

Efficient Transmission of Error Resilient H.264 Video over Wireless Links

by

Ashfiqua Tahseen Connie

B. Sc., Bangladesh University of Engineering and Technology, Bangladesh, 2004

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

Master of Applied Science

in

THE FACULTY OF GRADUATE STUDIES

(Electrical & Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

September 2008

© Ashfiqua Tahseen Connie, 2008

Abstract

With the advent of telecommunication technology, the need to transport multimedia content is increasing day by day. Successful video transmission over the wireless network faces a lot of challenges because of the limited resource and error prone nature of the wireless environment. To deal with these two challenges, not only the video needs to be compressed very efficiently but also the compression scheme needs to provide some error resilient features to deal with the high packet loss probability. In this thesis, we have worked with the H.264/ Advanced Video Coding (AVC) video compression standard since this is the most recent and most efficient video compression scheme. Also H.264 provides novel error resilient features e.g. slicing of the frame, Flexible Macroblock Ordering (FMO), data partitioning etc.

In this thesis, we investigate how to utilize the error resilient schemes of H.264 to ensure a good quality picture at the receiving end. In the first part of the thesis, we find the optimum slice size that will enhance the quality of video transmission in a 3G environment. In the second part, we jointly optimize the data partitioning property and partial reliability extension property of the new transport layer protocol, Stream Control Transmission Protocol (SCTP). In the third and last part, we focus more on the network layer issues. We obtain the optimum point of application layer Forward Error Correction (FEC) and Medium Access Control (MAC) layer retransmission in a capacity constrained network. We assume that the bit rate assigned for the video application is more than the video bit rate so that the extra capacity available can be used for error correction.

Table of Contents

Abstract.....	ii
Table of Contents	iii
List of Tables	vi
List of Figures.....	vii
List of Acronyms	ix
Acknowledgements	xiv
1 Introduction.....	1
2 Background	4
2.1 Video Compression	4
2.1.1 Basic of Video Compression	4
2.1.2 H.264/ Advanced Video Coding Standard.....	8
2.2 3G Networks: A Brief Overview	15
2.3 Stream Control Transmission Protocol (SCTP).....	16
2.4 Literature Review	22
2.4.1 Previous Works on Video Transmission over Wireless Environments	22
2.4.2 Previous Works on Video Transmission using SCTP as the Transport Layer Protocol	23
2.4.3 Previous Works on Video Transmission Employing Unequal Error Protection (UEP) by means of FEC and Retransmission	24
3 Video Packetization Techniques for Enhancing H.264 Video Transmission over 3G Networks.....	27

3.1	Overview	27
3.2	Packetization Steps in 3G Networks.....	27
3.3	Packet Size Optimization in 3G Networks.....	28
3.4	Summary.....	36
4	Transport of Data-Partitioned H.264 Video Using SCTP.....	38
4.1	Overview	38
4.2	Utilizing SCTP Reliability Features to Transport H.264 Video Data Partitions .	38
4.3	Performance Evaluations and Discussions	42
4.4	Summary.....	48
5	Efficient Utilization of Error Protection Techniques for Transmission of Data-Partitioned H.264 Video in a Capacity Constrained Network	49
5.1	Overview	49
5.2	Application Layer FEC: A Brief Description.....	50
5.3	Optimal Resource Allocation: FEC versus Retransmission.....	51
5.4	Optimal Resource Allocation: Non-Data-Partitioned Video.....	57
5.4.1	Loss-Distortion Model.....	57
5.4.2	Optimal Resource Allocation	59
5.5	Optimal Resource Allocation: Data-Partitioned Video	64
5.5.1	Loss-Distortion Model.....	64
5.5.2	Optimal Resource Allocation	64
5.6	Summary.....	73
6	Conclusions and Future Work.....	75

6.1	Conclusions.....	75
6.2	Future Work.....	76
	Bibliography	78

List of Tables

Table 2.1 Error concealment scheme where different types of partitions are available [4]	14
Table 5.1 Optimum point at different loss ratios and available extra capacity.....	67
Table 5.2 Optimum point at loss ratio of 40% and different capacity allocation.	73

List of Figures

Figure 2.1 General block diagram of a video encoder.....	5
Figure 2.2 Flexible Macroblock Ordering: Dispersed	12
Figure 2.3 Head of Line (HoL) blocking problem in TCP	17
Figure 2.4 Multihoming property of SCTP.....	19
Figure 2.5 Multistreaming property of SCTP	20
Figure 3.1 Packetization in 3GPP protocol stack.....	29
Figure 3.2 Framework of the experiment	31
Figure 3.3 For the video sequence Coastguard: (a) Bit rate versus slice size, (b) Loss ratio versus slice size and (c) PSNR(Y) versus slice size.	33
Figure 3.4 For the video sequence Foreman: (a) Bit rate versus slice size, (b) Loss ratio versus slice size and (c) PSNR(Y) versus slice size.	34
Figure 3.5 For the video sequence Hall Monitor: (a) Bit rate versus slice size, (b) Loss ratio versus slice size and (c) PSNR(Y) versus slice size.....	35
Figure 4.1 Simulation Topology.....	44
Figure 4.2 Impact of loss of partition B (a: Foreman, b: Coastguard).....	45
Figure 4.3 PSNR versus loss percentage (video: Foreman)	46
Figure 4.4 PSNR versus loss percentage (video: Coastguard)	47
Figure 5.1 RS FEC scheme.....	51
Figure 5.2 Framework of our scheme	53
Figure 5.3 Loss distortion model for non-DP video (Foreman sequence).....	58
Figure 5.4 (a) FEC erasure rate versus redundancy, (b) extra bit rate needed for FEC versus redundancy, and (c) extra bit rate needed for retransmission versus redundancy.	60

Figure 5.5 PSNR versus application layer redundancy and number of retransmissions for different loss ratios and capacity; (a) p=10% and C= 10% more, (b) p=20% and C= 24% more, (c) p=30% and C= 39% more, and (d) p=40% and C= 60% more.....	63
Figure 5.6 Loss distortion model for a DP video (Foreman Sequence): (a) Measured values, and (b) Model values.	65
Figure 5.7 (a) Retransmission bit rate versus number of retransmissions for partition BC, and (b) PSNR versus number of retransmissions for partition BC.....	68
Figure 5.8 (a) PSNR versus number of retransmissions for partition BC, and (b) PSNR versus application layer redundancy for partition A.....	70
Figure 5.9 PSNR versus number of retransmission for partition BC.	72
Figure 5.10 PSNR versus redundancy for partitions A and BC.	72
Figure 5.11 PSNR versus Retransmission for Partition BC.	73

List of Acronyms

3G	Third Generation
3GPP	Third Generation Partnership project
API	Application Programming Interface
ARQ	Automatic Repeat reQuest
ACK	Acknowledgement
AVC	Advanced Video Coding
B-frame	Bidirectional frame
BER	Bit Error Rate
BLER	Block Error Rate
CBP	Coded Block Pattern
CDMA	Code Division Multiple Access
CSMA/CA	Carrier Sense Multiple Access/Collision Avoidance
DCH	Dedicated Channel
DCT	Discrete Cosine Transform
DoS	Denial of Service

DP	Data Partitioning
DSCH	Downlink Shared Channel
FACH	Forward Access Channel
FEC	Forward Error Correction
FMO	Flexible Macroblock Ordering
GOP	Group Of Picture
HoL	Head of Line
I-frame	Intra frame
IDCT	Inverse Discrete Cosine Transform
IDR	Instantaneous Decoder Refresh
IP	Internet Protocol
ISO/IEC	International Standardization Organization/ International Electrotechnical Commission
ITU-T	International Telecommunication Unit- Telecommunication
JVT	Joint Video Team
MB	Macro Block

MC	Motion Compensation
ME	Motion Estimation
MPEG	Moving Pictures Experts Group
MSE	Mean Squared Error
MTU	Maximum Transfer Unit
MV	Motion Vector
NAL	Network Abstraction Layer
NALU	Network Abstraction Layer Unit
NS	Network Simulator
P-frame	Predicted frame
PDCP	Packet Data Convergence Protocol
PDU	Packet Data Unit
PLR	Packet Loss Ratio
PR-SCTP	Partial Reliability-SCTP
PSNR	Peak Signal to Noise Ratio
PSC	Packet Switch Conversational
QCIF	Quarter Common Intermediate Format

QP	Quantization Parameter
RLC	Radio link Control
RoHC	Robust Header Compression
RRC	Radio Resource Control
RS	Reed-Solomon
RTP	Real Time Protocol
RTT	Round Trip Time
SACK	Selective ACK
SCTP	Stream Control Transmission Protocol
SDU	Service Data Unit
SI	Switching Intra
SP	Switching Predicted
SSN	Stream Sequence Number
TCP	Transmission Control Protocol
TD-SCDMA	Time Division-Synchronous CDMA
TrCh	Transmission Channel
TSN	Transmission Sequence Number

UDP	User Datagram Protocol
UEP	Unequal Error Protection
UMTS	Universal Mobile Telecommunications System
VCEG	Video Coding Experts Group
VPLR	Video PLR
VLC	Variable Length Coding
WCDMA	Wideband CDMA
WLAN	Wireless Local Area Network

Acknowledgements

At first I would like to thank the Almighty Allah for giving me the chance of having such a good opportunity.

I would like to express my gratitude to the persons without whom it would not be possible for me to complete my thesis. At first I want to thank The Canadian Bureau of International Education (CBIE) for selecting me as a Commonwealth Scholar and fulfilling my dream to study abroad. Then most importantly, I want to express my sincere gratitude to my supervisors, Dr. Victor C. M. Leung and Dr. Panos Nasiopoulos. Dr. Leung gave me enough freedom to choose my research area and go on with my thinking. Thanks a lot to him to keep confidence on me. He always encouraged me and helped me a lot in proof-reading my thesis. Dr. Nasiopoulos was involved in every detail of my work. I must admit that his encouragement helped me to pass the deadlock period of my thesis. Another person who helped me profoundly with suggestions, encouragement and being available whenever I need him is Dr. Yaser Pourmohammadi Fallah. I can never express in words how much thankful I am to him.

I am really proud to be a member of the Signal, Image and Multimedia Processing (SIMPL) lab. My colleagues in lab are bright and helpful. Thanks to Qiang Tang, Di Xu, Xin Yi Yong, Zicogn Mai, Colin Doutre, Mahsa Pourazad, Angshul Majumder, and Joy Zhang for their constant support and encouragement. A very special thanks to Hassan Mansour for his insightful suggestions and the fruitful discussions I had with him.

Finally, I would like to thank my parents and sister for their support and encouragement.

I wish to express my earnest gratitude to my husband, who was always with me and always encouraged me throughout the whole journey of this study.

1 Introduction

Video transmission has become a major application for the wireless systems, such as Wireless Local Area Networks (WLAN), Third Generation (3G) cellular networks etc. The wireless environment is very error prone due to the effect of fading, shadowing, interference etc. and also the radio resources such as transmission bandwidth and power are limited. That is a big problem for video transmissions since the bandwidth of the video is usually large. Therefore the researchers turn to video compression to address this problem. It is obvious that good compression efficiency is the main requirement for a video coding standard. However, keeping in mind the lossy nature of the wireless environment and the fact that compressed video data is very sensitive to packet loss, the video compression standard should provide good error resilient features. H.264/ Advanced Video Coding (AVC) [1], the most recent video coding standard, fulfills these requirements. It provides very good compression efficiency, almost twice compared to the previous standards, facilitates the transmission of the video data in the packet-switched network through the concept of Network Abstraction Layer (NAL), and moreover it provides good error resilient features. In this thesis, we investigate how to effectively utilize the error resilient features in an error prone wireless environment.

In the first part of this thesis, we work with one of the error resilient features: slicing. Slicing allows a video frame to be divided into more than one slice or video packet and the slices can be decoded independent of each other. These slices or NAL units (NALU) are packetized according to the 3G Partnership Project (3GPP) protocol stack and the optimum slice size depends on the underlying structures of the 3GPP standard. In the Medium Access Control (MAC) layer, 3GPP allows aggregation (in case the slices are smaller than a Radio Link

Control (RLC) frame) and fragmentation (when the slice size is larger than the RLC frame size). These two mechanisms along with block error rate of the wireless link affects the effective loss ratio and as a result, the quality.

In the second part of the thesis, we work with another important error resilient property of H.264, data partitioning property. Data Partitioning allows the slice to be divided into 3 partitions of different importance. Partition A contains the most important information: the header information, motion vectors, etc. Partition B contains the intra information whereas partition C contains the inter information. Partition A can be decoded independent of partition B and C but the later two needs partition A to be decoded. This different importance levels motivates us to apply different reliability levels on their transport. We choose to realize the different reliability levels at the transport layer. The most widely used transport layer protocol, Transmission Control Protocol (TCP) cannot be used for multimedia content delivery because its stringent reliability can cause large delay variations. Usually User Datagram Protocol (UDP) is used for video transmission, but this results in sub-optimal performance since UDP is unreliable and has no congestion control or flow control mechanism. It needs some application layer mechanism on top of it to make the video transmission a success e.g. Real Time Protocol (RTP) is widely used on top of UDP. Considering all these facts, we choose Stream Control Transmission Protocol (SCTP) [2] as our transport layer protocol. SCTP has all the good features of TCP and UDP; in addition to this, it also possesses some novel features such as multihoming, multistreaming, and partial reliability extension. We use multistreaming along with the partial reliability extension to set different reliability levels to different partitions. We also compare this to the case when the same percentage of non-data-partitioned packets is protected by the partial reliability feature.

In the above, we assume that the capacity allocated for video application is sufficient to allow retransmissions of the packets that require reliable delivery. This motivates us to find a way to efficiently utilize the available capacity. We assume that the capacity allocated for video transmission is more than the video bit rate, so the extra capacity can be used for protecting the packets against the high loss rate of the wireless environment. For error protection, we consider both Forward Error Correction (FEC) and MAC layer retransmissions. We obtain the optimum point between these two schemes that will enhance the picture quality in a capacity constrained error prone wireless environment.

The contribution of this thesis work can be summarized as below:

1. Finding the optimal packet size for conversational video transmission in a 3GPP network.
2. Joint optimization between the error resilient property of H.264 and the PR-SCTP features.
3. Efficient utilization of the available extra capacity for video transmission by means of FEC and MAC layer retransmissions.

The rest of this thesis is organized as follows. Chapter 2 provides background information on H.264 video coding standards, 3G networks, SCTP and the existing work on video transmission over wireless networks. In Chapter 3, a video packetization technique to enhance video quality in 3GPP network is presented. Joint optimization between the error resilience features of H. 264 and the SCTP features are presented in Chapter 4. Chapter 5 describes the optimum point between FEC and MAC layer retransmissions in a capacity constrained network to obtain the highest achievable quality. Finally, conclusions are presented in Chapter 6.

2 Background

In this chapter, we provide the background information on the fundamental of video compression and the H.264 standard, the SCTP and also the existing works in video transmission in wireless environment. The orientation of this chapter is as follows: the basic of video compression with a brief summary of the latest video coding standard, H.264 is presented in Section 2.1. Section 2.2 and 2.3 depicts a brief overview of 3G networks and SCTP. A detailed summary of the existing research work is presented in Section 2.4.

2.1 Video Compression

The basic idea of compression is to remove redundancy. Video signals contain a large amount of redundancy in spatial (similarities in the pixel domain), temporal (similarities in time domain; e.g., two consecutive frames can be very similar if there is not much motion in the sequence and the frame rate is high) and statistical domains (statistical redundancy refers to the fact that for a given sequence, some pixel values are more likely; e.g., a natural image is more likely to have green or blue tones than bright red). To have a lossless compression, only statistical redundancy is removed. But lossless compression can only provide a moderate amount of compression. So video compression algorithms generally perform lossy compression by maintaining a good trade-off between distortion and compression gain.

2.1.1 Basic of Video Compression

A video encoder consists of three modules: a temporal module to reduce the redundancy between consecutive frames, a spatial module to reduce the redundancy in the same frame

and an entropy encoder to reduce the statistical redundancy. Figure 2.1 shows a general video encoder consisting of these three modules.

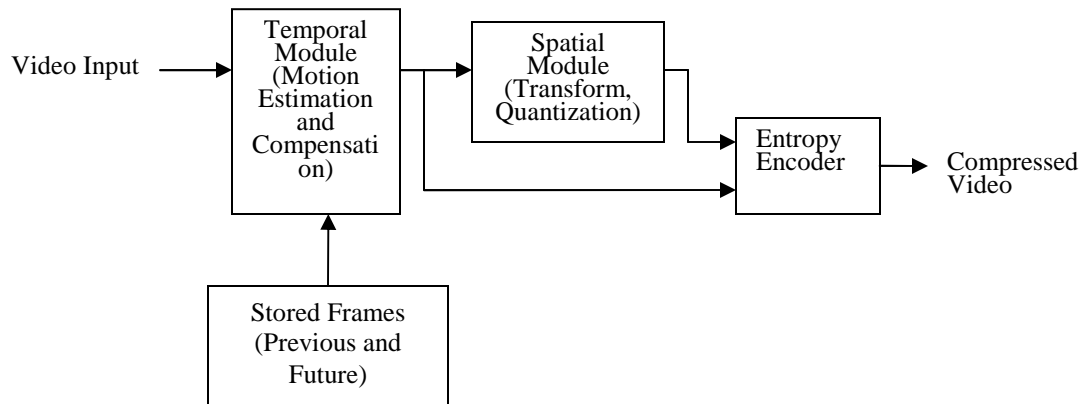


Figure 2.1 General block diagram of a video encoder

2.1.1.1 Temporal Module

The temporal module utilizes the redundancy of the successive frames in a video sequence. In the simplest form, the frame prior to the current frame (at time $t - 1$) can be considered as the predictor of the current frame (at time t). This predictor frame is subtracted from the current frame and a residual is formed. The residual frame is encoded and sent to the decoder. The efficiency of the compression depends on the energy available in the residual frame. This form of simple prediction works fine when the sequence involves little motions with an almost static background. However, if the consecutive frames involve a motion of a large part of the object in the sequence or camera panning, then the performance of this kind of simple prediction is not good. Due to motion, the pixel positions in the frames differ. To have minimum energy in the residual frame, the position of the corresponding pixel in the reference frame (motion estimation) should be determined properly. There are mainly two

approaches. The first one estimates the pixel trajectory between the successive frames, known as the optical flow. This scheme is computationally expensive and time consuming. Another efficient approach is the block based motion estimation. In this approach, for a block of $a \times b$ samples in the current frame, an area in the reference frame is searched to find the block which provides the best match. Usually the search area is centered on the block of $a \times b$ samples in the current frame. This process is known as the Motion Estimation (ME). Then the predictor block is subtracted from the current block and a residual block is formed. This is known as Motion Compensation (MC). The offset between the current block and the position of the best matched region is known as the motion vector and must be transmitted along with the coded residual block to the decoder.

2.1.1.2 Spatial Module

To reduce spatial redundancy efficiently requires fewer correlations between neighboring image samples. However, image samples of natural images are highly correlated and thus difficult to compress. The residual frame generated by the motion compensated prediction contains less energy than the original image and less correlations between the image samples, so that they can be compressed more efficiently. So the purpose of the spatial module is to decorrelate the image data so that they can be efficiently compressed. Transformation to the frequency domain serves this purpose. This transformation has another advantage since it reflects the property of the human visual system, which is more sensitive to the low frequency component. Also, natural images contain more low frequency components than high frequency components. Transformation to the frequency domain separates the low frequency information from the high frequency information.

The most commonly used transforms are block based (Discrete Cosine Transform: DCT) and image based (wavelet) transforms. DCT is used in the most recent video compression standard, H.264. DCT is generally applied to block sizes ranging from 8×8 to 64×64 and transforms the input pixel values to frequency coefficients. This transform is fully reversible and the pixel values can be reconstructed by performing the Inverse DCT (IDCT), but the advantage here is that, a subset of the DCT coefficients is enough to generate the pixel values to a reasonable accuracy.

The DCT coefficients are then quantized. In quantization, the coefficients are scaled by a given scaling factor and then rounded to the nearest integer value. This step removes the insignificant or near zero DCT coefficients. Since quantization is a lossy process, inverse quantization will not generate the exact coefficient values. The accuracy depends on the step size. For large step sizes, the accuracy is poor but compression gain is high. On the other hand, small step sizes result in better accuracy but reduced compression gain. So a trade-off should be maintained between these two.

2.1.1.3 Entropy Encoder

The quantized coefficients are reordered such that the significant values are clustered together at the beginning of the sequence. The advantage of having this reordering is that the zero sequence can be compressed using run length coding. Run length coding replaces the zero sequence by a special symbol and the length of the zero sequence. After this, the symbols are entropy coded and a compressed bitstream is generated suitable for transmission and storage. Variable Length Coding (VLC) is used for this purpose. VLC represents the most common symbols with short code words whereas rare symbols are represented with long code words. Huffman coding and arithmetic coding are the most widely used VLC

methods. One disadvantage of using VLC is that transmission errors can cause desynchronization of the decoder and errors will propagate until the next synchronization point is reached. As a measure of prevention for this problem, synchronization markers are inserted in the bitstream at regular intervals.

