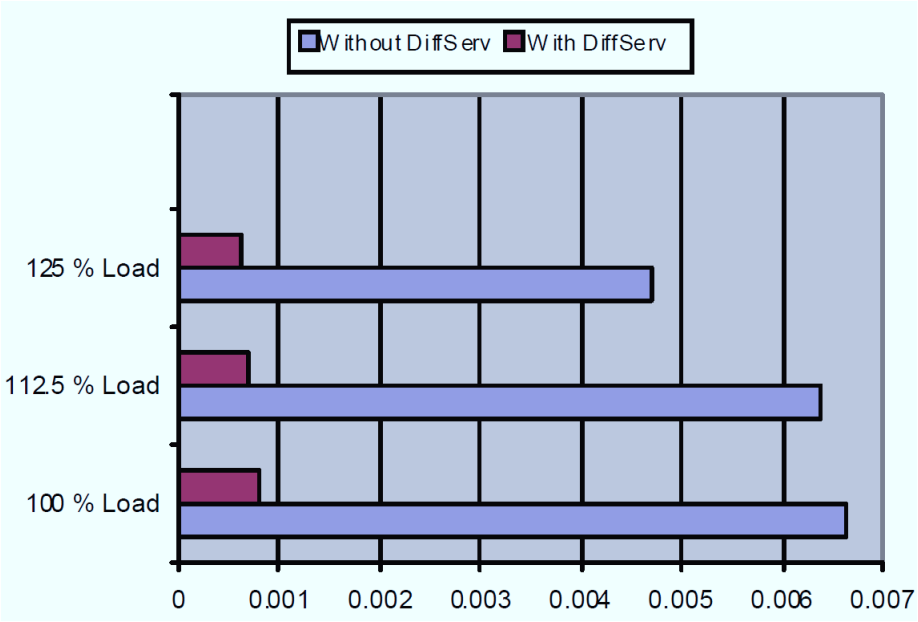


Figure 6.12: VoIP service jitters in networks with and without DiffServ.

6.5 Conclusions

The measurements in the test-bed network show that DiffServ network clearly provides better QoS for multimedia services. Though the effect of DiffServ in normal traffic conditions is not significant, it clearly performs better in congested network. Internet by default

is a Best Effort network and DiffServ improves the performance of the core network by providing the minimum required QoS level. Time sensitivity may not have a big impact on applications like FTP but it is a great limitation for real-time traffic. PQ should be used only when the amount of delay-sensitive traffic is small compared to overall traffic and needs to be processed as soon as possible, as in our test-bed where we were sending 64 Kbps of voice traffic while the total bandwidth of the core was around 5 Mbps.

Conclusions

In this PhD thesis several methodologies are proposed to provide better QoE to the end user for multimedia applications like video and voice. More emphasis was given to video for which R-D models and adaptation algorithms were developed for real-time and non real-time video applications, while error protection technique was also proposed for the non real-time R-D model to cope with the errors during transmission. For VoIP application, though WiMAX provide a dedicated class of service to prioritize the voice packets but it may lose significance if the intermediate networks cannot offer an equivalent preferential treatment. The R-D model proposed for non real-time video streams in Chapter 3 not only reduces the complexity by reducing a sequence dependent parameter but also the number of iterations and functions evaluations, thereby allowing minimum loss in the goodness parameters. A framework for the rate adaptation is also proposed which provides minimum rate required for each video in the transmission while providing fairness among the videos based on distortion. To cope with errors during transmission an unequal error protection scheme based on Reed-Solomon encoding with erasure correction is introduced in Chapter 4 with the aim to maintain high expected quality even in the presence of high packet error rate. This error protection framework is suitable for video application such as video on demand, IP-TV services and real-time streaming. A new R-D model for real time video streams is proposed in Chapter 5. This real-time R-D model is based on the R-D model proposed in Chapter 3, as it estimates the sequence dependent parameters through spatial and temporal indexes. These indexes are obtained from raw video sequences. Also

these index values are used to estimate the base layer and the highest enhancement layer rates. As the index values are obtained from raw videos the R-D model and the base layer and the enhancement layer rates are estimated before encoding the video stream thereby making it more suitable for real-time video streams. Chapter 6 discusses some network level issues related to prioritization of delay and time sensitive incoming VoIP traffic from WiMAX network. The measurement from the test bed network shows that by implementing DiffServ Network clearly provides better QoS. Prioritization is more beneficial when the bandwidth is limited and time critical data has to be transmitted within the minimum required standards.

Bibliography

[1] 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, no. 9, pp. 1103–1120, 2007.

[2] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13 No. 7, pp. 560–576, July 2003.