The entropy coded video is transmitted to the decoder. The decoder performs all these tasks in the reverse direction (entropy decoding, reordering, IDCT, ME and MC) and the video is reconstructed. The quality of the reconstructed video depends on several things e.g. Quantization Parameter (QP), algorithm used for ME and MC etc.

2.1.2 H.264/ Advanced Video Coding Standard

In the video compression world, two organizations, namely, International Telecommunication Union- Telecommunication (ITU-T) Video Coding Experts Group (VCEG) and International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) Moving Pictures Experts Group (MPEG), have been dominating the task of standardization. The standards developed by ITU-T VCEG are H.120 (1984-1988), H.261 (1990+), H.263 (1995-2000+) whereas MPEG-1 (1993) and MPEG-4 Visual (1998-2001+) are by ISO/IEC MPEG. H.262/MPEG-2 and the most recent standard, H.264/ AVC is developed by the Joint Video Team (JVT), which is a collaboration of VCEG and MPEG. H.264 achieves 50%, 47% and 24% coding gain over MPEG-2, H.263 baseline and H.263 high profile standards, respectively [3]. The cost that needs to be paid for this gain is the increased implementation complexity at the encoder (by a factor of more than 1) and decoder (by a factor of 2). Since H.264/AVC is the recent and most efficient standard, we will focus on this standard.

H.264 follows the principles of compression described in Section 2.1.1. It also incorporates a number of new coding features that enhance its performance. For example, H.264 uses an integer transform that approximates the DCT, whereas the previous standards are based on DCT calculated with floating point arithmetic. Again for the intra coding part, the H.264 supports different prediction modes, e.g., DC prediction mode where the block is predicted from the average value of the surrounding pixels, or each pixel can be predicted from a pixel of outside block residing directly above it, to the left or in a diagonal direction. Seven block types ranging from 4×4 to 16×16 can be used for inter prediction. In motion compensation, H.264 allows sub pixel motion accuracy; luma and chroma pixels have quarter and $1/8$ pixel resolution resulting in increased precision of motion vectors.

The above mentioned improvements are on the signal processing aspect of the codec. However, the high supremacy in this domain does not necessarily ensure the network friendliness of the coded video data. With this view in mind, the H.264 introduces the idea of two conceptual layers, the Video Coding Layer (VCL), responsible for efficient video compression and the NAL, responsible for efficient transmission of the coded data. In general, abstraction layer hides the implementation complexity of a particular set of functionality. The NAL works as an interface between the VCL and the outside world and enables efficient transmission of data from the VCL in a broad variety of systems and maps VCL to different kind of applications. The VCL data is transported in NAL Unit (NALU). Each NALU is comprised of a header and payload. The payload is, in most of the cases, the slice consisting of the Macro Blocks (MB). A MB is a 16×16 sample region in the frame containing the coded data. The NALU can be of (i) byte stream format, in which case a start code prefix indicates the start of a new data or of (ii) packet format. The NALU can be

categorized as VCL NALU and Non-VCL NALU. Non-VCL NALU contains meta data such as parameter sets or other supplemental enhancement information. Also The NALUs can be aggregated to form a compound packet or can be fragmented if the size is too big. NALUs of a specific format may be grouped together into an access unit. A sequence of access units forms a coded video stream, which can be decoded independent of any other coding sequence.

2.1.2.1 Error Resilience Features of H.264

Achieving a high compression efficiency and network friendliness have become the major goals for the H.264 standard. Apart from this, considering the lossy nature of the wireless environment, the standard should provide good error resilience features. A brief description of the error resilience features of H.264 is presented below:

- **Intra Placement:** The purpose of having intra placement is to reduce the error drifting problem in case of a packet loss. Intra coded pictures require a large number of bits and so there is a limitation on the number of intra coded pictures per video frame. H.264 has two types of intra placement: the Instantaneous Decoder Refresh (IDR) picture and the Intra (I) frame. The IDR pictures are made up of I or Switching Intra (SI) slices and clears the short term memory buffer. No pictures following the IDR picture can refer to a frame for inter prediction that is coded prior to the IDR picture. Thus they eliminate the error propagation and also provide a resynchronization point for the subsequent slices. The first picture of a coded sequence is always an IDR picture. An intra picture is composed of only I slices and cancels the error drift for the duration of that picture. However, if the subsequent pictures use any reference picture that is coded prior to the I frame, then error

propagation can again be re-established even for the error free transmission of the I and the subsequent pictures.

- **Slices:** A picture can be divided to one or more slices. Each slice has a header and a data portion. The data portion contains one MB or a sequence of MBs. The slices are independently decodable and thus important to prevent propagation of errors. A slice can be categorized as I, Predicted (P) or Bidirectional (B) slice depending on the nature of MBs belonging to the slices. In I slice, all the MBs are coded using intra prediction. In P and B slices, in addition to the coding types of I slices, MBs can be coded using inter prediction. P slices use the previously coded pictures as reference for inter prediction whereas B slices may use one or more than one reference pictures coded before and after the current picture in temporal order. In addition, the H.264 provides Switching Predicted (SP) and SI slices which enables efficient switching between video streams to cope up with different bit rates if needed. The SP slice allows the switching between different pre coded pictures whereas the SI slice permits the exact match of a MB in a P slice.
- **Parameter Set Concept:** The parameter sets contain information that is applicable to a large number of coded pictures. A sequence parameter set contains the information related to a sequence of pictures, e.g., sequence parameter set identity, limits of frame numbers, number of reference frames used for decoding, whereas a picture parameter set contains information related to the slice of a picture, e.g., picture parameter set identity, number of slice groups to be used, slice group map type. Since parameter sets contain important information to decode a picture, they should be sent reliably in an error prone environment. They can be sent to the decoder ahead of the slices or

they can be sent in band with the slices but with more protection to ensure that they reach the decoder accurately.

- **Flexible Macroblock Ordering (FMO):** The FMO mode refers to the grouping of slices in a flexible way. Without FMO, slices are scanned in a raster order. But use of FMO opens the possibility of different scanning orders, e.g., interleaved, dispersed, foreground and background, box out, raster and wipe maps. The MBs in a slice are scanned in raster order. The allocation of MB to different slice groups is determined by 'MB to slice group map'. FMO allows two neighboring MBs to be allocated in different slice groups and MBs belonging to each slice groups are packed in different NALUs. For example, in Fig. 2.2, in dispersed FMO, the MBs are assigned to two slice groups, 0 and 1. If the packet containing MBs of slice group 0 is lost; then the lost MBs have correctly received neighboring MBs belonging to slice group 1 and so can be concealed more efficiently. The downside of using this mode is lower coding efficiency and a higher delay.

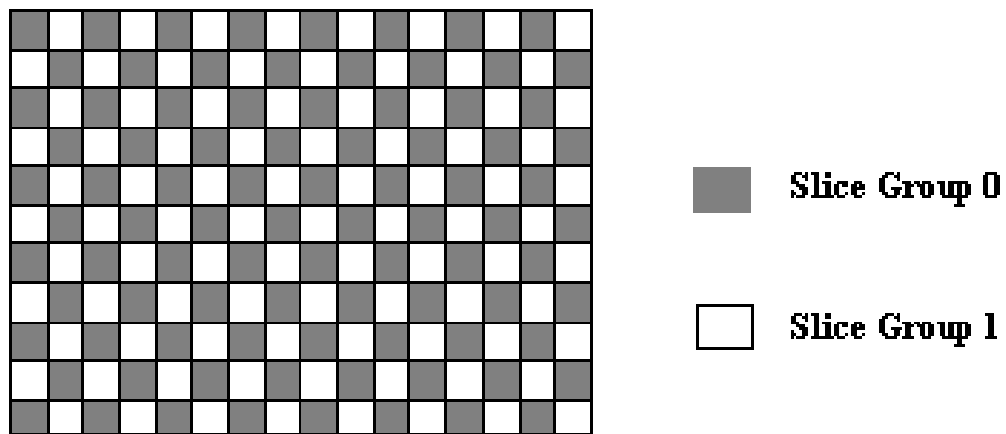


Figure 2.2 Flexible Macroblock Ordering: Dispersed

- **Redundant Slices (RS):** Redundant representation of the MBs may be sent in addition to the coded MB of the slice itself. Redundant slices use a different coding parameters e.g. the quantization parameter for RS is higher than that of the primary slices. It is used by the decoder only when the primary slice is lost and discarded otherwise.
- **Data Partitioning (DP):** All the information available in a slice is not equally important. Header information is more important in the sense that if the coded MBs are lost but the header information is available, then the decoder can perform a better concealment compared to the case where the whole slice (both header and coded MBs) is lost. H.264 allows a slice to be partitioned into three parts:
 - **Partition A:** Partition A is the most important and it contains the header information, motion vectors, quantization parameters, MB types. Without the information contained in partition A, the data contained in partition B and C cannot be utilized.
 - **Partition B:** Partition B contains the intra coded block patterns (CBP) and the intra coefficients. Partition B is more important than partition C since intra information can stop error propagation and also can provide a synchronization point.
 - **Partition C:** Partition C contains the inter CBPs and the inter coefficients. It is the least important partition among the three since it neither contains any header information nor can stop error drift. Since most of the frames are coded as P frames, partition C forms the largest part of the coded video data. The

decoding of partition C requires the availability of partition A; but not of partition B.

Partition B and C are useless if they are received without partition A, because they cannot be decoded without the information contained in partition A. However, if partition B and/or partition C is lost, then the decoder can perform enhanced error concealment using the information available in partition A. Table 2.1 describes the error concealment schemes in case of loss of different type of partitions [4]:

Table 2.1 Error concealment scheme where different types of partitions are available [4]

Available Partitions	Concealment Method
A and B	Conceal using Motion Vectors (MV) from partition A and texture from partition B; intra concealment is optional.
A and C	Conceal using MVs from partition A and inter information from partition C; inter texture concealment is optional.
A	Conceal using MVs from partition A
B and/or C	Drop partitions B and C. Perform frame copy or motion copy according to the requirement.

2.2 3G Networks: A Brief Overview

The 3G cellular communications standard is designed to enable multimedia communications and consists of Universal Mobile Telecommunications Service (UMTS), Code Division Multiple Access (CDMA) 2000 and Time Division Synchronous CDMA (TD-SCDMA) variants. Currently, UMTS has emerged as the most widely adopted air interface. For this reason, we focus on this interface, which is also known as Wideband CDMA (WCDMA).

Without dwelling too much into the details, below we overview the layers of the UMTS air interface and their functionalities.

Layer 1, known as the physical layer, is responsible for the characteristics and the way data is transferred. This layer provides a carrier service for transport channels without caring about the content transferred. Its main jobs include modulation/demodulation, scrambling, interleaving, estimating Transport Channel Block Error Rate (TrCh BLER) and the ratio of the transmitted or received user data bit (E_b) to noise spectral density (N_0) i.e., E_b/N_0 , power control, multiplexing, and FEC.

Layer 2, known as Data Link Layer, is divided into RLC/MAC parts. This layer is responsible for the transmission format. It provides flow control function; including segmentation and reassembly of variable-length upper layer Packet Data Units (PDU) into/from smaller RLC PDUs, match RLC PDUs to transport channels e.g., Dedicated Channel (DCH), Downlink Shared Channel (DSCH), and Forward Access Channel (FACH). The MAC layer also performs some multiplexing and RLC can provide error correction functionalities by Acknowledgement (ACK) and retransmission (ACK Mode).

Layer 3 is called Radio Resource Control (RRC) which deals with resource management issues.

2.3 Stream Control Transmission Protocol (SCTP)

Multimedia application is a delay sensitive but loss tolerant application. Once the user start viewing the media, the subsequent packets should reach the destination within a specified time interval. This requirement sets an important condition on the transport layer protocols to be used for multimedia content delivery. The two most widely used transport layer protocols are TCP and UDP. TCP is very reliable for data transmission but is not suitable for multimedia content delivery. The shortcomings of TCP are described briefly below:

- TCP is a reliable connection oriented protocol ensuring ordered data delivery. This stringent nature causes problem for a delay sensitive application or in the case where more than one application are sharing the same TCP session. For instance, if one packet is lost for a particular application, then although the following packets are received correctly in the receiver end, they are not delivered to the application unless the lost packet is retransmitted. As a result, the applications that the correctly received packets belong to, suffer unwanted delay because of the loss of a packet that belongs to a separate application. This situation is known as Head of Line (HoL) blocking. Figure 2.3 depicts this problem. The consequence of this strict order of delivery is additional delay. For multimedia applications, each packet has a delivery deadline for it to be useful at the receiver and if that is passed, that packet is discarded. So HoL blocking problem can be destructive for multimedia applications since subsequent packets will experience additional delays and if the retransmission

interval is large, then a large number of packets can be discarded rendering a very low picture quality and also wastage of bandwidth.

- TCP works on byte oriented fashion. The application needs to provide some markers to delineate the messages.
- TCP has a three way handshake mechanism which makes it prone to Denial of Service (DoS) attack. In three way handshake scheme, the client sends a SYN message and the server responds to that with a SYN-ACK message. The client again responds with an ACK message and the connection is established. But the server cannot have any identity of the client. So if any rogue client sends a lot of SYN packets with invalid IP addresses, the server cannot detect this and will send SYN-ACK to each of these requests and will run out of bandwidth. Then any valid client will not get the service although bandwidth is available.

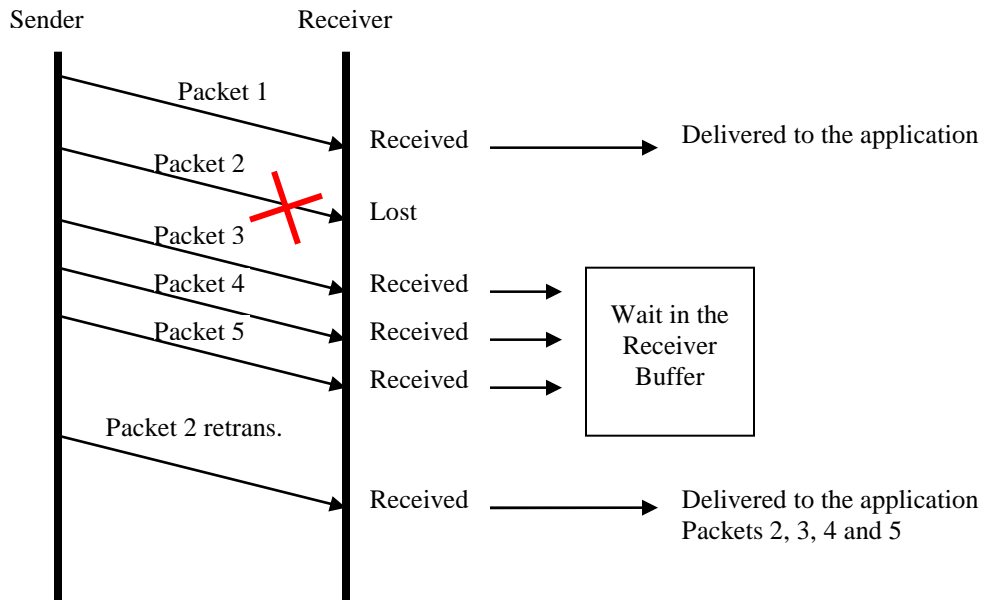


Figure 2.3 Head of Line (HoL) blocking problem in TCP

Apart from these shortcomings, TCP is one of the best transport layer protocols. It has efficient congestion and flow control mechanisms that adapt to the dynamic nature of the network.

The second widely used transport layer protocol is UDP. UDP is an unreliable protocol which does not provide any guarantee that data has been delivered to the other end. It has no congestion control or flow control mechanism. So in a congested network, UDP makes the situation worse by continuously injecting traffic. However, since UDP is unreliable, the delay incurred in this case is less than TCP and thus it's suitable for multimedia traffic. Also UDP works in message oriented fashion which is an advantage. The applications are required to incorporate some features on top of UDP to detect the loss of packets, etc., which adds to the complexity of the system.

From the above discussion, it is apparent that neither TCP nor UDP is best suited for multimedia traffic. This fact gives the motivation to seek for an alternative transport layer protocol which will alleviate the problems of TCP and UDP. SCTP [2], which is a relatively new transport layer protocol, devised in 2001, mainly for transporting telephone signaling messages, is a good candidate to be considered for multimedia application. SCTP has the good features of TCP and UDP and in addition to that, it has some novel features of its own. Following is a brief description of the SCTP features.

- **Multihoming:** The multihoming feature allows assigning more than one Internet Protocol (IP) addresses to an SCTP port. In multihomed SCTP, each end point is represented by a list of IP addresses that are assigned to the end point. The connection between two multihomed end points is referred as an 'association' in contrast to the term 'connection' used for TCP end points. All the IP addresses

can be assigned to the same port or different port depending on the application requirement. Figure 2.4 shows SCTP association between two multihomed hosts. Here the association between end point A and B can be described as $(\{A_1, A_2, A_3\}, \{B_1, B_2\})$.

Multihoming property provides redundancy so that if a path fails then data packets can be routed through the alternate paths. Usually one of the paths works as the primary path and data is transmitted using that path. When the primary path fails, then the secondary one is used for data transmissions until the primary path is restored. The present SCTP standard does not allow load sharing between two paths, but research is ongoing to address the problems that will arise as a consequence of transmitting data simultaneously (mostly reordering problem) using all the available paths.

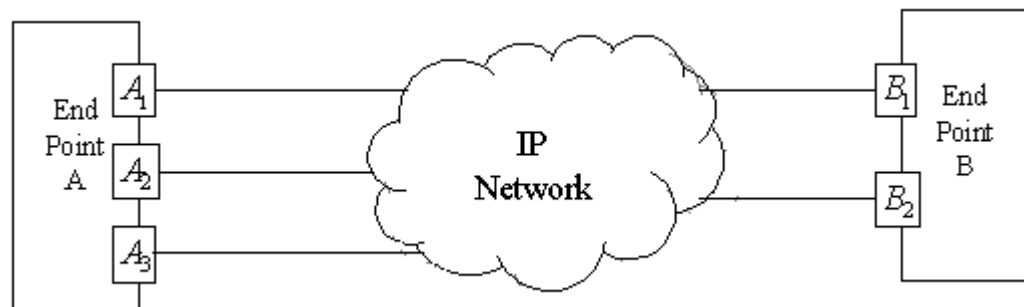


Figure 2.4 Multihoming property of SCTP

- **Multistreaming:** Multistreaming is another novel feature of SCTP. By virtue of this property, SCTP avoids the HoL blocking problem. Multistreaming property allows SCTP to multiplex the data from several applications onto one association.

Each stream can be defined as a unidirectional channel of data transmission (like a TCP connection) within an association. The streams are independent from each other and so data loss in one stream does not affect other streams. For example, in Fig. 2.5, if a data packet is lost in stream number 1, then only the packets following that lost packet in stream 1 needs to wait in the receiver buffer, but the packets arriving in stream 0 and 2 can be delivered to the upper layer application. In this way, HoL blocking problem is avoided.

It can be mentioned here that the two properties, multihoming and multistreaming are orthogonal to each other. That is multiple streams and multiple interfaces are logically independent. Data from any stream can use any path to reach any destination address.

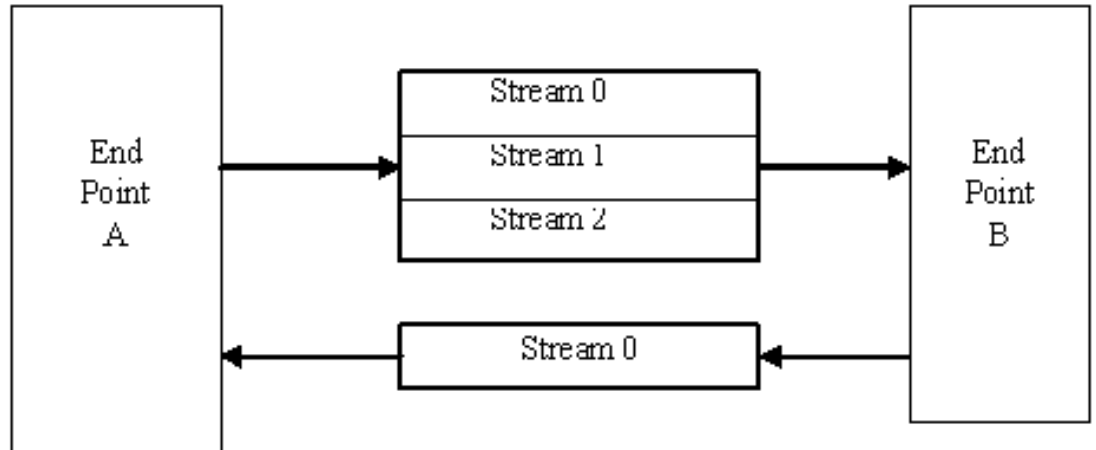


Figure 2.5 Multistreaming property of SCTP

- **Packet Structure:** The unit of data delivered to the other end is the SCTP packet or Protocol Data Unit. Each SCTP packet consists of a header and a chunk, either control or data. Multiple chunks can be aggregated up to the Maximum

Transmission Unit (MTU) or can be fragmented if necessary. Each data chunk is assigned a 32-bit Transmission Sequence Number (TSN) in order to ensure reliable delivery of data. The TSN assigned to data chunk is independent of the Stream Sequence Number (SSN) assigned in the stream level.

- **Congestion Control:** The congestion control scheme of SCTP is similar to TCP. It employs a window based congestion control mechanism. SCTP maintains three variables for this purpose, the sender's congestion window (cwnd) which is maintained on a per destination basis, the receiver's advertised window (rwnd) which is shared across an association and the slow start threshold (ssthreshold). It also keeps track of the Round Trip Time (RTT) estimates and maintains a retransmission timer per destination. SCTP uses the Selective Acknowledgement (SACK) mechanism to have a better performance. Use of SACK is mandatory in SCTP to ensure a more robust response to multiple losses from a single windowed data.
- **Partial Reliability extension of SCTP (PR-SCTP):** Partial reliability extension property enables SCTP to have different reliability on different messages. The reliability can be imposed in terms of time and also number of retransmissions. In the former case, each message is assigned a transmission deadline. The message will be retransmitted if the deadline has not expired; otherwise it will be discarded. In the retransmission based reliability, it is possible to fix the number of retransmissions each message is allowed to have. If the number of allowable retransmissions is zero that means the message will be discarded if it is lost.

This property is very important for the application that has messages of varying importance level. Partial reliability provides differential service for such applications without adding any extra complexity.

2.4 Literature Review

2.4.1 Previous Works on Video Transmission over Wireless

Environments

The fundamental work for analyzing H.264 in wireless environments was laid down by Stockhammer, Hannuksela and Wiegand in [5]. The authors provide an overview of coding and error-resilience tools, which are likely to be used in wireless environments. Experimental results are given for selected system concepts based on the common test conditions. It is found that, for different error concealment methods, the introduction of shorter packets significantly increases the decoded quality. In [6], Liu, S. Zhang, Ye and Y. Zhang did further research on error-resilience tools and analyzed their usability in a 3G environment. This work showed that encoding with simple FMO mode and extra intra block refreshing achieves the best trade-off between error correction performance and bit rate.

Other works have focused on enhancing the H.264 video transmission over bursty packet-loss 3G cellular networks [7]-[8]. In [7], an audio/video frame interleaving scheme is presented, which is based on priority based scheduling using feedback. Experimental results showed that interleaving achieves superior performance in the presence of link outages. In [8], the authors investigated a combined use of passive error concealment together with FEC coding and periodic intra-updating to improve the performance in the presence of bursty packet losses. However, a cross-layer optimization may provide better overall performance.

A robust cross-layer architecture that relies on a data partitioning technique at the application layer and an appropriate priority mapping at the 802.11e MAC layer is described in [9]. This solution is complemented by the schemes proposed in [10]. In [10], a cross-layer design is presented that optimizes the encoded packet sizes to improve H.264 video transmission over 802.11 WLAN. Instead of doing the fragmentation in the MAC layer, packet sizes are adjusted by slicing the video frame at the application layer, achieving better performance for a given packet-loss condition.

2.4.2 Previous Works on Video Transmission using SCTP as the Transport Layer Protocol

Video transmission using SCTP as a transport layer protocol is an active research area and a lot of experiments have been conducted to verify whether SCTP can perform better than UDP in terms of quality and delay. Most of these studies are concerned with the reliability level, testing MPEG-4 video transmission using SCTP in different platforms [11]-[13]. In these cases, the reliability level of I frames is set to be higher than the P and B frames so that only I frames are retransmitted in a lossy environment. H.264 video transmission using SCTP is investigated in [14], where similar experiments as those conducted in [11], [12], and [13] showed that SCTP performs better than UDP. A novel idea of frame dropping filter based on the Partial Reliability extension of SCTP (PR-SCTP) is also introduced. In PR-SCTP, each message can have different reliability value and is retransmitted according to that value. The sender gives up retransmission if the reliability value is set to 0.

Another novel scheme where reliability level can be set depending on the network condition is proposed in [15]. The authors proposed that the reliability should become a function of the congestion state, instead of always sending I frames in a reliable stream. In case of low and

moderate congestion, an I frame is sent with reliability 1 whereas in the congested case, reliability is set to zero to avoid the high probability of ineffective retransmission.

While the above schemes provide enhancements for transmission of video using SCTP, to the best of our knowledge, no solution has been proposed for mapping the error resiliency features of H.264 to the reliability features of SCTP. Also, the existing solutions for transmission of error resilient H.264 video are mostly focused on MAC and physical layer issues in wireless networks. The works presented in [16]-[17] and [9] investigate the efficient transmission of H.264 in IEEE 802.11 wireless networks, focusing on the MAC and physical layer issues.

2.4.3 Previous Works on Video Transmission Employing Unequal Error Protection (UEP) by means of FEC and Retransmission

The high loss ratio of wireless environment calls for employing efficient error correction schemes to maintain a minimum base quality of video to the end user. FEC [18] and the packet retransmission are the two widely used schemes to deal with packet errors. In application layer FEC, the sender uses redundancy against the packet loss ratio to recover the lost information whereas the lost packet is retransmitted in the packet retransmission scheme. But in case of large scale multicast or broadcast video transmission, retransmission of lost packets to individual users is not possible and FEC is the only available error correction method.

Considering the delay constraint and the bandwidth usage, several variants of FEC and Automatic Repeat reQuest (ARQ) have been proposed. Due to the time varying nature of the wireless network, channel adaptive FEC schemes are more effective than the static FEC schemes. An Enhanced Adaptive FEC (EAFEC) scheme is proposed in [19]. The EAFEC

scheme determines the degree of redundancy depending on traffic load (estimated by the queue length) and channel state (estimated by the packet retransmission time). This scheme performs better than the static FEC scheme but for optimum performance, the threshold values for queue length and retransmission time should be judiciously determined. A unified scheme for optimal video streaming combining scheduling, error correction and error concealment for layered video coding where network bandwidth is not known a priori is presented in [20]. The authors showed that their scheme performs better whereas static error protection schemes results in near-optimal performance. For layered video, several UEP scheme based on FEC have been proposed [21]-[23], where redundancy is considered as a function of the importance of the layer. The most important base layer is provided higher redundancy to ensure a minimum quality level. The performance of hybrid ARQ scheme (employing both FEC and ARQ) has also been studied in [24]-[25].

The idea of UEP can be used more efficiently if the video slice/frame can be divided into layers of different importance. This can be achieved by virtue of data partitioning property. Several UEP schemes have been studied for data-partitioned video [26]-[28]. An UEP scheme considering the varying length of the partitions is presented in [26]. The work in [27] depicts a novel scheme where the data is prioritized in a tree structure depending on their importance and then protected unequally. A variant of this work is presented in [28] where the protection is allocated to the key pictures. A combination of FEC and hierarchical quadrature amplitude modulation to protect partitions differentially is presented in [29]. The work in [30] presents joint utilization of the source and channel coding scheme. In this paper, the authors subdivided partition C depending on the impact factor (quality degradation as a consequence of losing a particular partition) and protected each subdivision differentially by

the FEC scheme. Although the authors showed improved results compared to the conventional H.264, the extent of utilization will depend on the complexity of implementation and also on the improvement of relative quality by subdividing partition C.

3 Video Packetization Techniques for Enhancing H.264

Video Transmission over 3G Networks

3.1 Overview

In this chapter we study the performance of a cross-layer optimization scheme for video packet size adjustment in 3G cellular networks. Our motivation is based on the performance results achieved by the cross-layer approach proposed in [10], but also the significant differences between the MAC layers of 802.11 WLAN and a 3G system. The former uses Carrier Sense Multiple Access/ Collision Avoidance (CSMA/CA) whereas the 3G cellular systems employ dedicated channels to transport the packets. In this work, we focus on video transmission over UMTS. UMTS frames have a fixed length which is determined when the bearer is set up, unlike the WLAN case where the MAC can flexibly adjust its frame size. These two significant differences between the two types of wireless networks may lead to different solutions for enhancement of video transmissions over UMTS cellular networks and is the motivation for our work.

The rest of the chapter is organized as follows. Section 3.2 presents the steps for video packetization targeting 3G networks. Our proposed scheme is described in Section 3.3, which also includes analysis and discussion of the simulation results. Conclusions are given in Section 3.4.

3.2 Packetization Steps in 3G Networks

As discussed in Chapter 2, the H.264/AVC VCL data are passed to the outside world in the form of NALUs. Each of the NALUs is then encapsulated in a RTP packet. For real time data

transfer we use UDP in the transport layer. The underlying network layer protocol is IP. Both UDP and IP are unreliable protocols. UDP does not guarantee end to end data delivery and IP only performs best effort packet routing. For this reason, there is a need to use RTP, which runs over UDP. With the help of some features such as sequence numbering and time stamps, RTP makes the transmission of video data feasible over unreliable transport and network layer protocols. An RTP packet is encapsulated in a UDP and IP packet. Each packet has its associated header which will overload the network. The header lengths for RTP, UDP and IP are 12 bytes, 8 bytes and 20 bytes (for IPv4, 40 bytes for IPv6) respectively. Therefore, immediately above the link layer, the Packet Data Convergence Protocol (PDCP) performs Robust Header Compression (RoHC). After RoHC, the IP/UDP/RTP packet is encapsulated into a PDCP or Point-to-Point Protocol (PPP) packet, thus forming the Service Data Unit (SDU) of the RLC protocol. This RLC-SDU needs to be fitted in the link layer frame, RLC PDU, which is of a fixed length that is determined depending on some physical layer parameters such as the spreading factor, etc. If the RLC SDU is larger than the RLC PDU, then it is fragmented to fit in the PDU. If the size is smaller than an RLC PDU, then there are two options: (a) the remaining space is filled by padding bits at the cost of increased overhead, and (b) the remaining part is filled by the start of the next RLC PDU. That is, each PDU contains the same amount of information bits. In our performance evaluations we use case (b). Figure 3.1 depicts the packetization steps in 3G networks.

3.3 Packet Size Optimization in 3G Networks

The objective of our work is to determine the packet size of H.264 video that offers the best performance over a 3G network. Our motivation is derived from the discussion of the impact

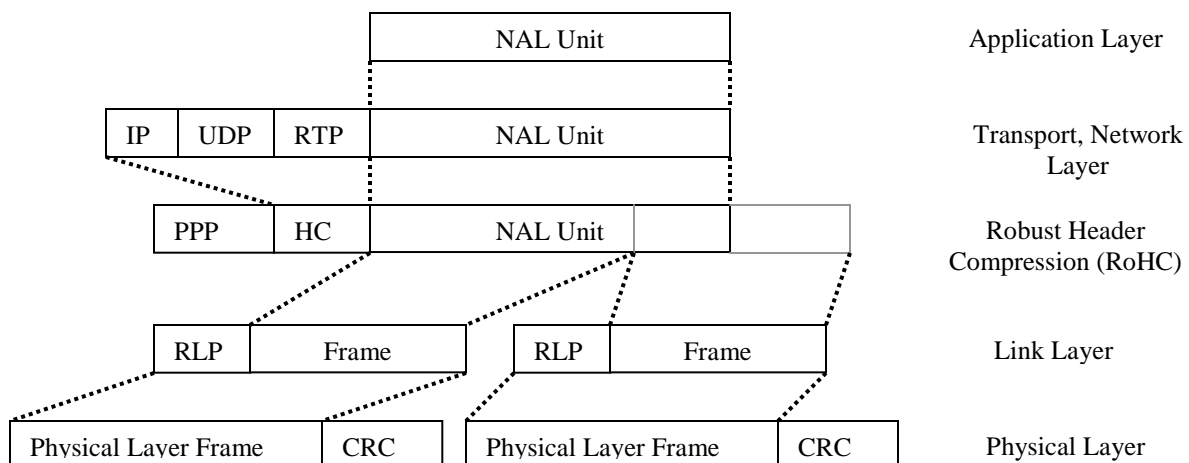


Figure 3.1 Packetization in 3GPP protocol stack

of packet length on Bit Error Rate (BER) in [5] and a similar experiment conducted in [10] for WLAN.

In the case of a 3G network, however, conditions are different, starting with the size of the RLC PDU, which is fixed. Thus, the MAC layer actually performs two tasks: 1) aggregation (if the RLC SDU is smaller), and 2) fragmentation (if the RLC SDU is larger than the RLC PDU). In order to characterize the quality of the channel, 3G uses BLER instead of BER. If we want to find an “optimum” packet size for video transmission over 3G, the effect of the aggregation and fragmentation mechanisms on the application layer Packet Loss Ratio (PLR), denoted by Video PLR (VPLR), should be investigated first. In the case of fragmentation, loss of one fragment will cause the total SDU to be discarded. If the application layer video frame size is M and L is the PDU size in bits (L is fixed), then there are $n = \lceil M / L \rceil$ packets (the last packet will be zero padded if the amount of information bits is not enough to fill the PDU length L). Since n denotes the number of packets in case of

fragmentation, n is always greater than or equal to one. Thus, the VPLR in the presence of fragmentation is:

$$\begin{aligned}
 VPLR_{frag.} &= 1 - \Pr.(no\ fragment\ is\ lost) \\
 &= 1 - (1 - \Pr.(L\ sized\ fragments\ are\ lost))^n \\
 &= 1 - (1 - e_b)^n
 \end{aligned} \tag{3.1}$$

where e_b is the BLER.

However, in case of aggregation, one RLC PDU is filled with more than one application layer PDUs. If L is the length of the RLC PDU in bits and M is the application layer PDU size in bits, then on average each RLC PDU consists of $n' = L/M$ application layer PDUs. The probability of loss of one PDU is characterized by the BLER, e_b , which corresponds to the loss of n' application layer PDUs. Given that there are n' times more application layer PDUs than the RLC PDUs, the VPLR in case of aggregation is:

$$VPLR_{aggr.} = e_b \tag{3.2}$$

Now a direct comparison between (3.1) and (3.2) shows that

$$VPLR_{frag.} > VPLR_{aggr.} \tag{3.3}$$

According to (3.3), it can be said that, in general, smaller packets should generate better picture quality. However, the limiting factor here is the bit rate. As the packet size decreases, the video compression efficiency decreases as well, resulting in a higher bit rate. Furthermore, smaller packet sizes cause an increased amount of overhead. The total bit rate (that is the video bit rate and the bit rate due to the overhead) should be less than the available channel capacity in order to ensure proper video transmission. Thus, an arbitrarily small sized packet will not necessarily generate the best quality picture.

To reconfirm the above arguments, we have simulated the transport of H.264 video over 3G networks. Figure 3.2 depicts the framework used for our tests. We use H.264/AVC reference software JM 10.2 [31] to encode and decode the video sequences, while the 3GPP simulation software [32] is used to simulate the characteristics of 3G wireless networks. The 3GPP simulator is an offline simulator based on the common test conditions suitable for transmission in 3GPP/3GPP2 networks as described in [33]. This simulator assumes that packets are delivered in order and it drops a packet depending on the error masks (specified in the simulator). No bit erroneous or corrupted packet is delivered to the upper layer.

We simulated three Quarter Common Intermediate Format (QCIF) video streams (coastguard, foreman and hall monitor [34]) at different BLERs (1.5%, 2% and 5%) for a channel capacity of 128 kbps. We selected these three videos in accordance with [33]. At the encoding side, we used the ‘Dispersed’ FMO mode, since [10] showed that this mode has the best performance. Frame rate and quantization parameters were being changed to adjust the video bit rate with the channel capacity. All other parameters were set to the default values. For the 3GPP simulator, we simulated the Packet Switched Conversational (PSC) scenario

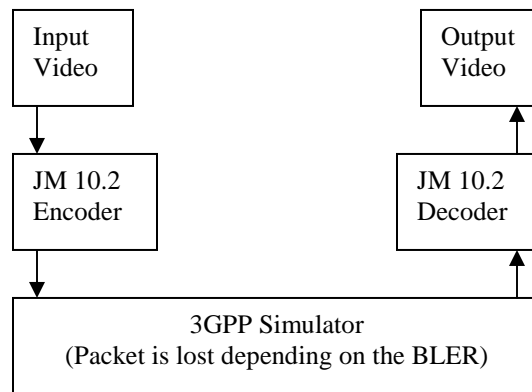
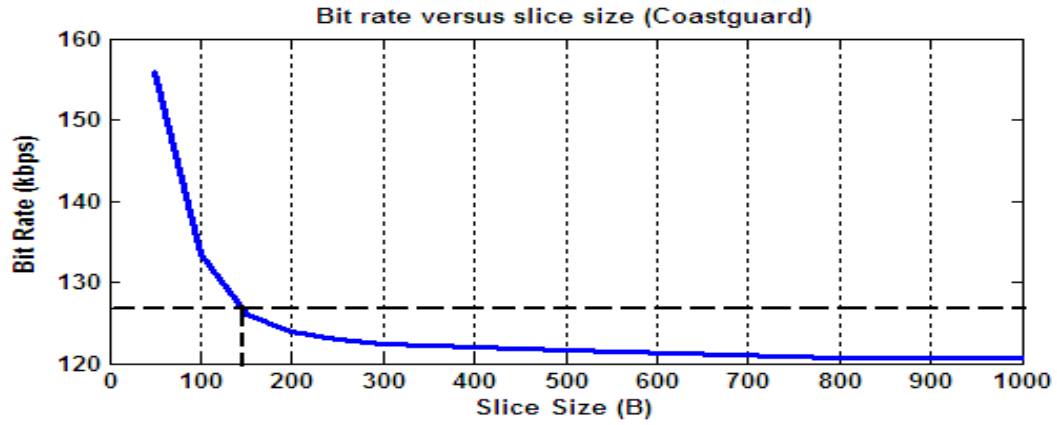


Figure 3.2 Framework of the experiment

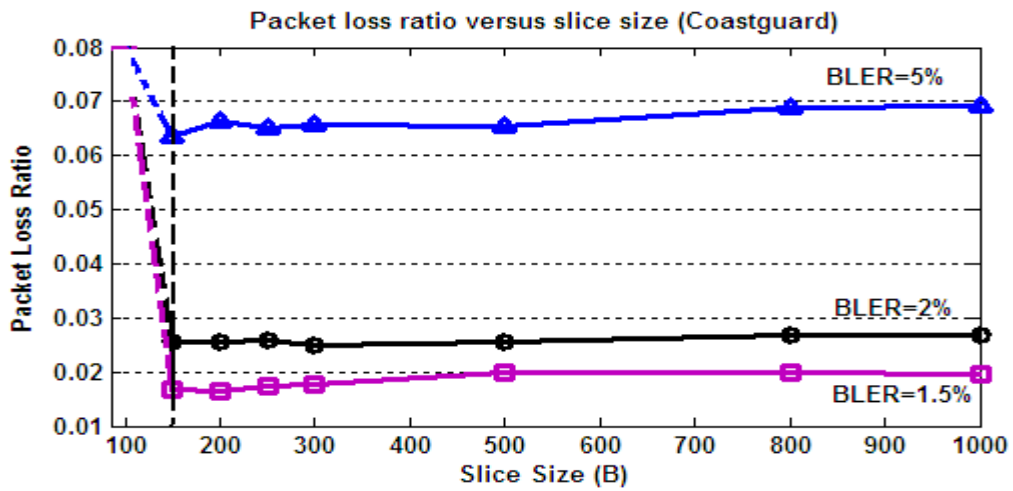
for a mobile speed of 3 kilometers per hour. PSC services have very strict delay constraint and they are actually meant for real time communications. So the simulator was modified to impose a maximum bound on delay, which is 250 ms [5]. Due to the characteristics of the PSC services, we have not used retransmission mode in simulating conversational services. Figures 3.3, 3.4, and 3.5 show the simulation results for the three videos, coastguard, foreman and hall monitor respectively.

Figures 3.3a, 3.4a and 3.5a show the graphs of bit rate versus slice size. As expected, the bit rate decreases as the slice size increases. To transmit video over the 3G network, the video bit rate and the bit rate due to the overhead associated with each packet should match the channel capacity. The dashed line in Fig. 3.3a, 3.4a and 3.5a indicates the threshold (128 kbps) above which the video is not deliverable or is delivered with significant delay (a major limitation in case of conversational services).