[3] *JSVM 9.19.11 Reference Software Febrary 2011*.

[4] B. Gorkemli, Y. Sadi, and A. M. Tekalp, "Effects of MGS fragmentation, slice mode and extraction strategies on the performance of SVC with medium-grained scalability," in *Proc. 17th IEEE Int Image Processing (ICIP) Conf*, 2010, pp. 4201–4204.

[5] 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 Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1186–1193, 2007.

[6] J. D. Cock, S. Notebaert, P. Lambert, and R. V. de Walle, "Architectures for fast transcoding of H.264/AVC to quality-scalable SVC streams," *IEEE Transactions on Multimedia*, vol. 11, no. 7, pp. 1209–1224, Nov. 2009.

[7] C. T. Hewage, Z. Ahmad, S. T. Worrall, S. Dogan, W. A. C. Fernando, and A. Kondoz, "Unequal error protection for backward compatible 3-D video transmission over WiMAX, circuits and systems," *IEEE Int. Symp*, pp. 125–128, 2009.

[8] Y.-K. Wang, S. Wenger, and M. M. Hannuksela, "System and transport interface of H.264/AVC scalable extension image processing," *Proc. IEEE Int. Conf. Digital Object Identifier*, pp. 165–168., 2006.

[9] J. Ohm, "Three dimensional subband coding with motion compensation," *IEEE Trans. Image Process*, vol. 3, no. 9, pp. 559–571, Sept. 1994.

[10] S. Choi and J. Woods, "Motion compensated 3-D subband coding of video," *IEEE Trans. Image Process*, vol. 8, no. 2, pp. 155–167, Feb. 1999.

[11] T. Wiegand, L. Noblet, and F. Rovati, "Scalable video coding for IPTV services," *IEEE Transactions on Broadcasting*, vol. 55, no. 2, pp. 527–538, 2009.

[12] Y. Wang, L.-P. Chau, and K.-H. Yap, "Joint rate allocation for multiprogram video coding using FGS," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 6, pp. 829–837, 2010.

[13] X. M. Zhang, A. Vetro, Y. Q. Shi, and H. Sun, "Constant quality constrained rate allocation for FGS-coded video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 2, pp. 121–130, 2003.

[14] M. Jacobs, J. Barbarien, S. Tondeur, R. Van de Walle, T. Paridaens, and P. Schelkens, "Statistical multiplexing using SVC," in *Proc. IEEE Int Broadband Multimedia Systems and Broadcasting Symp*, 2008, pp. 1–6.

[15] H. Mansour, V. Krishnamurthy, and P. Nasiopoulos, "Rate and distortion modeling of medium grain scalable video coding," in *Proc. 15th IEEE Int. Conf. Image Processing ICIP 2008*, 2008, pp. 2564–2567.

[16] M. Cesari, L. Favalli, and K. M. Folli, "Quality modeling for the medium grain scalability option of H.264/SVC," *Mobimedia*, September 7-9, 2009, London.

[17] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE Transactions On Selected Areas in Communications*, vol. 18, no. 6, pp. 1012–1032, 2000.

[18] H. Mansour, V. Krishnamurthy, and P. Nasiopoulos, "Channel adaptive multi-user scalable video streaming with unequal erasure protection," *Eight International Workshop on Image Analysis for Multimedia Interactive Services*, 2007.

[19] K. Do-Kyoung, S. Mei-Yin, and K. C. C. Jay, "Rate control for H.264 video with enhanced rate and distortion models," *IEEE Trans. Circuits Syst. Video Techn*, vol. vol. 17, no.5, pp. 517–529, 2007.

[20] T. Thang, J. Kim, J. Kang., and J. Yoo, "SVC adaptation: Standard tools and supporting methods," *Journal, Image Communication archive*, vol. Volume 24 Issue 3, March, 2009.

[21] R. Gupta, A. Pulipaka, P. Seeling, L. J. Karam, and M. Reisslein, "H.264 coarse grain scalable (cgs) and medium grain scalable (mgs) encoded video: A trace based traffic and quality evaluation," *Submitted 2011*.

[22] Y. Liu, Z. G. Li, and Y. C. Soh, "Rate control of H.264/AVC scalable extension," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 1, pp. 116–121, 2008.

[23] T. Schierl, H. Schwarz, D. Marpe, and T. Wiegand, "Wireless broadcasting using the scalable extension of H.264/AVC," in *Proc. IEEE Int. Conf. Multimedia and Expo ICME 2005*, 2005, pp. 884–887.