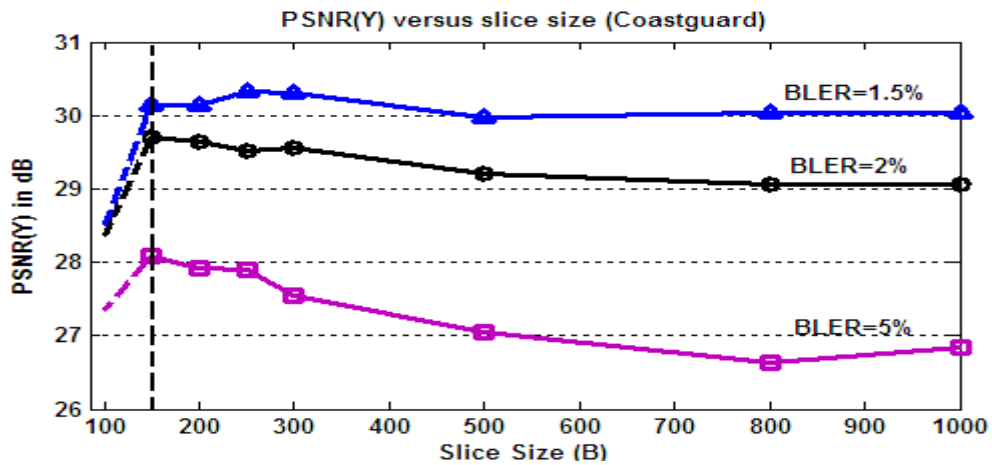
Figures 3.3b, 3.4b and 3.5b depict the relation of loss ratio with packet size. We can split the curve into two regions: prior to the threshold (video bit rate \geq channel capacity) and after the threshold (video bit rate $<$ channel capacity). For successful video transmission, the channel capacity and video bit rate must be matched. If the video bit rate is more than the channel capacity, it will incur indefinitely increasing and unbounded delay, resulting in the packet loss ratio of 1 for applications with bounded delay requirements. After reaching the threshold point, VPLR should increase with the packet size according to (3.3), something that is evident from the simulation results too. A slight deviation from this behavior is observed in the PLR versus slice size curve for the video sequence, 'coastguard' at BLER=1.5%. Though the experiment was repeated 10 times with different starting positions in the error mask file to have statistical reliability, this deviation might be considered a simulation related fact and



(a)

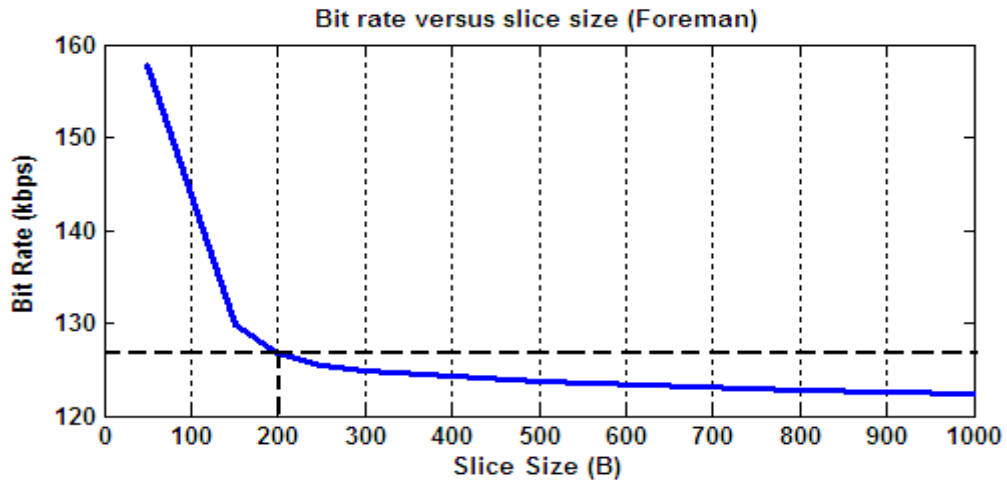


(b)

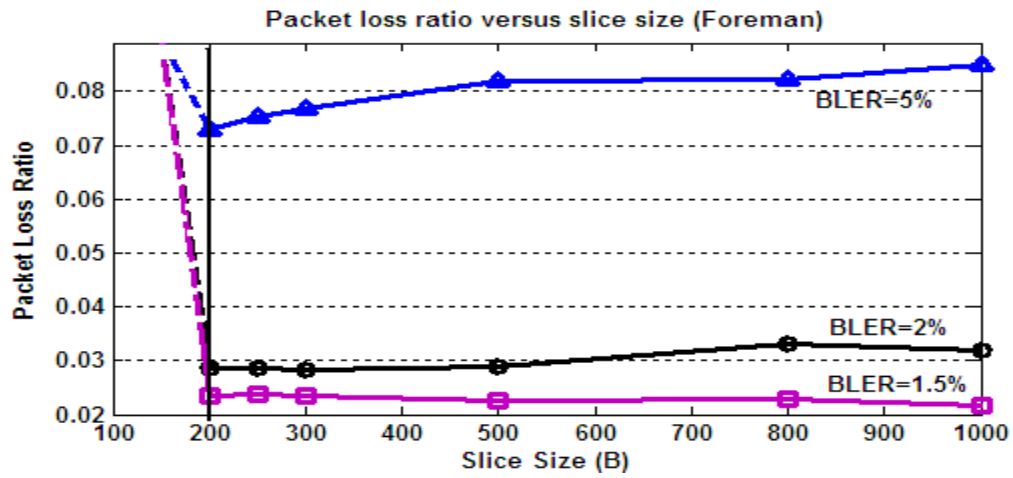


(c)

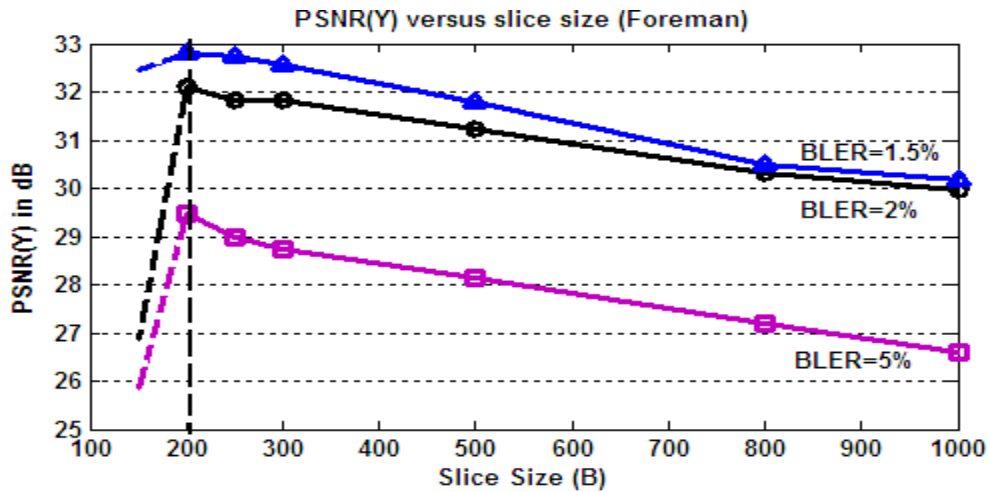
Figure 3.3 For the video sequence Coastguard: (a) Bit rate versus slice size, (b) Loss ratio versus slice size and (c) PSNR(Y) versus slice size.



(a)

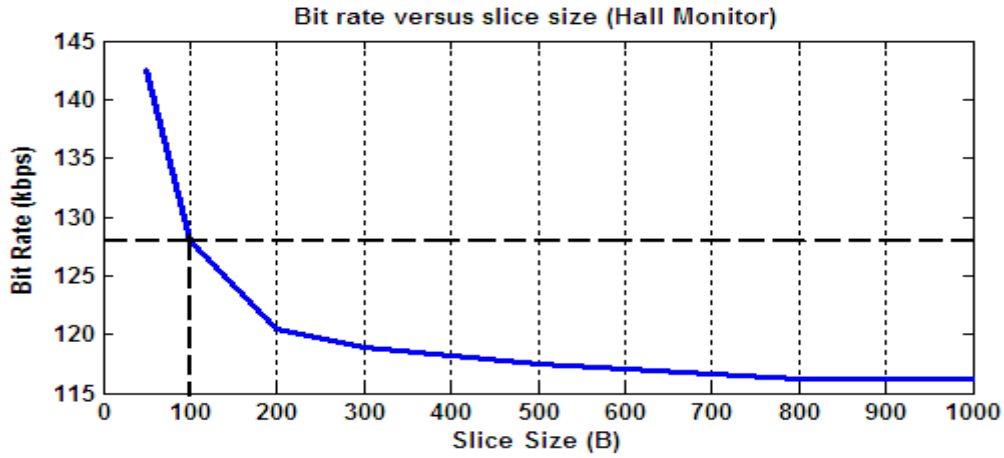


(b)

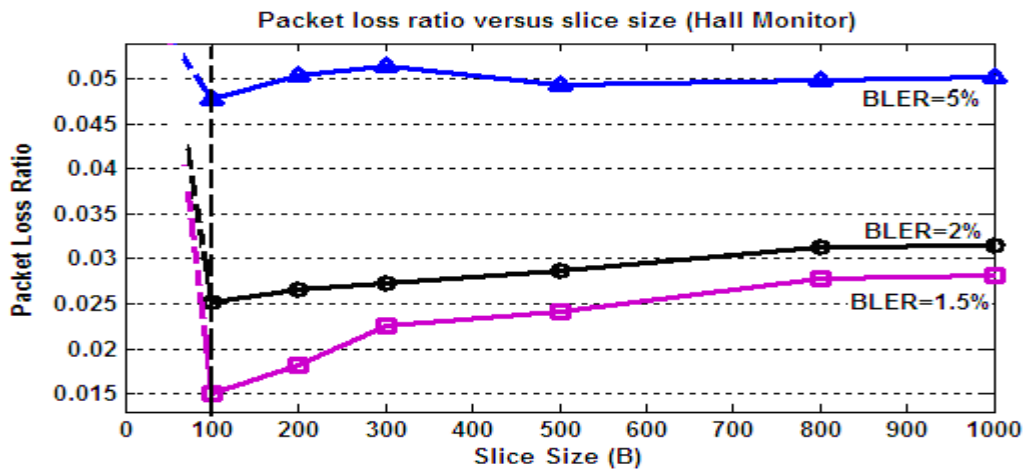


(c)

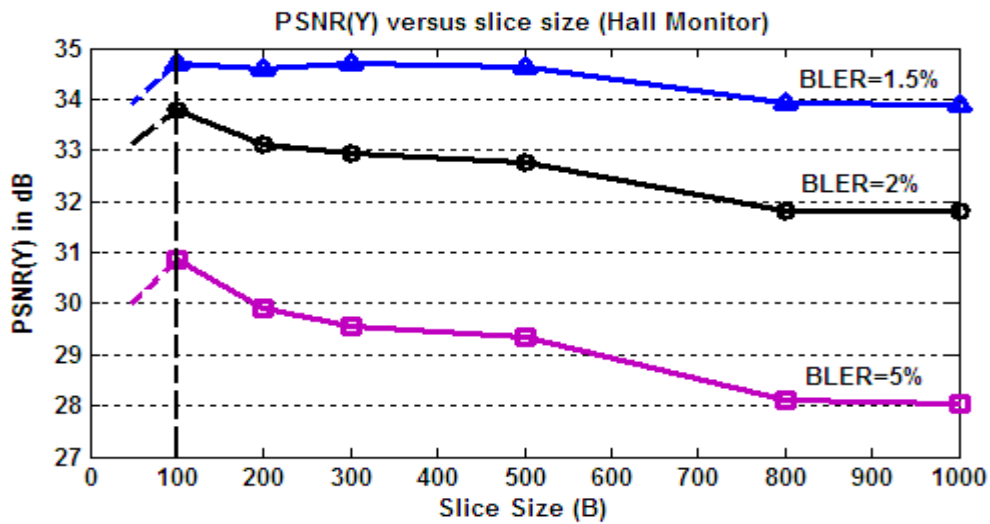
Figure 3.4 For the video sequence Foreman: (a) Bit rate versus slice size, (b) Loss ratio versus slice size and (c) PSNR(Y) versus slice size.



(a)



(b)



(c)

Figure 3.5 For the video sequence Hall Monitor: (a) Bit rate versus slice size, (b) Loss ratio versus slice size and (c) PSNR(Y) versus slice size.

would have shown the expected behavior if the experiment was repeated a large number of times (e.g. 100 times). Because of the hybrid nature of the codec, the effect of a lost slice/frame is not limited to that particular region the slice/frame belongs to rather it propagates throughout the Group of Picture (GOP). Also loss of slices/frames from different positions of a GOP differentially affects the quality. However, if the simulation is being run many times, then effectively the loss is distributed throughout the GOP and this effect is minimized. Again, in this case the PLR decrease is so small that it can be safely ignored in terms of having any effect on the perceptual picture quality. Figures 3.3c, 3.4c, and 3.5c show the graphs of luma Peak Signal to Noise Ratio (PSNR) versus slice sizes. As expected, we observe that the PSNR is the highest at the threshold point and afterwards it has a decreasing characteristic. This is in accordance with the PLR versus slice size graph. Since the PLR increases after the threshold value, PSNR should decrease in that region. Actually this is due to the overhead or reduced compression efficiency in the case of smaller slice sizes, with the picture quality not being the best at the smaller slice size but rather depending on the bit rate.

3.4 Summary

In this chapter, we have studied the impact of changing slice sizes on the overall performance of video streaming applications in 3G networks. We observed that in general, smaller packet sizes result in lower packet loss ratios and better video quality, as long as the bit rate matches the channel capacity. However, using smaller slice sizes reduces the efficiency of encoding and introduces more overhead, resulting in increased bit rate for the video. As a result, slice sizes cannot be chosen arbitrarily small and a lower limit on the slice size is determined by the achieved bit rate for the video sequence.

While we assumed RTP/UDP as the transport protocol and studied data link layer issues for the delivery of H.264 video over 3G networks, further improvements may be achieved if reliable transport protocols are used instead of RTP/UDP. For example, SCTP may be used at the transport layer. This protocol has all the potentials to be proved as an effective transport layer protocol to transmit video over wireless network. In the next chapter, we study the transmission of data partitioned H.264 video using the partial reliability extension of SCTP.

4 Transport of Data-Partitioned H.264 Video Using SCTP

4.1 Overview

In this chapter we consider the effects of transport layer reliability (using PR-SCTP) in the transmissions of H.264 video over IP networks. As we discussed in Chapter 2, DP is one of the very important error resilience features of H.264. With DP, each video slice is encoded into three different classes of data with different importance. The encoded partitions containing the most important information should be protected against transmission errors to ensure a good picture quality. By virtue of the multistreaming feature of SCTP and the partial reliability feature of the PR-SCTP extension, we can set different priority or reliability levels for different DPs. In this chapter, we investigate the impacts of the loss of DPs on picture quality. We present a comparative study of the possible solutions for transmission of H.264 video using SCTP, considering both partitioned and non-partitioned H.264 video. We demonstrate how the reliability features of SCTP can be efficiently mapped to the error resiliency features of H.264 video.

The rest of the chapter is organized as follows. Section 4.2 illustrates our approach to transport H.264 video DPs using SCTP, performance evaluations and discussions are demonstrated in Section 4.3 and Conclusions are presented in Section 4.4.

4.2 Utilizing SCTP Reliability Features to Transport H.264 Video

Data Partitions

The objective of this work is to jointly optimize the data partitioning property of H.264 and reliability features of SCTP incorporating the PR-SCTP extension. One of the reliability

features of SCTP is multistreaming, which enables SCTP to send data in multiple streams. This property has been described in Chapter 2. For video applications, which are delay sensitive and loss tolerant, multistreaming is very useful. Using this property, a single video sequence can be sent through several streams, in effect allowing the use of differentiated services provided by SCTP and lower layers. Moreover, the partial reliability feature of the PR-SCTP extension enables us to set different priority or reliability levels for different packets. For example, I and P frames can be transmitted in separate streams. Since I frames are more important, the stream carrying I frames can use reliability level 1. On the other hand, reliability of P frames can be set to 0 so that they will not be retransmitted once they are lost. This is very important because loss of an I frame degrades the quality of the picture, and protecting every packet is expensive in terms of resource utilization. Therefore, PR-SCTP gives us a trade-off between quality and resource utilization; by setting I frames as priority 1, an acceptable picture quality may be achieved while reducing use of resources.

DP is a very important error resilient feature of H.264, as described in Chapter 2. The DP feature of H.264 allows NALUs to have different levels of importance. Utilizing PR-SCTP, different data partitions can be transported in different streams and have different reliability levels. Since partition A is the most important, it should have reliability 1 and since partition C contains less important information, it can be sent through a less reliable stream. Partition B contains intra information. Before deciding whether partition B should be sent reliably or unreliably, the impact of loss of partition B on picture quality should be investigated first.

Ideally, in streaming applications where moderate delay is acceptable, it is preferred that all packets are transported reliably. However, the limited available capacity or transmission resource of the underlying network prevents us from enabling retransmission features for all

packets (or partitions). To have a valid comparison between all the solutions that allow partial retransmissions of packets, we assume that the same amount of resources is used in all scenarios. Since DP divides a slice/frame into three partitions, total number of packets increases compared to a non-DP video. Also, the packet lengths of different partitions and whole frame/slice are different. So, while enabling retransmissions of packets, the average number of bytes transmitted per GOP should remain the same in all the cases. We derive this value as follows:

$$G = N_{rel}(1 + p)L_{rel} + N_{unrel}L_{unrel} \quad (4.1)$$

where G is the number of bytes transmitted per GOP, p is the packet loss ratio, N_{rel} and N_{unrel} are the number of packets transmitted in a reliable and unreliable stream, respectively, and L_{rel} and L_{unrel} are the average lengths of packets that are sent respectively in the reliable and unreliable streams.

In general, determining the number of packets that can be retransmitted is done based on the assigned channel capacity and the actual video bit rate. If the channel capacity dedicated to the video stream is C bits per second (bps), the video bit rate is C_v bps, and $C > C_v$, then the capacity available for retransmitted packets is $C - C_v$. We define this as the extra capacity available for retransmissions and denote it as E . The number of packets that can be retransmitted will depend on E . If there is no retransmission, the number of bytes transmitted per GOP is given by,

$$\hat{G} = N_{rel}L_{rel} + N_{unrel}L_{unrel} \quad (4.2)$$

Therefore, E is found as:

$$E = G - \hat{G}$$

$$= pL_{rel}N_{rel} \quad (4.3)$$

When setting the reliability value we have to ensure that the number of packets retransmitted satisfies (4.3).

The combinations of using DP or not, and the possibilities to set different reliability levels to different DPs result in the following five scenarios:

Case A: The video is coded with no DP and only I slices have priority 1,

Case B: The video is coded with no DP and all I slices and some of the P slices have priority 1,

Case C: The video is coded with DP and partitions A and B have priority 1,

Case D: The video is coded with DP and partitions A and B that belong to an I frame have priority 1,

Case E: The video is coded with DP and only partition A has priority 1.

The impact of these five scenarios on video quality will obviously be different. For case A, we do not use any DP and only I frames are protected; we expect that it will give us the lowest picture quality among the five cases since some of the available extra capacity is not used for retransmissions. We included this scenario only as a reference. In case B, in addition to I frames, some of the P frames are also sent in a reliable stream to ensure that G remains the same for case B and case C, in which DP is used. Both cases B and C will result in a better picture quality than case A because of the utilization of the available extra capacity.

The comparison between case B and case C is crucial. In case C, partition C, which contains less important information (inter information) of a slice/ frame, is subject to packet losses, whereas in case B, whole P frame or P slice may be lost. We expect that the decoder will conceal error more efficiently where some information about the slice is available (as in case

C) than the case where the whole slice is lost (as in case B), and so case C should result in a better picture quality than case B.

The remaining two scenarios investigate whether partition B should be sent reliably or unreliably. Partition B contains intra information. Also, both I frames and P frames contain intra information. However, loss of intra information from an I frame may result in a lower picture quality than the loss of intra information from a P frame. To investigate this fact, we can divide partition B into two groups: partition B that belongs to a P frame and partition B that belongs to an I frame. In case D, we set different priorities for them. Partition B belonging to an I frame is sent in a reliable stream along with partition A. So partition C and B belonging to a P frames can be lost. Since two types of partitions may be lost compared to case C, the quality will be worse but since we are losing information belonging to a P frame, the quality will be better than case B and will be very close to case C.

In the last case, we consider only partition A as reliability 1 message. Partitions B and C are sent unreliably. Here partition B packets belonging to both P and I frames may be lost. It is not clear, in this case, whether using DP results in better quality than in the case where data partitioning is not used and the available protection is dedicated to I frames. We examine this case in the next section.

4.3 Performance Evaluations and Discussions

To confirm the arguments presented in Section 4.2, we conducted several simulation experiments. We observed the quality of H.264 video delivered using SCTP in different packet loss scenarios. We used the JM 13.1 codec [31] to encode and decode the video stream. In this study, we focus on how the SCTP features can be utilized for video transmission. Any simulator that simulates the end to end scenario of a network will serve

our purpose. Network Simulator-2 (NS-2) is a good choice and we used NS-2.29 [35] in our experimentation. However, the legacy applications in NS-2 do not use any properties like multistreaming and PR-SCTP. We use an SCTP aware application so that we can use the SCTP application programming interface (API) in NS-2. Figure 4.1 depicts our simulation scenario. This simple scenario consists of a sender, two routers and a receiver. The bandwidth and delay between the nodes and the routers are 0.5 Mbps and 10 ms and those between the routers are 1Mbps and 10ms, respectively. We consider this simple scenario since at this point we are only concerned with the picture quality by using the partial reliability extension of SCTP and not concerned with any other performance measure such as delay, bandwidth usage. Packet losses such as those that may be introduced in a wireless environment are incorporated. A wireless transmission error is usually detected by a failed checksum at the receiver and causes the receiver to drop the erroneous packet. In our simulation scenario, a composite packet loss ratio taking into account both transmission and queuing losses has been considered and packets are being dropped according to a random uniform loss model at the receiver end.