[24] H. Mansour, V. Krishnamurthy, and P. Nasiopoulos, "Channel aware multiuser scalable video streaming over lossy under-provisioned channels: Modeling and analysis," *IEEE Transactions on Multimedia*, vol. 10 No. 7, no. 7, pp. 1366–1381, 2008.

[25] E. Maani and A. K. Katsaggelos, "Unequal error protection for robust streaming of scalable video over packet lossy networks," pp. 407–416, 2010, circuits and Systems for Video Technology, IEEE Transactions on.

[26] Y. Sanchez, T. Schierl, C. Hellge, T. Wiegand, D. Hong, D. De Vleeschauwer, W. Van Leekwijck, and Y. Lelouedec, "Improved caching for HTTP-based video on demand using scalable video coding," in *Proc. IEEE Consumer Communications and Networking Conf. (CCNC)*, 2011, pp. 595–599.

[27] T. Schierl, C. Hellge, S. Mirta, K. Gruneberg, and T. Wiegand, "Using H.264/AVC-based scalable video coding (svc) for real time streaming in wireless IP networks," in *Proc. IEEE Int. Symp. Circuits and Systems ISCAS 2007*, 2007, pp. 3455–3458.

[28] D. Munaretto, D. Jurca, and J. Widmer, "A fast rate-adaptation algorithm for robust wireless scalable streaming applications," in *Proc. IEEE Int. Conf. Wireless and Mobile Computing, Networking and Communications WIMOB 2009*, 2009, pp. 246–251.

[29] H. Cheng-Hsin and M. Hefeeda, "On the accuracy and complexity of rate-distortion models for fine grained scalable video sequences," *ACM Transaction on Multimedia Computing, Communications and Applications*, December 2006.

[30] S. Cicalo, A. Haseeb, and V. Tralli, "Fairness-oriented multi-stream rate adaptation using scalable video coding," *Elsevier Signal Processing: Image Communication*, Feb. 2012.

[31] P. Chaffe-Stengel and D. N. Stengel, *Working With Sample Data: Exploration and Inference*. Business Expert Press, August 05, 2011.

[32] Y. Dodge and J. Jureckova, *Adaptive Regression*, BPOD, Ed. Springer, 1 April, 2000.

[33] F. De Simone, M. Tagliasacchi, M. Naccari, S. Tubaro, and T. Ebrahimi, "A H.264/AVC video database for the evaluation of quality metrics," in *Proc. IEEE Int Acoustics Speech and Signal Processing (ICASSP) Conf*, 2010, pp. 2430–2433.

[34] ITU-T, "Subjective video quality assessment methods for multimedia application, recommendation," *ITU-T*, p. 910, September 1999.

[35] M. Koivula, M. Taramaa, and P. Ruska, "Differentiated services and vertical handovers supporting multimedia in heterogeneous networks," *18th annual IEEE international symposium on Personal, Indoor and Mobile Radio Communication*, 2007.

[36] N. C. et al, "A QoS model based on NSIS signaling applied to IEEE 802.16 networks," *IEEE CCNC proceedings*, 2008.

[37] Y.T.Mai, C.C.Yang, and Y. Lin, "Design of the cross-layer QoS framework for the IEEE 802.16 PMP networks," *IEICE Trans. Commun*, vol. E91-B, NO.5, MAY 2008.

[38] H. Haffajee and H. A. Chan, "Low-cost QoS-enabled wireless network with interworked WLAN and WiMAX," *Proceedings of the First IEEE International Conference on Wireless Broadband and Ultra Wideband Communications*, March 13-16 2006.

[39] E. Angori and et al, "Extending WiMAX technology to support end to end QoS guarantees," www.ist-weird.eu/documents., Tech. Rep.

[40] N. Muhammad, D. Chiavelli, D. Soldani, and M. Li, *QoS and QoE Management in UMTS Cellular Systems*. John Wiley & Sons, Ltd. ISBN: 0-470-01639-6, pp. 1-8, 2006.

[41] D. Sharma and R. Singh, "QoS & QoE managment in wireless communicatin system," *International Journal of Engineering Science and Technology*, vol. 3 No. 3, pp. 2385–2391, March 2011.

[42] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An architecture for differentiated services," *RFC 2475*, Dec 1998.

[43] *IEEE Standard for Local and Metropolitan Area Networks. Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, IEEE STD 802.16 2004, October 2004 Std.

[44] *OMNeT++ Discrete event Simulator www.omnetpp.org*.

[45] V. S. Kaulgud and S. Mondal, "Exploiting multihoming for low latency handoff in heterogeneous networks," *Proceedings of 8th International Conference on Telecommunications*, vol. Volume 1, pp. 49–55, 2005.