We used two representative QCIF video sequences, Foreman and Coastguard [34], for the evaluation of simulated video transmissions with different effective packet loss ratios. Use of DP increases the number of packets and protects almost half of the packets (even if only partition A is protected). Therefore, we set some of the P packets as priority 1 as long as (4.1) and (4.3) are satisfied. The videos are encoded using the H.264/AVC extended profile and the IPPPP GOP structure. We evaluated the five cases described in Section 4.2. In the first part of our experiments, we investigated the impact of the loss of partition B on picture

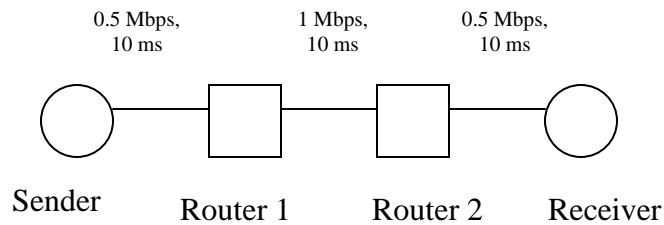
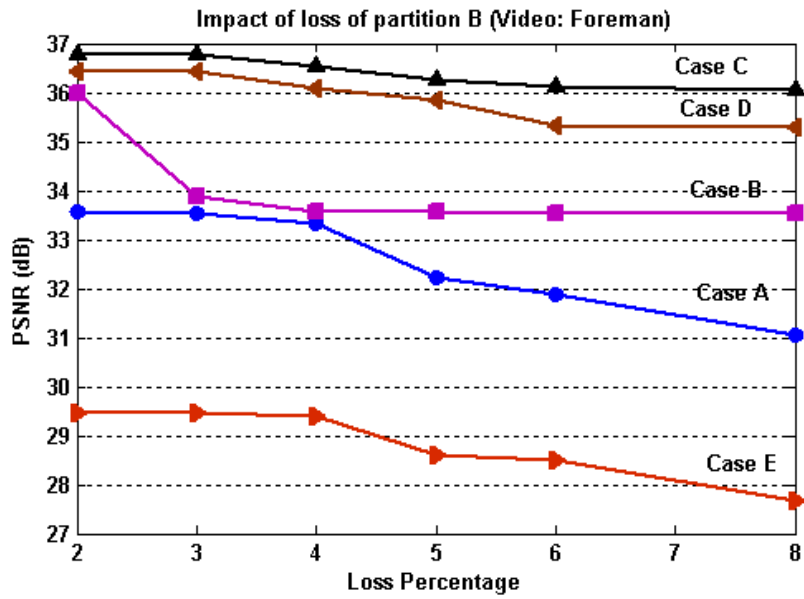


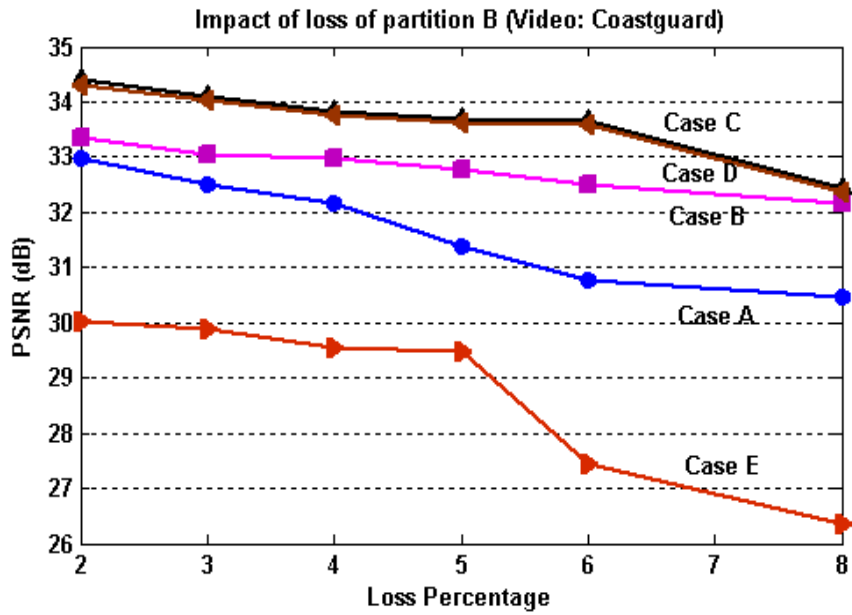
Figure 4.1 Simulation Topology

quality. Figure 4.2 shows the simulation results for the two videos we have used. We have used a slice size of 500 Bytes in this experimentation. We observed that case C resulted in the highest PSNR, since only partition C was sent in an unreliable stream. Case D has PSNR values very close to case C but the quality of case E is the worst, even worse than case A, in which no DP is used. This is in accordance with our discussion in section 4.2. In case A, we do not lose any information of an I frame/slice but in case E, we lose the intra information of an I frame/slice. Although partition B is only the intra information, not the whole slice; we can conclude that the loss of any information from an I slice significantly degrades the quality of the picture.

Since loss of partition B packets belonging to I frames results in the worst picture quality among all the cases, in the second part of the experiment we consider that these packets are transported in a reliable stream. To conclude concretely that case C always performs better than the cases where available protection is dedicated to I frames, we ran the simulation for two video sequences, Foreman and Coastguard, for different slice sizes and different FMO combinations: (i) slicing: 500 bytes, FMO: Dispersed; (ii) Slicing: 500 bytes and FMO: none and (iii) slicing: none and FMO: none. Figures 4.3 and 4.4 show the simulation results.

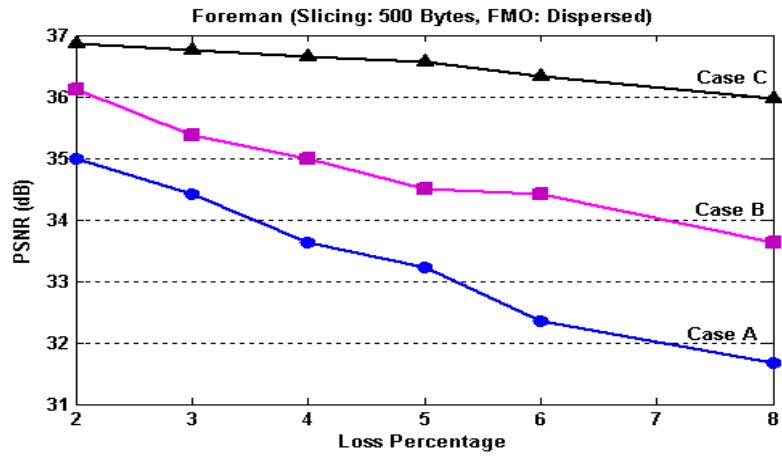


(a)

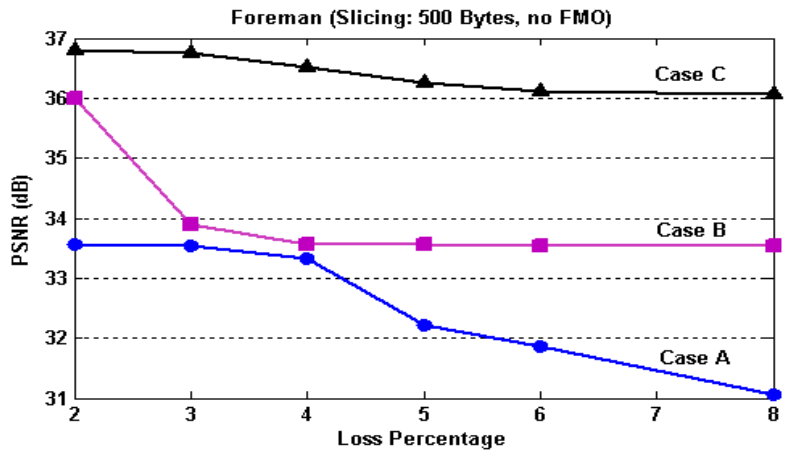


(b)

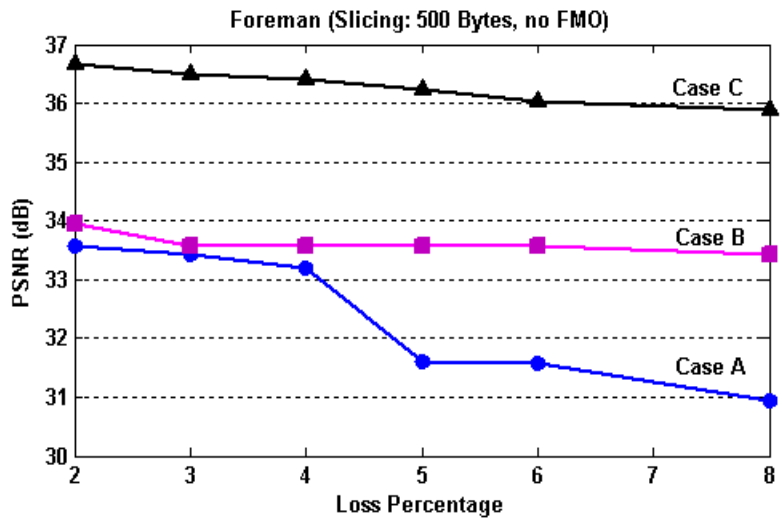
Figure 4.2 Impact of loss of partition B (a: Foreman, b: Coastguard)



(a)

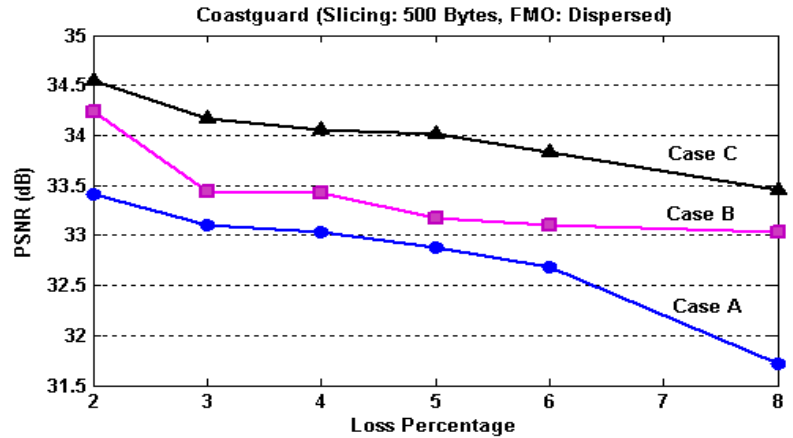


(b)

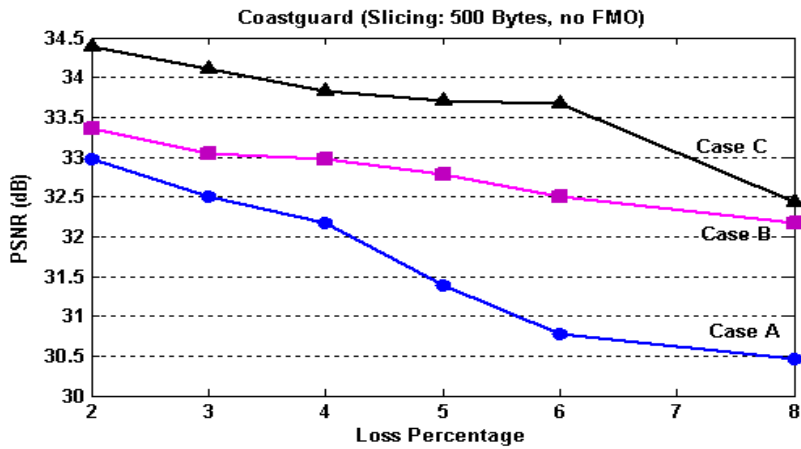


(c)

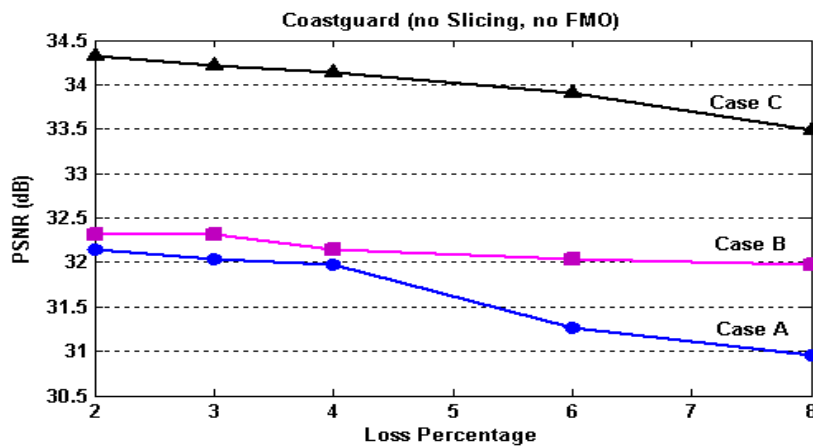
Figure 4.3 PSNR versus loss percentage (video: Foreman)



(a)



(b)



(c)

Figure 4.4 PSNR versus loss percentage (video: Coastguard)

We observe that in all cases data partitioning always provides a more efficient way of using the available reliability features of SCTP. As the loss ratio increases, PSNR decreases drastically in case A since only I slices are protected there. Case B performs better than case A but is lower in PSNR value than case C. This is due to the fact that, if we lose partition C, we still have some data available for those slices to be reconstructed. However, in case B, although some of the P frames are protected in addition to I frames, when we lose a P frame, then no information about that frame is available and the decoder has to perform concealment based on the surrounding information. So case C performs the best, achieving a PSNR gain in the range of 0.2dB-1dB compared to case B, and 1dB-3dB compared to case A.

4.4 Summary

In this chapter, we have examined how reliability features of SCTP can be assigned to different partitions and frame types of H.264 video. To set different priority levels on different partitions and frame types, we used the multistreaming and partial reliability properties of SCTP. We observe that the loss of partition B (belonging to an I frame) degrades the picture quality considerably, and may result in degradation even worse than when data partitioning is not used. If only partition C is lost, data partitioning always performs better than no data partitioning.

We have demonstrated that when extra capacity is available for retransmission of packets using SCTP, the data partitioning feature of H.264 can provide better results than solutions based on protecting all I frames and some P frames. In the next chapter, we consider alternatives to utilize the available capacity in error protecting schemes. Instead of using the available extra capacity for transport layer retransmission, the extra capacity is used for FEC and MAC layer retransmissions.

5 Efficient Utilization of Error Protection Techniques for Transmission of Data-Partitioned H.264 Video in a Capacity Constrained Network

5.1 Overview

Multimedia transmissions in wireless environment are governed by two factors: channel capacity and the average picture quality. The average picture quality depends on both video bit rate and wireless link loss rate. The video bit rate is determined by the encoder parameters and so the distortion can be referred to as encoder distortion. Choosing encoder parameters carefully can reduce this distortion at the cost of increasing the video bit rate. For successful video transmissions, the video bit rate should be less than or equal to the channel capacity assigned to the video stream. Given the fact that the average video bit rate is less than the assigned capacity, the extra capacity available can be used to improve the quality of the picture and for packet loss recovery. Since both encoder parameters and packet loss rate contribute to distortion, the effect of each of them to distortion should be minimized. In this chapter, we are not concerned about the coding layer. We consider that the coding parameters are not being changed to achieve the minimum encoder distortion; i.e., we consider the encoder distortion to have a fixed value. Here we are concerned about the transmission layer. We first find the optimum configuration of FEC and ARQ schemes in a capacity constrained network, and then utilize these findings in devising an optimal scheme for the transmissions of data-partitioned H.264 video over wireless networks.

This chapter is organized as follows. Section 5.2 briefly describes application layer FEC. Our proposed framework for optimal resource allocation is discussed in Section 5.3. Sections 5.4

and 5.5 present the loss-distortion model and optimal resource allocation for non-data-partitioned and data-partitioned video, respectively. Conclusions are presented in Section 5.6.

5.2 Application Layer FEC: A Brief Description

The key idea of FEC is to add structured redundancy to the transmitted data so that the receiver can reconstruct some amount of missing data. In application layer or packet level redundancy, parity packets are added to the source packets. In an (n, k) FEC codes, k source packets are encoded to n FEC packets, where $n > k$ and $\frac{n}{2} < k < n$ and any combination of the received k packets can retrieve the source information. Figure 5.1 depicts the process.

Reed-Solomon (RS) code is one of the most common FEC codes based on finite fields (Galois field). RS code is linear and systematic. The downside of RS code is that the implementation complexity increases if the block length, n , is large (e.g., larger than 255). However, here we are concerned with video applications (real time or streaming) which have delay and buffering constraints that make them incompatible with large blocks. Therefore, RS code is sufficient for multimedia applications and we are also considering RS code as our tool for FEC in the following analysis.

Employing FEC causes an increase in the transmission rate. If the video bitrate is r_v , then the protected bitstream rate is,

$$R_{FEC} = \frac{n}{k} r_v \quad (5.1)$$

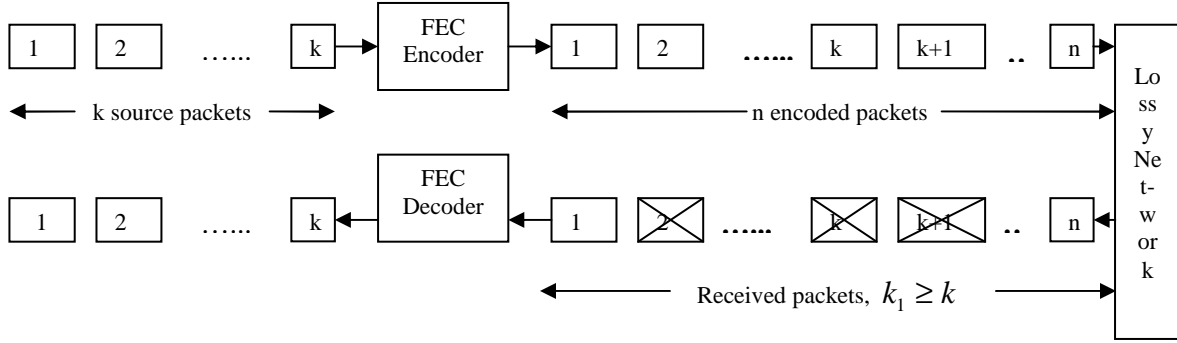


Figure 5.1 RS FEC scheme

The erasure failure rate for the RS FEC scheme is a function of the redundancy and also the packet loss ratio of the network and is given by,

$$P_{FEC}(k, p_{rtp}) = \sum_{i=n-k+1}^n C_i^n p_{rtp}^i (1 - p_{rtp})^{n-i} \quad (5.2)$$

where p_{rtp} is the application layer packet loss ratio. Since we are dealing with video transmissions, video frames/slices are packetized in RTP packets and so this also corresponds to the RTP packet loss ratio.

5.3 Optimal Resource Allocation: FEC versus Retransmission

As discussed in Section 5.1, distortion in video can be categorized into two groups: encoder distortion and distortion due to the loss of video packets. The encoder distortion is the result of choosing the encoder parameters and can be reduced by careful choice of these parameters. The second kind of distortion is due to the lossy nature of the wireless environment. Since most of the video codecs employ hybrid motion compensation, the effect of loss will propagate throughout the GOP resulting in a poor picture quality. So it is very important to provide sufficient error correction schemes to reduce the wireless loss ratio to a minimum value in order to have a good quality video. This is also our focus in this chapter;

i.e., we want to minimize the total distortion by minimizing the distortion due to loss of packets. The encoder distortion is considered constant in our analysis.

We are assuming here that the capacity allocated by admission control for video application is C kbps and the video bit rate is r_v kbps where $C > r_v$. Now the extra bandwidth available, $E = C - r_v$, can be used for error correction. As described in Section 5.1, FEC and ARQ are the two candidates for this purpose. In this section, we propose that there lies an optimum point between the application layer redundancy and MAC layer retransmission in a capacity constrained network. The enhancement in video quality achieved by employing any of them will be governed by bandwidth consumption and also by the loss ratio. If the loss ratio is too high, then even adding a high level of redundancy may not produce a good picture quality. On the other hand, the capacity and delay constraint prevent having a large number of retransmissions possible in real time communications. So a combination of these two schemes will result in optimum performance.

The framework of our scheme is depicted in Fig. 5.2. In application layer, the video packets are encoded using RS code and are passed to the transport and network layer. In the link layer retransmission is performed in case of any lost packets or erroneous decoding of the RS decoder. The rate needed for retransmission depends on the expected number of packets retransmitted and the maximum number of retransmissions allowed and is given by,

$$E_{rt} = \sum_{i=1}^X i(1-p)p^i \quad (5.3)$$

$$R_{rt} = E_{rt} \times R_{FEC} \quad (5.4)$$

where p is the data link layer loss ratio (i.e. if the physical layer BER is e , then $p = 1 - (1 - e)^L$, where L is the length of a packet), X is the maximum number of allowable retransmissions, E_{rt} is the expected number of packets retransmitted, R_{FEC} is the protected bitstream rate (from 5.1) and R_{rt} is the rate due to MAC layer retransmissions.

The loss ratio after X retransmissions becomes,

$$p' = (p)^{X+1} \quad (5.5)$$

For simplicity, we assume that one RTP packet is encapsulated in one RLC packet, i.e., $p_{rp} = p$. This is a reasonable assumption because by virtue of the ‘slicing’ feature of H.264, the video packet size can be adjusted according to the RLC size. After we have employed

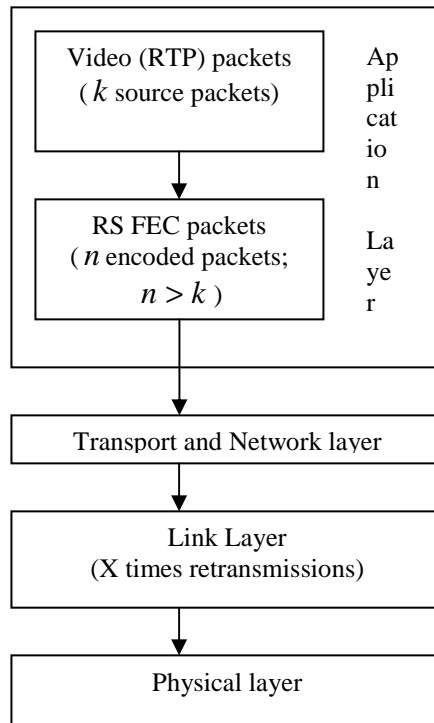


Figure 5.2 Framework of our scheme

both the FEC and MAC layer retransmissions, the effective loss ratio or effective erasure failure rate becomes,

$$P_{FEC} = \sum_{i=n-k+1}^n C_i^n (p')^i (1-p')^{n-i} \quad (5.6)$$

and the total rate for video transmissions,

$$R = R_{FEC} + R_{rt} \quad (5.7)$$

The above analysis is a general one and valid for a non-data-partitioned video. Here all the application layer packets are considered in the same layer. However, the concept of utilizing both FEC and ARQ will be more promising if we consider an application having packets/layers of different importance levels. Data-partitioned video is a good choice for this. DP is described in detail in Chapter 2.

These different importance levels of a data-partitioned video calls for employing different protection to different priority levels in a lossy environment. Another important fact is that, the bit rates of different partitions are different in a data-partitioned video, i.e., $r_A \neq r_{BC}$, where r_A and r_{BC} are the bit rates for partitions A and BC, respectively. In this chapter we consider partitions B and C in the same layer for simplicity and they will be referred as BC throughout this chapter. So for a capacity constrained network where we have a limited amount of extra capacity available for performance enhancement, we can use that capacity differentially to protect the partitions of varying importance.

The equations representing the bit rates and effective loss ratios for a data-partitioned video can be expressed as:

Partition A:

$$E_{rt_A} = \sum_{i=1}^{X_A} i(1-p)p^i$$

$$R_{rt_A} = E_{rt_A} \times R_{FEC_A} \quad (5.8)$$

$$p'_A = (p)^{X_A+1}$$

$$P_{FEC_A} = \sum_{i=n-k_A+1}^n C_i^n (p'_A)^i (1-p'_A)^{n-i}$$

Partition BC:

$$E_{rt_BC} = \sum_{i=1}^{X_{BC}} i(1-p)p^i$$

$$R_{rt_BC} = E_{rt_BC} \times R_{FEC_BC} \quad (5.9)$$

$$p'_{BC} = (p)^{X_{BC}+1}$$

$$P_{FEC_BC} = \sum_{i=n-k_{BC}+1}^n C_i^n (p'_{BC})^i (1-p'_{BC})^{n-i}$$

So the total rate for a data-partitioned video transmission is,

$$R = R_{FEC_A} + R_{FEC_BC} + R_{rt_A} + R_{rt_BC} \quad (5.10)$$

The notations have the same meaning as those for non-data-partitioned video with the addition of subscripts A and BC, denoting that they are for partitions A and BC, respectively.

The objective of our study is to minimize the effective loss ratio experienced by the video, i.e., to minimize the distortion caused by the transmission errors. We can formulate our study as an optimization problem that minimizes the video distortion subject to the capacity

constraint where the excess capacity is utilized by FEC and retransmission. We want to find the trade-off of joint utilization of FEC and ARQ that will maximize the capacity. For data-partitioned video, since different combinations of FEC and ARQ can be applied to partitions A and BC, it can also be said that we want to maximize the video quality subject to the total capacity constraint by choosing the optimal partition between different video layers (in this case, A and BC) given that FEC only, both FEC and ARQ, or only ARQ can be applied. It is worth mentioning here that we are concerned about the transmission layer only and are leaving the coding layer as it is. Since the capacity allocation between FEC and ARQ will be governed by the number of redundant packets $(n - k)$ for FEC and the maximum number of allowable retransmissions (X) , we can define our problem as,

$$\begin{aligned}
 & \min_{k, X} D \\
 & \text{subject to } R \leq C \\
 & (n/2) \leq k \leq n \\
 & 0 \leq X \leq m
 \end{aligned} \tag{5.11}$$

where D is the distortion (representing both non-data-partitioned and data-partitioned video). Here the range of k is selected to maintain a reasonable throughput at the receiver. The lower limit for X is chosen as 0 which corresponds to the broadcasting application. The value of the upper limit, m , is critical since retransmitting a packet several times introduces a large amount of delay and also is not reasonable considering the amount of overhead each retransmission consumes. If the encoded video is of size M bits and the total overhead including application, transport, IP and MAC layer is H bits, then the total bits used for

retransmission is, $B = E_r \times (M + H)$ bits. For 3G and WLAN, H corresponds to 60 and 72 Bytes respectively which correspond to the usage of 6% (3G) or 7.2% (WLAN) of the bandwidth for header transmission (for a 1000 Byte packet) and can be considered a wastage taking into account the limited availability of resources in a wireless environment. Considering these facts, we have limited m to 3.

5.4 Optimal Resource Allocation: Non-Data-Partitioned Video

5.4.1 Loss-Distortion Model

As described in Section 5.3, the effect of packet losses on picture quality is complex due to the hybrid nature of the codecs. It depends on the error concealment algorithm and also on the position of the lost frames in the GOP. In this section, we devise a loss distortion model for a non-data-partitioned video stream. For non-data-partitioned video, we consider all the packets in the same layer and have used a simple linear model to describe the loss-distortion relationship. The motive behind this model is that if there is no loss, then we have the minimum distortion. But in the presence of loss, distortion is added to the no-loss case distortion, i.e.,

$$D_{ndp} = D_{no_loss} + (p_{effective} \times D_{loss}) \quad (5.12)$$

where D_{ndp} is the total distortion for a non-data-partitioned video, measured in Mean Squared Error (MSE), D_{no_loss} is the MSE value in the no loss case, $p_{effective}$ is the effective loss ratio (after FEC and ARQ have been employed, as in (5.6)), and D_{loss} incorporates the effect of both concealment and inter frame error propagation. We have used curve fitting to the actual values to determine these parameters. Figure 5.3 shows the model values for the

Foreman QCIF sequence. The value of D_{no_loss} and D_{loss} are video sequence and codec parameters dependent. The figure shows that this linear model follows the actual values for a loss ratio smaller than 10%. This is obvious since the distortion depends largely on the position of the error and the error propagates throughout the GOP rendering a non-linear relation with loss ratio. Again, this linear model assumes that the individual errors are uncorrelated, which is true if the errors are temporally and/or spatially separated in the decoded video sequence. When there is a high loss ratio, the errors become correlated and so, our model provides inaccurate results. However, for a loss ratio higher than 10%, the distortion becomes so prominent that it is very difficult to achieve a reasonable picture quality [36]. Although employing a more sophisticated loss-distortion model will certainly enhance the effectiveness of this study, this simple model in no way deters the objective of our work.

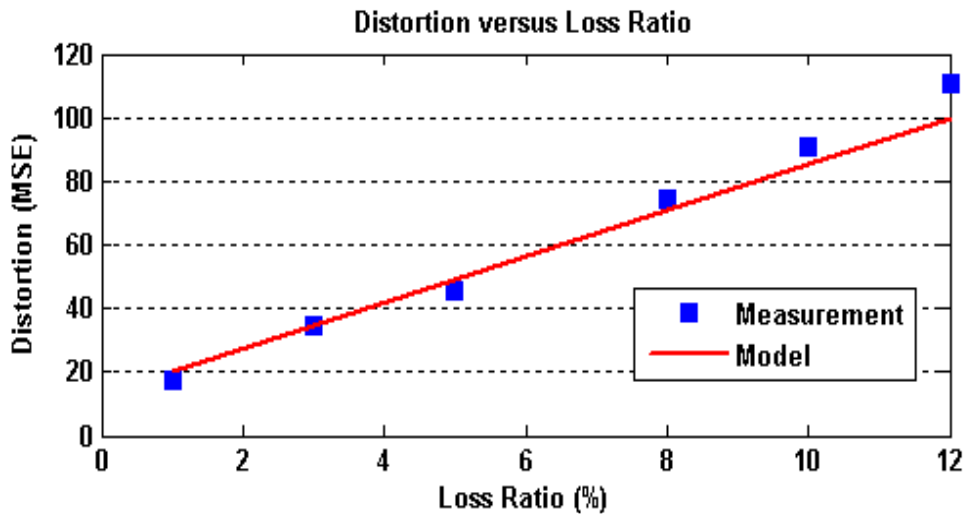
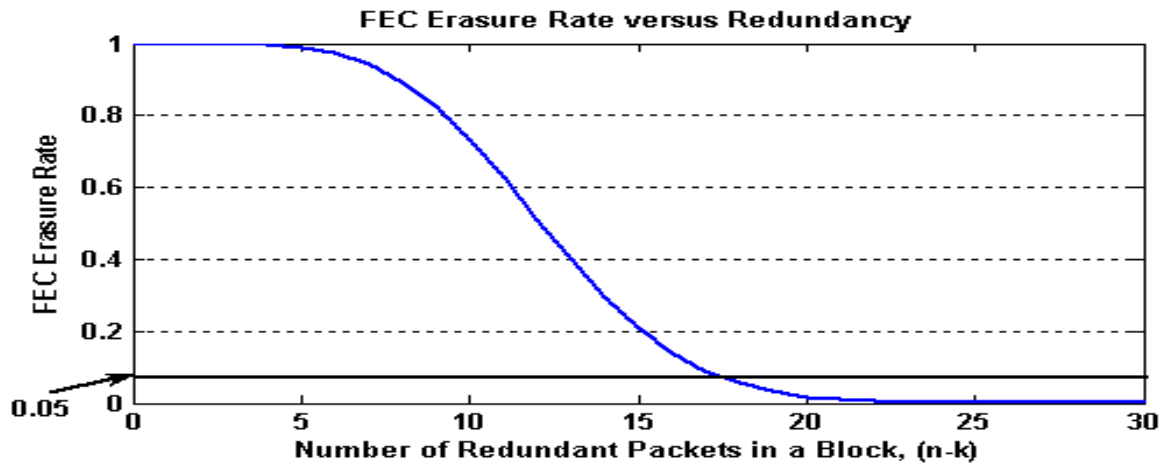


Figure 5.3 Loss distortion model for non-DP video (Foreman sequence).

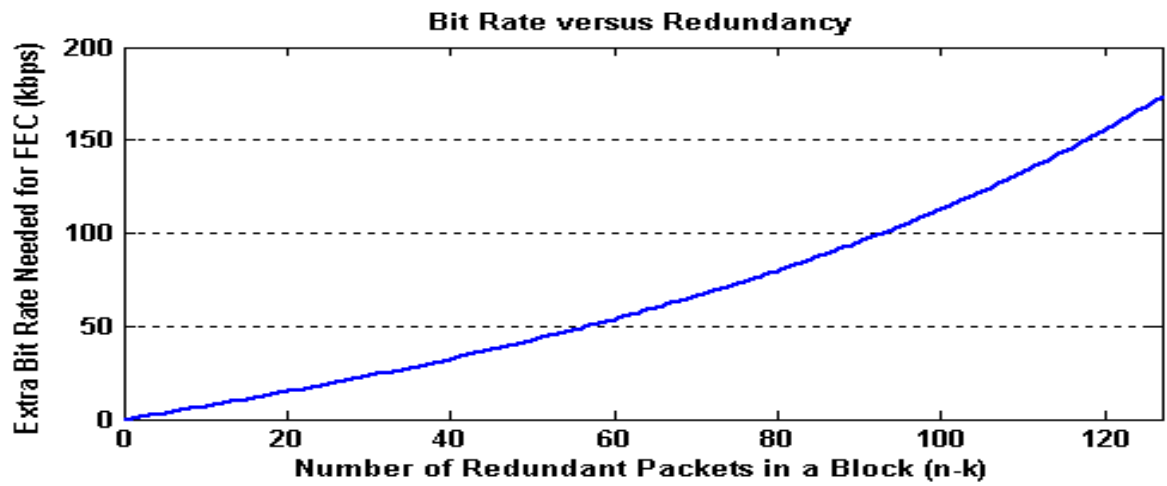
5.4.2 Optimal Resource Allocation

To find the optimal partition point between application layer redundancy and MAC layer retransmissions, we have simulated the transmissions of H.264 video over a network abstracted as depicted in Section 5.3. We consider a streaming application and assume that the receiver buffer is sufficiently large. The Foreman video sequence [34] is encoded using the JM 13.1 software [31] with quantization parameter 28. For application layer redundancy, we use a block size, $n = 255$. We perform the optimization to find the values of k and X , that will reduce the effective loss ratio to a minimum value and hence maximize the video quality for a given capacity and loss ratio.

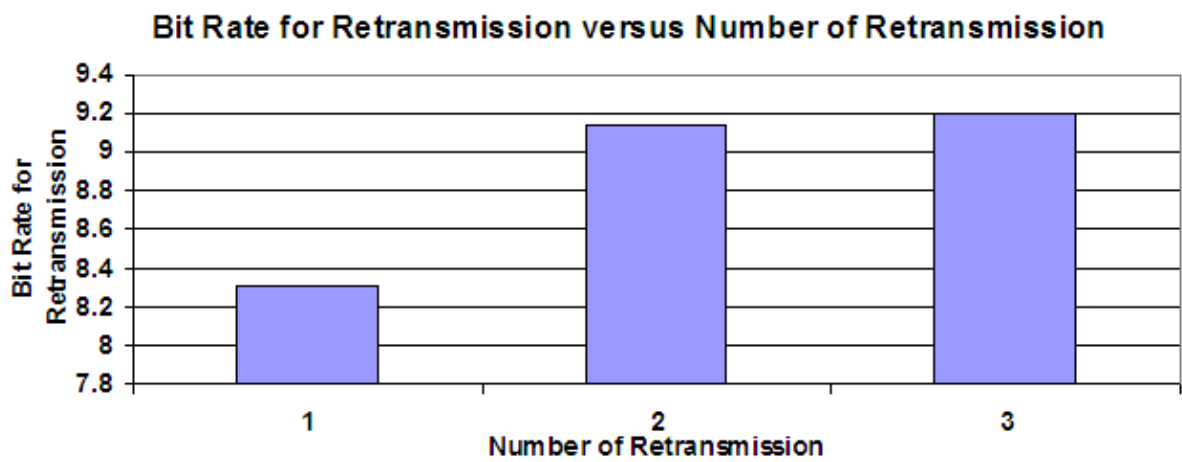
As discussed in Section 5.3, the optimum division between redundancy and retransmission largely depends on the capacity and loss ratio. We know that application layer FEC increases the bandwidth requirement and for each loss ratio, there should be a sufficient number of redundant packets provided. On the other hand, retransmissions also increase the bandwidth usage. Figure 5.4 shows the bandwidth requirements for both ARQ and FEC case. The FEC erasure rate versus number of redundant packets is shown in Fig. 5.4a for a loss ratio of 5%. To have a loss ratio of 5% or below, we need at least approximately 20 redundant packets which corresponds to the requirement of 18 kbps of extra bandwidth. On the other hand, retransmitting a lost packet three times (the maximum number of retransmissions allowed) require 9.2 kbps extra bandwidth while reducing the loss ratio to a value of $(0.05)^4 = 6.25 \times 10^{-4}$. So in such a capacity constrained network, retransmission will perform better. However, as the loss ratio increases, to have a good picture quality, we need to



(a)



(b)

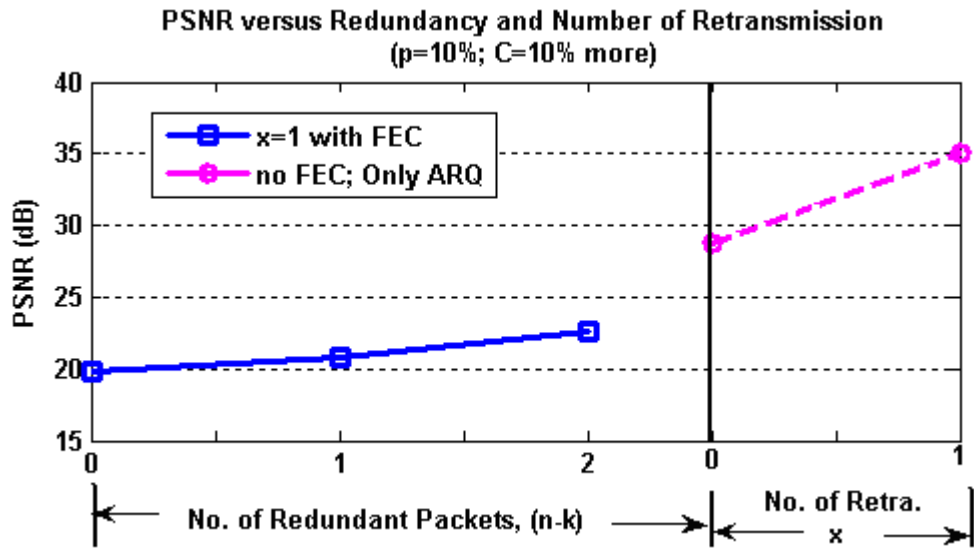


(c)

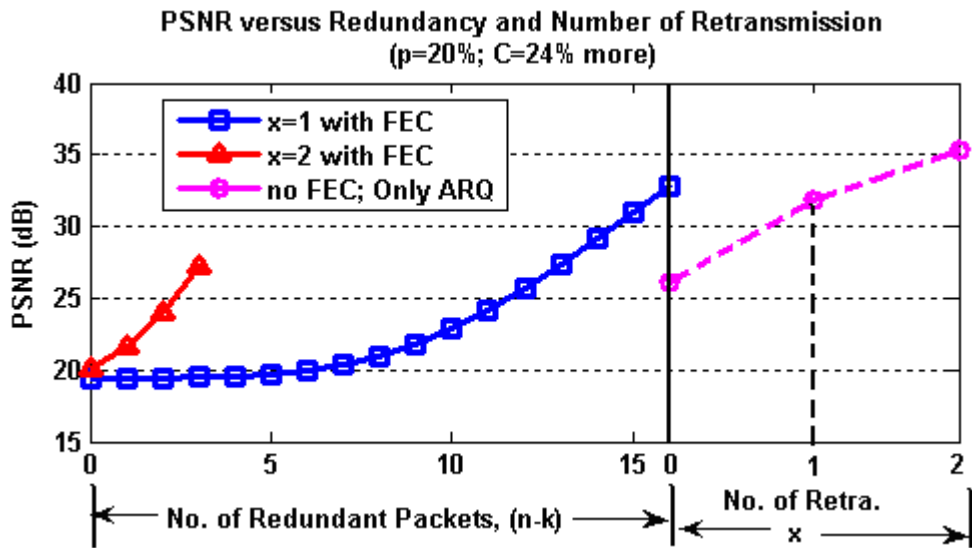
Figure 5.4 (a) FEC erasure rate versus redundancy, (b) extra bit rate needed for FEC versus redundancy, and (c) extra bit rate needed for retransmission versus redundancy.

allocate more bandwidth. For a high loss ratio (e.g., 30% or 40%), 3 retransmissions will consume a huge amount of bandwidth. Certainly FEC alone cannot give an optimal performance because a high loss ratio will require a very high redundancy as well as a huge amount of extra capacity. So in that case, an optimum point will be reached in a combination of 1 or 2 retransmissions and FEC.