[46] A. Haseeb and V. Tralli, "Qos performance analysis for voip traffic in heterogeneous networks with wimax access," *13th IEEE Conference on Advance Communication Technologies (ICACT), Phoenix Park*, February 2011, South Korea.

[47] S. Cicalo, A. Haseeb, and V. Tralli, "Multi-stream rate adaptation using scalable video coding with medium grain scalability," *7th International ICST Mobile Multimedia (MOBIMEDIA) Communication Conference*, September 2011, Cagliari, Italy.

[48] A. Haseeb, M. G. Maritini, S. Cicalo, and V. Tralli, "Rate distortion modeling for real-time mgs coding and adaptation," *8th IEEE Conference on Wireless communication-Wireless Advanced (WIAD)*, June 2012, London. (Submitted).

Il tuo indirizzo e-mail

abdul.haseeb@unife.it

Oggetto:

Dichiarazione di conformità della tesi di Dottorato

Io sottoscritto Dott. (Cognome e Nome)

Haseeb Abdul

Nato a:

Karak

Provincia:

Pakistan

Il giorno:

10/04/1982

Avendo frequentato il Dottorato di Ricerca in:

Scienze dell'Ingegneria

Ciclo di Dottorato

24

Titolo della tesi (in lingua italiana):

"Quality of Experience" e Tecniche di Adattamento per le Comunicazioni Multimediali

Titolo della tesi (in lingua inglese):

Quality of Experience and Adaptation Techniques for Multimedia Communications

Tutore: Prof. (Cognome e Nome)

Tralli Velio

Settore Scientifico Disciplinare (S.S.D.)

ING-INF/03

Parole chiave della tesi (max 10):

Multimedia, Adaptation, Video, Quality, Communication

Consapevole, dichiara

CONSAPEVOLE: (1) del fatto che in caso di dichiarazioni mendaci, oltre alle sanzioni previste dal codice penale e dalle Leggi speciali per l'ipotesi di falsità in atti ed uso di atti falsi, decade fin dall'inizio e senza necessità di alcuna formalità dai benefici conseguenti al provvedimento emanato sulla base di tali dichiarazioni; (2) dell'obbligo per l'Università di provvedere al deposito di legge delle tesi di dottorato al fine di

assicurarne la conservazione e la consultabilità da parte di terzi; (3) della procedura adottata dall'Università di Ferrara ove si richiede che la tesi sia consegnata dal dottorando in 2 copie di cui una in formato cartaceo e una in formato pdf non modificabile su idonei supporti (CD-ROM, DVD) secondo le istruzioni pubblicate sul sito: http://www.unife.it/studenti/dottorato alla voce ESAME FINALE – disposizioni e modulistica; (4) del fatto che l'Università, sulla base dei dati forniti, archivierà e renderà consultabile in rete il testo completo della tesi di dottorato di cui alla presente dichiarazione attraverso l'Archivio istituzionale ad accesso aperto "EPRINTS.unife.it" oltre che attraverso i Cataloghi delle Biblioteche Nazionali Centrali di Roma e Firenze; DICHIARO SOTTO LA MIA RESPONSABILITA': (1) che la copia della tesi depositata presso l'Università di Ferrara in formato cartaceo è del tutto identica a quella presentata in formato elettronico (CD-ROM, DVD), a quelle da inviare ai Commissari di esame finale e alla copia che produrrò in seduta d'esame finale. Di conseguenza va esclusa qualsiasi responsabilità dell'Ateneo stesso per quanto riguarda eventuali errori, imprecisioni o omissioni nei contenuti della tesi; (2) di prendere atto che la tesi in formato cartaceo è l'unica alla quale farà riferimento l'Università per rilasciare, a mia richiesta, la dichiarazione di conformità di eventuali copie; (3) che il contenuto e l'organizzazione della tesi è opera originale da me realizzata e non compromette in alcun modo i diritti di terzi, ivi compresi quelli relativi alla sicurezza dei dati personali; che pertanto l'Università è in ogni caso esente da responsabilità di qualsivoglia natura civile, amministrativa o penale e sarà da me tenuta indenne da qualsiasi richiesta o rivendicazione da parte di terzi; (4) che la tesi di dottorato non è il risultato di attività rientranti nella normativa sulla proprietà industriale, non è stata prodotta nell'ambito di progetti finanziati da soggetti pubblici o privati con vincoli alla divulgazione dei risultati, non è oggetto di eventuali registrazioni di tipo brevettale o di tutela. PER ACCETTAZIONE DI QUANTO SOPRA RIPORTATO

Firma del dottorando

Ferrara, li 13/03/2012 (data)

Firma del Dottorando A. Haseeb

Firma del Tutore

Visto: Il Tutore Si approva Firma del Tutore