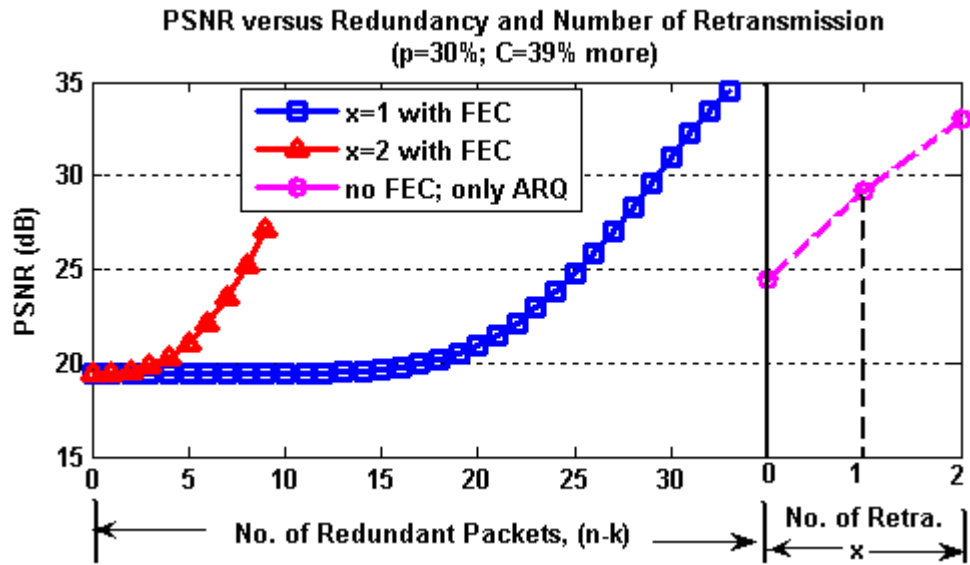
Figure 5.5 depicts the simulation results for different capacities and loss ratios. In Fig. 5.5(a), for 10% loss ratio and 10% more capacity allocation, the first retransmission without any FEC results in the optimum performance. When FEC is incorporated with the first retransmission, the capacity allocated does not let the FEC to have sufficient redundancy and so results in sub-optimal PSNR values. However, as the loss ratio increases, the capacity allocation needs to be increased and only retransmission or FEC proves to be insufficient to achieve the optimum value. For 20% loss ratio and 24% more capacity allocation, a second retransmission without FEC results in the optimum point but for 30% (39% more capacity) and 40% (60% more capacity) loss ratio, the optimum point is found with first retransmission and $k = 222$ and $k = 198$ respectively. In such a high loss ratio, second or even third retransmission cannot give any improvement in the video quality, signifying the importance of employing both FEC and retransmission.



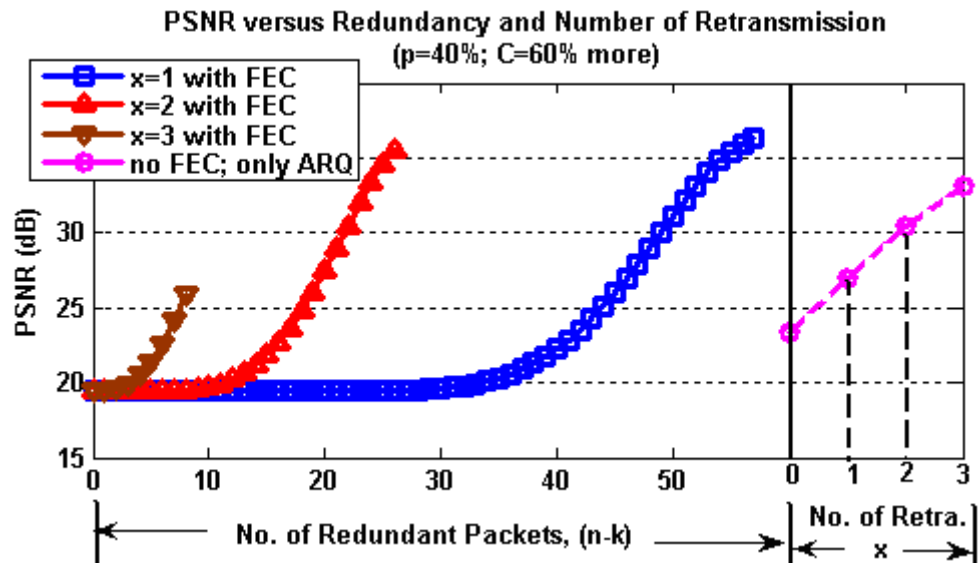
(a)



(b)



(c)



(d)

Figure 5.5 PSNR versus application layer redundancy and number of retransmissions for different loss ratios and capacity; (a) $p=10\%$ and $C=10\%$ more, (b) $p=20\%$ and $C=24\%$ more, (c) $p=30\%$ and $C=39\%$ more, and (d) $p=40\%$ and $C=60\%$ more.

5.5 Optimal Resource Allocation: Data-Partitioned Video

5.5.1 Loss-Distortion Model

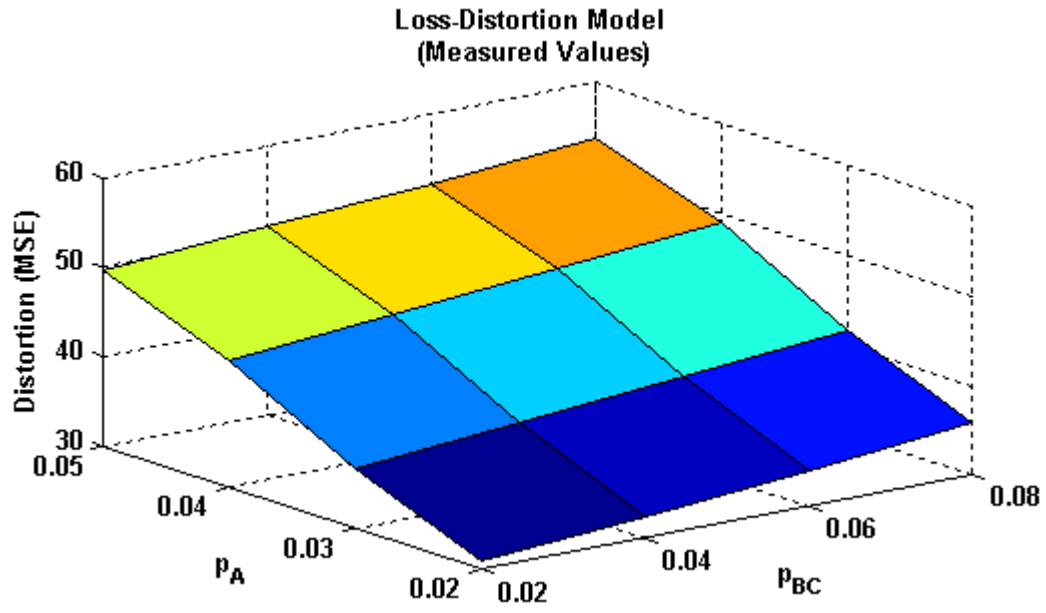
The loss-distortion model for a data-partitioned video is not as simple as (5.12) due to the fact that loss of different partitions affects the picture quality differentially. If partition B and/or C is lost, the information received in partition A can be used to conceal the error effectively. However, receiving partition B and/or C without partition A will cause the whole frame/slice to be discarded. Remembering these facts, the loss-distortion model can be represented as:

$$D_{dp} = (1 - p_A)(1 - p_{BC})D_{enc} + (1 - p_A)p_{BC}D_A + p_AD_{ec} \quad (5.13)$$

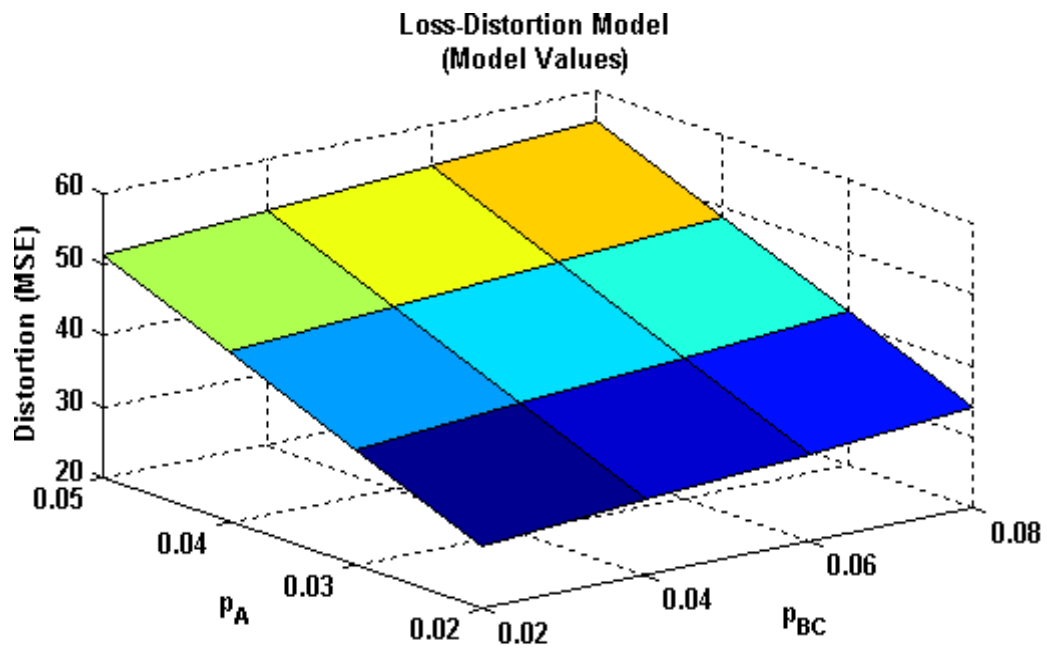
Here D_{dp} represents the total distortion for a data-partitioned video, p_A and p_{BC} are the effective loss ratios of partitions A and BC, respectively (as in (5.8) and (5.9)), D_{enc} , D_A and D_{ec} represent the encoder distortion (when all the partitions are received correctly), distortion due to loss of partition BC and error concealment distortion (when all partitions or partition A is lost), respectively. We find the values of these parameters using the best fit approach on the experimental data. Figure 5.6 shows that our model closely approximates the actual values.

5.5.2 Optimal Resource Allocation

The data partitioning property enables us to employ unequal error protection by virtue of having packets of varying importance. Since partition A is the most important, it should be protected with a higher priority. However, in a capacity constrained network, protecting only partition A may not result in the highest achievable picture quality. Again, we have two tools



(a)



(b)

Figure 5.6 Loss distortion model for a DP video (Foreman Sequence): (a) Measured values, and (b) Model values.

to perform the error correction: FEC and ARQ. They can be used to protect partitions A and BC in different combinations:

Case i: Only retransmissions are employed for all the partitions,

Case ii: Partition A is protected by both FEC and retransmissions whereas partition BC is protected only by retransmissions,

Case iii: All the partitions are protected by both FEC and retransmissions.

Before describing the dynamics of each of these three cases in detail, we present the optimum point at different loss ratios and available excess capacities. The simulations use a similar configuration as we have used for non-data-partitioned video. Table 5.1 shows the optimum point at different loss ratios and different available capacities.

From Table 5.1, we observe that the optimum point changes with the loss ratios and capacity allocation. When the loss ratio is not that high, and the available extra capacity is limited, then the use of retransmissions only without application layer redundancy performs the best. This is in accordance with our discussion in the no data partitioning case. This is due to the fact that, application layer FEC requires sufficient amount of redundancy (and hence capacity) to reduce the loss ratio to a lower value. Insufficient redundancy may cause the loss ratio to reach a value as high as 1. However, as the loss ratio increases, use of retransmissions only proves to be insufficient to achieve the optimal value. In the case of 20% loss ratio (25% extra available capacity), we need to apply retransmissions and FEC for partition A and only retransmissions for partition BC. If the capacity allows, then certainly providing sufficient FEC for partition BC would have resulted in the optimal value. For an even higher loss ratio (40%), the maximum quality is achieved when both FEC and retransmissions are employed

Table 5.1 Optimum point at different loss ratios and available extra capacity.

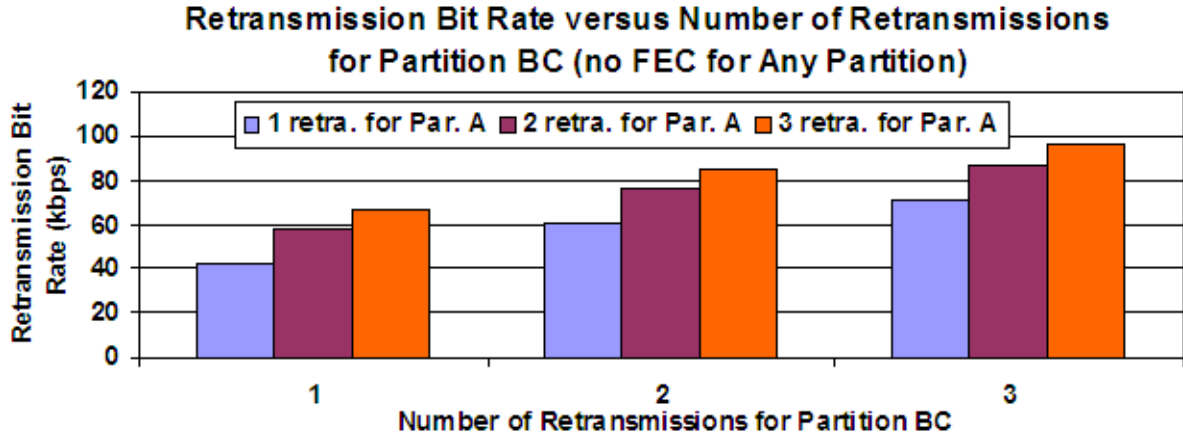
Loss Ratio (p)	Available extra capacity (E)	Partition A		Partition BC	
		Number of Retransmission (x_A)	Application Layer Redundancy ($n-k_A$)	Number of Retransmission (x_{BC})	Application Layer Redundancy ($n-k_{BC}$)
5%	5%	3	No	1	No
10%	10%	3	No	1	No
20%	25%	3	7	2	No
40%	60%	1	61	1	54

[‘No’ means no application layer redundancy has been applied]

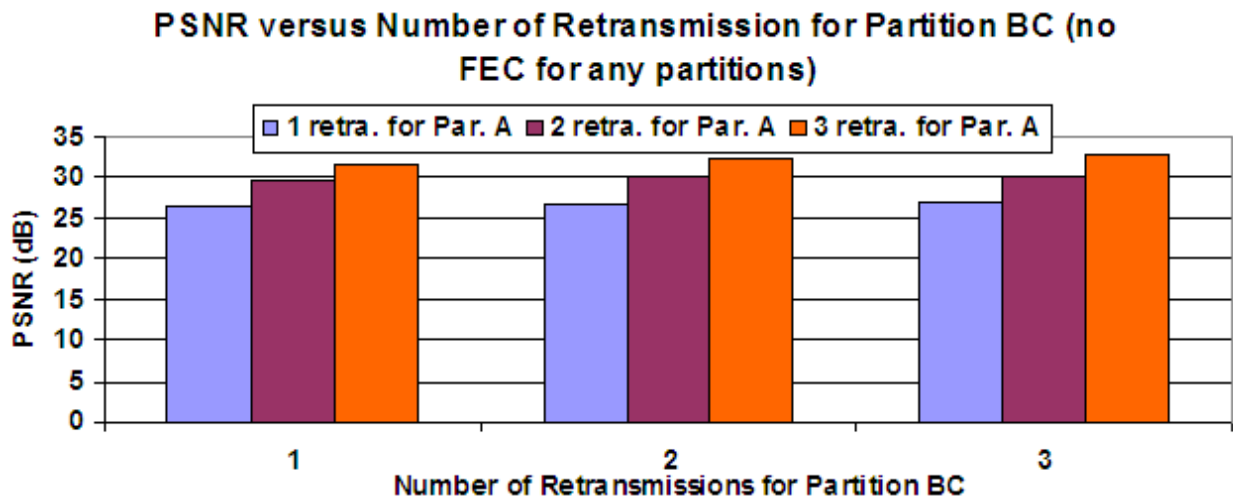
for all the partitions.

Now to have a detailed analysis, we discuss cases (i), (ii), and (iii) in the following paragraphs. We present the analysis for a loss ratio of 40% and 60% available extra capacity. We have assigned such a high capacity to understand the dynamics of changing the optimal point properly.

Case i. Only retransmissions are employed for both partitions A and BC: The extra capacity needed for retransmissions for partitions A and BC and the achieved PSNR are shown in Fig. 5.7. Figure 5.7a shows the extra bit rate needed for retransmissions of partitions A and BC. It shows that increasing the number of retransmissions, although, causes huge bandwidth consumption, the gain in PSNR is not that much for higher retransmissions for partition BC. This is obvious since partition BC contains the less important information than partition A. So the capacity used for retransmitting partition BC, can be employed for protection of partition A by means of FEC or retransmission and can result in good picture quality.



(a)



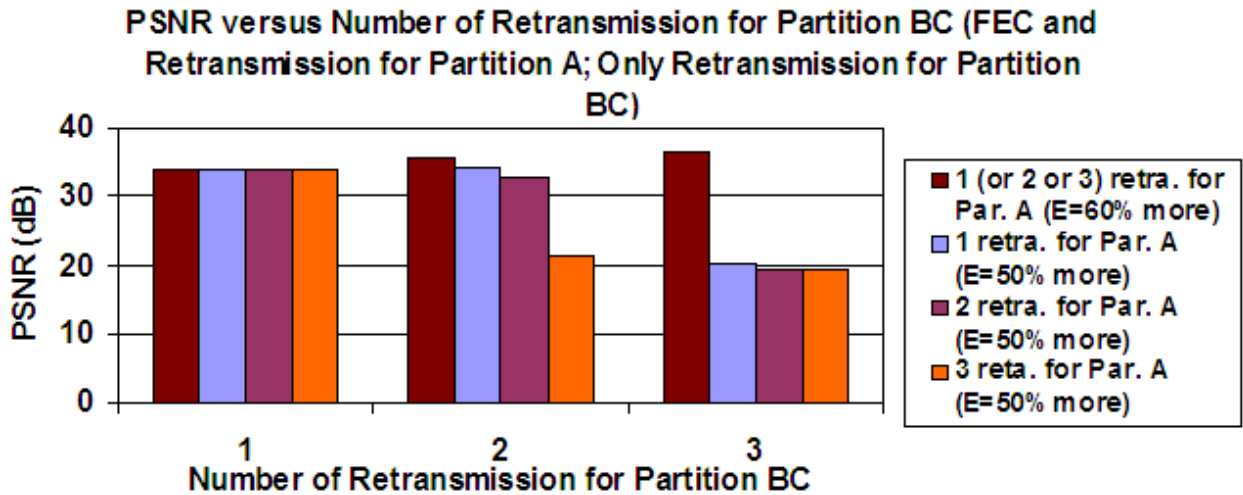
(b)

Figure 5.7 (a) Retransmission bit rate versus number of retransmissions for partition BC, and (b) PSNR versus number of retransmissions for partition BC.

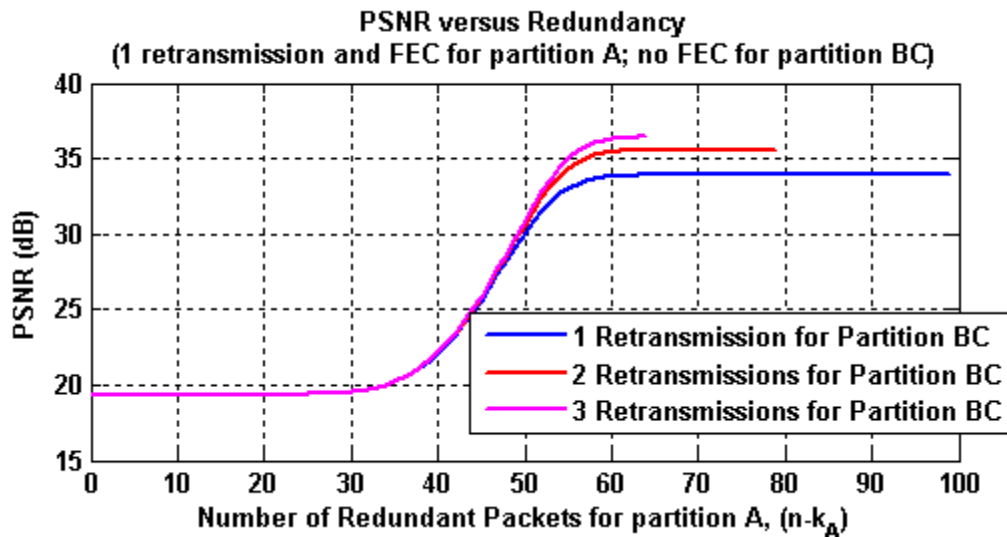
Case ii. Partition A is protected by both FEC and retransmissions whereas partition BC is protected only by retransmissions: In this case, we protect partition A by means of FEC and retransmissions. We employ only retransmissions for partition BC. We assume that in a capacity constrained network, this will result in optimal performance since we are allowing

sufficient bandwidth for the most important partitions, but if the available extra capacity is large, then probably it will not result in the maximum PSNR value. Figure 5.8 shows the variation of PSNR. In Fig. 5.8a, we show the variation of PSNR with number of retransmissions for BC for two capacities. In the first case when we have 60% more capacity allocation, retransmitting partition A once, twice or thrice along with FEC results in almost the same PSNR values. This is because the capacity is large enough to allow sufficient protection in each case. In the second case, when we have 50% extra capacity available, we found that the optimum point is reached when partition A is retransmitted once with FEC and partition BC is retransmitted twice. Retransmitting partition A twice or thrice consume too much bandwidth, thus preventing FEC to have sufficient redundancy.

In Fig. 5.8b, the change in PSNR with FEC for partition A is observed for different retransmissions of partition BC. We observe here that when 1 retransmission is allowed for partition BC, the rest of the available capacity is allocated for partition A. However, we see that after adding 60 redundant packets, the PSNR value becomes constant which means that adding more redundancy for partition A does not help to improve the video quality, so that capacity can be better used for the protection of partition B. In the second case, when 2 retransmissions are allowed for partition BC, we see that again the PSNR becomes constant signifying the need to allocate that protection for partition BC. So we can conclude from our observation that, rather than applying all the protection to partition A, if capacity allows, it is always better to provide sufficient protection for partition BC.



(a)



(b)

Figure 5.8 (a) PSNR versus number of retransmissions for partition BC, and (b) PSNR versus application layer redundancy for partition A.

Case iii. All the partitions are protected by both FEC and retransmissions: Here all the partitions are protected by both FEC and retransmissions. This case will require a large amount of bandwidth to provide the maximum PSNR values. Figure 5.9 depicts PSNR versus number of retransmissions for partition BC for this case. We see here that 1 retransmission

with FEC (for both partition A and BC) performs the best. The reason is that when more retransmissions are allowed for each partition, bandwidth allocated to FEC is reduced. So, the redundancy in that available bandwidth is not sufficient to reduce the effective loss ratio to a low value. Again, in a loss ratio as high as 40%, use of retransmissions only is not good enough to achieve the optimal performance. So it needs FEC for both of them and this plays the major role to reduce the effective loss ratio if one retransmission is allowed for both the partitions. Figure 5.10 shows the PSNR values at different redundancy level after 1 retransmission has been applied. The blue plane indicates the lowest PSNR value which corresponds to the case when the codec conceals the lost frame/slice. The highest point is reached when sufficient redundancy is provided for both of the partitions; in this particular case this corresponds to $k_A = 194$ and $k_{BC} = 201$.

Now let us compare all the three cases described above. Figure 5.11 shows all of them together. It seems that the second case will achieve the highest quality in most of the capacity constrained scenarios. To reach a conclusion about this, we show the optimal point for a loss ratio of 40% and different capacity allocation. The results are shown in Table 5.2. Here we see that except the last scenario (where a huge amount of bandwidth is available for protection), case (ii) performs the best. This is due to the fact that case (ii) allows a sufficient protection for the most important partition A and also it ensures reasonable protection for partition BC.

PSNR versus Retransmission for Partition BC (All the Partitions' are Protected by FEC and ARQ)

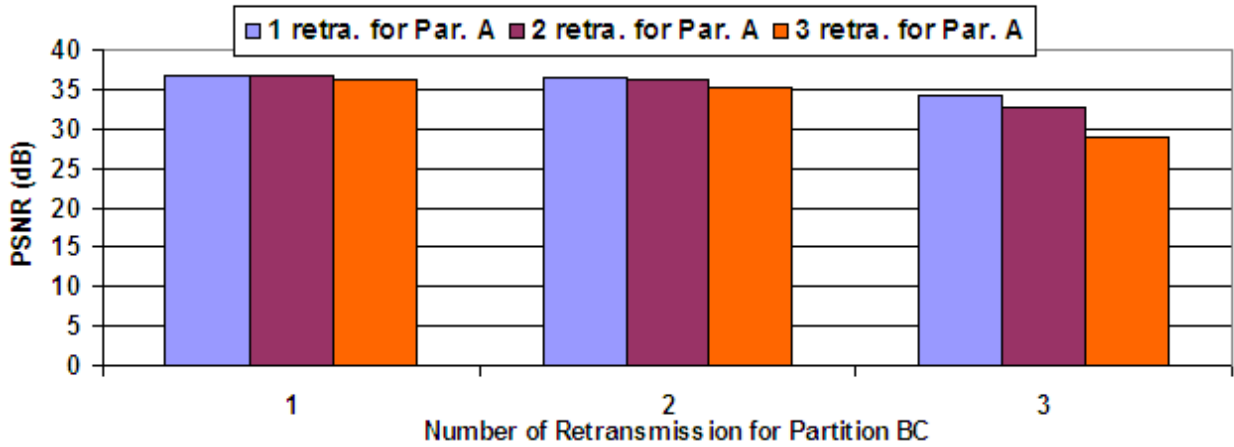


Figure 5.9 PSNR versus number of retransmission for partition BC.

PSNR versus Redundancy for Partition A and BC

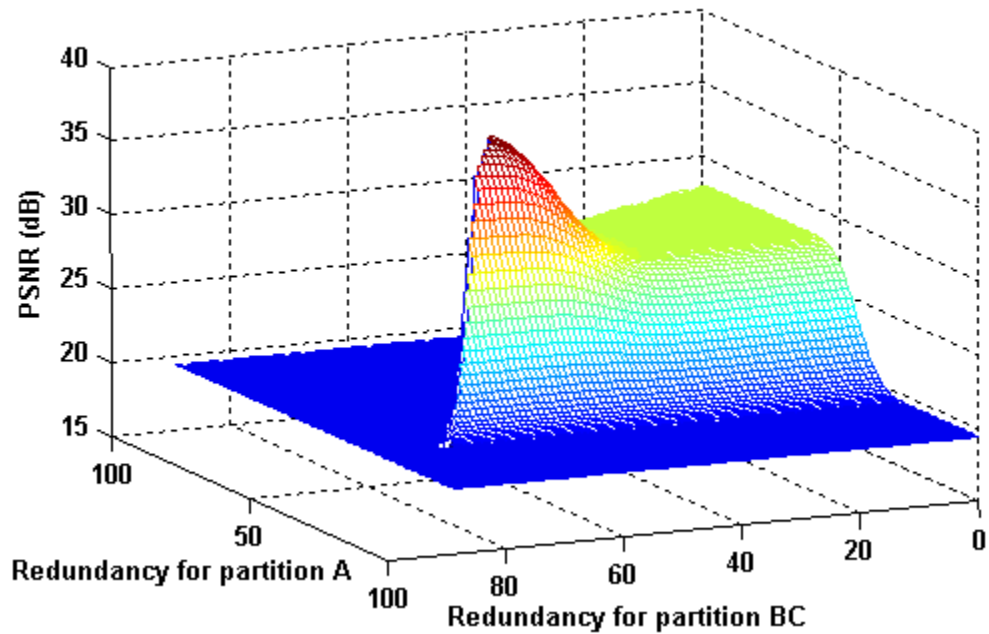


Figure 5.10 PSNR versus redundancy for partitions A and BC.

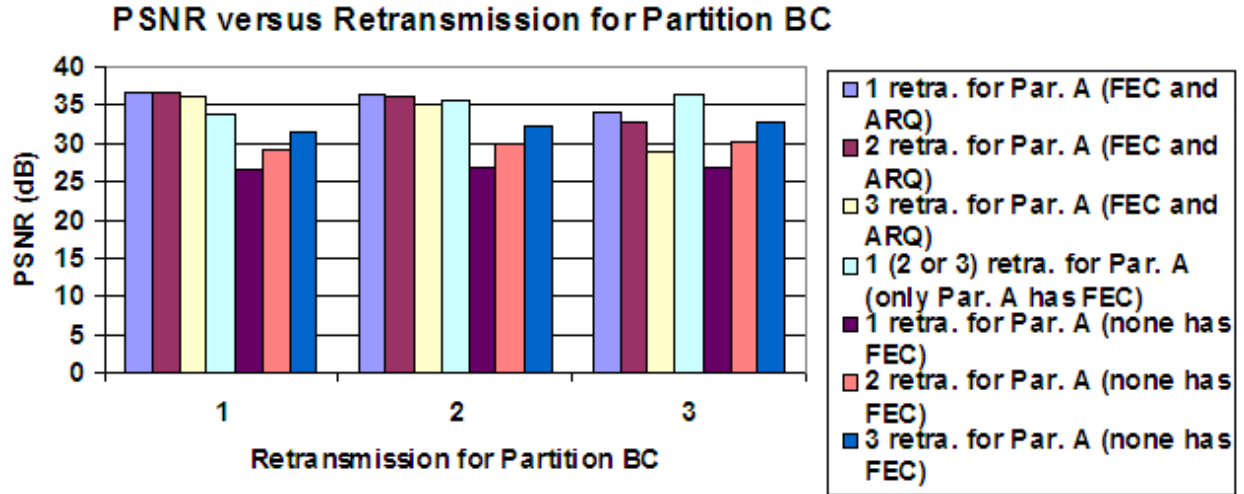


Figure 5.11 PSNR versus Retransmission for Partition BC.

Table 5.2 Optimum point at loss ratio of 40% and different capacity allocation.

Loss Ratio (p)	Available extra capacity (E)	Partition A		Partition BC	
		Number of Retransmission (x_A)	Application Layer Redundancy ($n-k_A$)	Number of Retransmission (x_{BC})	Application Layer Redundancy ($n-k_{BC}$)
40%	50% more	1	55	2	No
	54% more	1	64	2	No
	56% more	1	70	2	No
	60% more	1	61	1	54

5.6 Summary

In this chapter, we maximize the video quality by choosing the optimal point in application layer and MAC layer redundancy. We have shown that, in a capacity constrained network, especially in a highly lossy environment, neither FEC nor ARQ alone results in optimum performance. Rather a combination of these two techniques is effective to reduce the loss ratio to a low value. This is true for both data-partitioned and non-data-partitioned video but is more challenging in the data-partitioned case. The different impacts of loss of different

partitions on quality enable different combinations of FEC and retransmissions to be used for different partitions. In most cases, use of FEC and retransmissions for partition A and retransmissions only for partition BC results in the maximum PSNR values.

6 Conclusions and Future Work

6.1 Conclusions

In this thesis, efficient utilization of the error resilient features of H.264 video in a capacity constrained error prone wireless network has been studied. The study has determined: (1) The optimum slice size for a 3G network that will enhance the picture quality, (2) Joint optimization between the data partitioning property of H.264 and the partial reliability extension feature of SCTP, and (3) The optimum utilization of FEC and MAC layer retransmissions that will maximize the video quality (for both non-data-partitioned and data partitioned video) in a capacity constrained network.

In Chapter 3, the video packetization scheme for conversational video service over 3G networks has been examined. It is found that in general a smaller slice size produces a better picture quality. However, the extent in which the slice size can be reduced is governed by the video bit rate. Since the overhead associated with smaller slices cause the video bit rate to increase, and for successful video transmissions, the video bit rate should be less than the channel capacity, the smallest slice size that should be used for the best video quality is the one that will match the video bit rate with the channel capacity.

In Chapter 4, the transmissions of data-partitioned video using SCTP as the transport layer protocol has been studied. By virtue of the multistreaming and partial reliability features of SCTP, different reliability levels can be set for different partitions of H.264 video according to their importance. To have a valid comparison between non-data-partitioned and data-partitioned video, the average number of bytes transmitted per GOP is kept the same in both of the cases. It is found that sending both partitions A and B in a reliable stream results in the

maximum quality. However, if partition B is sent unreliably and intra information (partition B) is lost from an I frame, then the quality is poor and possibly worse than the non-data-partitioned video since in the later case I frames are always sent reliably and thus affected less in terms of error propagation.

In Chapter 5, an optimization scheme has been proposed that will maximize the video quality in a capacity constrained network. It is found that, if the loss ratio is not that high, then use of retransmissions only results in the best performance. However, as the loss ratio increases, then neither FEC nor retransmissions alone are sufficient to maximize the video quality. A combination of these two mechanisms can reduce the loss ratio to the minimum value. This is true for both data-partitioned and non-data-partitioned video.

6.2 Future Work

In our scheme of devising the trade-off between FEC and ARQ in Chapter 5, we have not considered the encoder distortion. However, inclusion of coding parameters will add a new dimension to this scheme. For example, we can consider the QP as a variable to obtain the highest possible quality. Increasing QP increases distortion and decreases bit rate whereas decreasing QP does the reverse. Also the bit rates allocated for partitions A, B and C change with the change of QP. If the capacity allocated for video transmissions is very restricted and the environment is lossy, then it might be a good idea to increase QP (increasing the encoder distortion) and use the saved bandwidth for error protection. For a more detailed analysis, let us consider the case of 10% packet loss rate. For QP=27, 28 and 29, the video bit rates are 197.91 kbps, 175.61 kbps and 153.14 kbps, respectively. If the available capacity is $C_1 = 193$ kbps, then it may be appropriate to apply QP=28 (the result for this case is presented in Section 5.5; the highest PSNR achieved is 37.085 dB). However, if the available

capacity is $C_1 = 180$ kbps (which is 17% more than the bit rate needed for QP=29 but only 2.5% more than the bit rate needed for QP=28); then certainly use of QP=28 will not result in optimum performance. In fact such a small amount of extra capacity will accommodate neither sufficient application layer redundancy nor even one retransmission. So the quality achieved is 28.4 dB. However, if we use QP=29, then the available extra capacity (17%) is large enough to allow sufficient redundancy to be employed. The PSNR achieved in this case is 36.36 dB which is certainly the best achievable quality for this scenario. So for future work we can develop a framework where depending on the information of available capacity, suitable coding parameters (e.g., QP) will be chosen so that the scheme can minimize both the coding and loss distortion simultaneously.

Bibliography

- [1] “Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264/ISO/IEC 14 496-10 AVC,” in Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVTG050, 2003.
- [2] R. Stewart, Q. Xie, K. Morneault, C. Sharp, H. Schwarzbauer, T. Taylor, I. Rytina, M. Kalla, L. Zhang, and V. Paxson, Stream Control Transmission Protocol (SCTP), RFC 2960, 2000.
- [3] N. Kamaci, and Y. Altunbasak, “Performance Comparison of the Emerging H.264 Video Coding Standard with the Existing Standards,” in *proc. of the Intl. Conf. on Multimedia and Expo (ICME '03)*, vol. 2. pp. 345–348, Washington, DC, USA.
- [4] S. Kumar, L. Xu, M. K. Mandal, S. Panchanathan, “Error Resiliency Scheme in H.264/AVC Standard,” *Journal of Vis. Commun. Image R.*, vol.17, pp.425-450, 2006.
- [5] T. Stockhammer, M. M. Hannuksela, and T. Wiegand, “H.264/AVC in Wireless Environments,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol.13, no.7, pp. 657-673, July 2003.
- [6] L. Liu, S. Zhang, X. Ye, Y. Zhang, “Error Resilience Schemes of H.264/AVC for 3G Conversational Video Services,” in *proc. of the 5th Intl. Conf. on Computer and Information Technology (CIT'05)*, pp. 657 – 661, Sept. 2005.
- [7] T. Schierl, M. Kampmann, and T. Wiegand, “H.264/AVC Interleaving for 3G Wireless Video Streaming,” in *proc. of IEEE Intl. Conf. on Multimedia and Expo. (ICME 2005)*, 6-8 July 2005.

- [8] Q. Qu, Y. Pei, and J.W. Modestino, "Robust H.264 Video Coding and Transmission Over Bursty Packet-Loss Wireless Networks," *IEEE Vehicular Technology Conf. (VTC 2003-Fall)*, vol. 5, pp. 3395 – 3399, Oct. 2003.
- [9] A. Ksentini, M. Naimi, and A. Gueroui, "Toward an Improvement of H.264 Video Transmission over IEEE 802.11e through a Cross-Layer Architecture," *IEEE Communications Magazine*, vol. 44, issue 1, pp. 107 – 114, Jan. 2006.
- [10] Y. P. Fallah, D. Koskinen, A. Shahabi, F. Karim, and P. Nasiopoulos, "A Cross Layer Optimization Mechanism to Improve H.264 Video Transmission Over WLANs," in *proc. of Consumer, Communications and Networking Conference (CCNC) 2007*, pp. 875-879, Jan 2007.
- [11] Z. Lifan, S. Yanlei, and L. Ju, "The Performance Study of Transmitting MPEG4 over SCTP," in *proc. of IEEE Intl. Conf. on Neural Networks & Signal Processing*, pp. 1639-1642, December 2003.
- [12] A. Balk, M. Sigler, M. Gerla, and M. Sanadidi, "Investigation of MPEG-4 Video Streaming over SCTP," in *proc. of IIS World Multiconference on Systemics, Cybernetics, and Informatics (SCI 2002)*, pp. 337-340, Orlando, FL, USA, July 2002.
- [13] H. Wang, Y. Jin, and W. Wang, "The performance Comparison of PRSCTP, TCP and UDP for MPEG-4 Multimedia Traffic in Mobile Network," in *proc. of ICCT 2003*, pp. 403-406.
- [14] A. Argyriou, "A Novel End-to-End Architecture for H.264 Video Streaming over the Internet," *Telecommunication Systems*, vol. 28, issue. 2, pp. 133-150, Feb 2005.

- [15] M. N. E. Derini, and A. A. Elshikh, "MPEG-4 Video Transfer with SCTP-Friendly Rate Control," in *proc. of the 2nd Intl. Conf. on Innovations in Information Technology*, 2005.
- [16] Y. P. Fallah, P. Nasiopoulos, and H. Alnuweiri, "Efficient Transmission of H.264 Video over MultiRate IEEE 802.11e WLANs," *EURASIP Journal on Wireless Communications and Networking*, vol. 2008, Article ID: 480293, 14 pages, 2008, doi: 10.1155/2008/480293.
- [17] Y. P. Fallah, P. Nasiopoulos, and H. Alnuweiri, "Scheduled and Contention Access Transmission of Partitioned H.264 Video over WLANs," *IEEE Global Telecommunications Conference 2007*, pp. 2134 – 2139, Nov. 2007.
- [18] L. Rizzo, "Effective Erasure Codes for Reliable Computer Communication Protocols," *Computer Communication Review*, vol. 27, no. 2, pp. 24-36, April 1997.
- [19] C. H. Lin, C. H. Ke, C. K. Shieh, and N. K. Chilamkurty, "An Enhanced Adaptive FEC Mechanism for Video Delivery over Wireless Networks," in *proc. of Intl. Conf. on Networking and Services 2006 (ICNS'06)*, p. 106, Santa Clara, Calif, USA, July 2006.
- [20] P. de Cuetos, and K. W. Ross, "Unified Framework for Optimal Video Coding," in *proc. of INFOCOM 2004*, pp. 1479-1489, March 2004.
- [21] B. Girod, K. Stuhlmuller, M. Link, and U. Horn, "Packet Loss Resilient Internet Video Streaming," in *proc. of SPIE Visual Communications and Image Processing*, vol. 3653 (2), pp. 833-844, San Jose, CA, January 1999.
- [22] H. Mansour, V. Krishnamurthy, and P. Nasiopoulos, "Channel Adaptive Multi-user Scalable Video Streaming with Unequal Erasure Protection," in *Intl. Workshop on Image Analysis and Multimedia Interactive Services (WIAMIS 2007)*, p. 54, June 2007.

- [23] H. Mansour, P. Nasiopoulos, and V. Krishnamurthy, "Real-Time Joint Rate and Protection Allocation for Multi-User Scalable Video Streaming," to appear in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'08)*.
- [24] H. Liu, and M. El Zarki, "Performance of H.263 Video Transmission over Wireless Channel using Hybrid ARQ," in *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 9, pp. 1775-1786, December 1997.
- [25] J. Wen, Q. dai, and Y. Jin, "Channel-Adaptive hybrid ARQ/FEC for Robust Video Transmission over 3G," in *proc. of IEEE Intl. Conf. on Multimedia and Expo (ICME'05)*, pp. 580-583, July 2005.
- [26] M. G. Martini, and M. Chiani, "Proportional Unequal Error Protection for MPEG-4 Video Transmission," *IEEE Intl. Conf. on Communication (ICC 2001)*, vol. 4, pp. 1033-1037, doi. 10.1109/ICC.2001.936799.
- [27] Y. Wang, and M. D. Srinath, "Error Resilient Video Coding with Tree Structure Motion Compensation and Data Partitioning," *12th Intl. Packet Video Workshop (PV 2002)*, Pittsburg, PY, May 2002.
- [28] Y. -K. Wang, M. M. Hannuksela, and M. Gabbouj, "Error Resilient Video Coding using Unequally Protected Key Pictures," in *proc. of Intl. workshop VLBV03*, Sept. 2003.
- [29] M. M. Ghandi, B. Barmada, E. V. Jones, and M. Ghanbari, "H.264 Layered Coded Video over Wireless Network: Channel Coding and Modulation Constraints," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 85870, pages: 1-8, doi: 10.1155/ASP/2006/85870.

- [30] S. Xiao, C. Wu, J. Du, and Y. Yang, "Reliable Transmission of H.264 Video over Wireless Network," *20th Intl. Conf. on Advanced Information Networking and Applications 2006 (AINA 2006)*, vol. 2, pages: 5, doi:10.1109/AINA.2006.285, April 2006.
- [31] K. Suehring, H.264/AVC Software Co-ordination, available at <http://iphome.hhi.de/suehring/tml/>
- [32] 3GPP doc S4-040803, Video Network Simulator and Error Masks for 3GPP Services, 3GPP TSG-SA-WG4 #33 Meeting, November 22-26, 2004.
- [33] ITU-T VCEG-M77, Common Test Conditions for RTP/IP over 3GPP/3GPP2, VCEG (SG16/Q6), Thirteenth Meeting, Austin, TX, April 2001.
- [34] Video Sequences, available at http://www.img.lx.it.pt/~discover/test_conditions.html
- [35] McCanne, and S. Floyd, Network Simulator, available at <http://www.isi.edu/nsnam/ns>
- [36] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, "Analysis of Video Transmission over Lossy Channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1012-1032, June 2000.