Scalable Video Streaming with Fountain Codes

Scalable Video Streaming with Fountain Codes

By

Sajid Nazir

Cambridge Scholars Publishing

Scalable Video Streaming with Fountain Codes

By Sajid Nazir

This book first published 2016

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-4438-7757-3
ISBN (13): 978-1-4438-7757-2

I dedicate this book to my brother, Shahid Nazir, who initiated me on the journey of image and video coding and inspired me to learn computer programming.

TABLE OF CONTENTS

FOREWORD

The area of video streaming has seen tremendous growth due to the enhanced processing power, better compression algorithms and increased bandwidth in the emerging networks. However research challenges exist to overcome effects of video packet losses and delays to provide a better user experience. Fountain codes are used in this book to address some of the challenges in this research area, and are combined in innovative ways with the different importance classes of compressed video data. This book represents a useful reference point for researchers, academics, research students, and industry developers interested in utilizing the error correction codes for ensuring better video quality.

The book is organized in five chapters. The outline of the chapters is as follows:

Chapter 1 **Introduction to Video Communications and Streaming** describes the basic techniques employed for the encoding of video data and the latest video coding standards. The error resilience features present in the latest video coding standards, Multi Description Coding (MDC), transport protocols and multimedia communication standards are also described. Some applications of video streaming are also provided.

Chapter 2 **Fountain Codes for Error Correction** covers the need for the error correction codes and demonstrates their use for protecting the video data. State-of-the-art fountain coding schemes namely LT codes, Raptor codes and Random Linear Codes are described. An application of Fountain codes is presented for DVB-H network.

Chapter 3 **Rate Adaptation Techniques for Video Streaming** highlights the importance for the video application to adapt its rate to the underlying channel bandwidth. This chapter proposes a novel adaptive scheme, called RASSA, for unicast video transmission.

Chapter 4 **Video Delivery over Wireless Relay Networks** covers the cooperative relay based multi hop wireless network to symbolize the 4G communication standard, LTE-A. The transmission of layered video data for H.264/AVC and H.264/SVC is compared over relay networks.

A multiple description coding (MDC) scheme based on the slicing and data partitioning (DP) is proposed. Expanding window-random linear codes are used to simulate the transmission of the layered descriptions for LTE-A standard.

Chapter 5 **Future of Video Streaming** points to some of the promising technological developments for video streaming.

PREFACE

Most of the latest communication standards are Internet protocol (IP) based, whereas Internet provides only a best-effort service model and the priority-based service models are only gradually being realized for real-time data. Multimedia communication over such wireless channels is difficult due to the fact that the channel conditions are generally poor and also the channel characteristics can change very rapidly. Considering the importance of the issues highlighted above, this book focuses on designing techniques to exploit different importance classes in compressed video data to design adaptive solutions to support multimedia traffic over wireless channels.

The video content is sent and received in compressed form as small packets and the video playback can start after some data has been buffered at the receiver. Thus, video streaming enables playing of video as more data is still being received and the user does not have to wait for the entire file to be downloaded.

The video streaming over IP networks has seen tremendous growth and this traffic is likely to increase as new platforms and applications evolve. The area of video communications over current heterogeneous networks is very important and challenging. Internet was never designed to work with such real-time traffic. Considering a TCP/IP network, the receipt of each packet is acknowledged to ensure the guaranteed delivery but this works against video data characteristics where it can tolerate some loss but not the delay in delivery. Mobile wireless channels and IP based communications are inherently prone to errors and packet losses which highlight the need for better error protection and recovery mechanisms.

Error resilience features and Forward Error Correction (FEC) at the application layer are often used to protect the video data against losses. One class, Fountain codes are rateless codes which can be used to potentially generate an unlimited number of encoded packets from a limited number of source packets. The decoding is possible if the number of received encoded packets at the receiver is just a little more than the source packets. The delivery and recovery of the original encoded information thus is independent of the network reliability. However, there is a small penalty to be paid for the encoding and decoding complexity.

As each portion of compressed video data does not have equal importance for the video re-construction, this characteristic can also be advantageously exploited while designing FEC solutions by providing more protection to more important portions. Thus efficient FEC solutions can be designed which take advantage of the scalable nature of current video standards and the unequal importance of video data. The amount of video data to be transmitted and the degree of protection by the rateless codes can be adaptively applied to match the channel conditions and is important for transmission to heterogeneous receivers.

Scalable video coding has come into prominence with H.264 Scalable Video Coding (SVC) standard. Scalable video means that it is possible to extract a subset of compressed video data which can be used to playback the video at a lower quality. The advantage with the use of scalable video coding is that the data gets coded as an important base layer, sufficient to create a low quality representation, and one or more enhancement layers, each of which refine the video quality. This property of video data makes it possible to use the fountain codes to provide Unequal Error Protection (UEP), that is, more protection to the more important data. UEP could also be used for other importance classes that can be realised in AVC standard through data partitioning or slices or by assigning a higher protection to more important intra (I) frame as compared to P and B frames.

In order to ensure quality of experience, FEC is the scheme proposed in this book. Fountain codes have emerged as a dominant scheme where the degree of protection does not depend on which packets are received but rather how many are received. This characteristic of fountain codes can result in many simplifications to the data transport techniques.

The video coding standards H.264/AVC and SVC together with the latest standard High Efficiency Video Coding (HEVC) provide mechanisms to match the video content to the available channel bandwidth through selecting a subset of compressed video for transmission with resulting reconstruction at a reduced quality. It is thus imperative to design FEC solutions which are adaptive to the varying wireless channel conditions, i.e., bandwidth and packet loss rate. This adaptive behaviour becomes even more important for transmission to heterogeneous receivers.

The book covers the Luby Transform (LT codes), Random Linear Codes (RLC) and Raptor coding based techniques for the mobile television broadcasting standards like Digital Video Broadcasting-Handheld (DVB-H) and DVB-T2 (Second Generation Terrestrial). A reliable unicast video communication solution based on LT codes is proposed by exploiting Unequal Error Protection (UEP) for encoded video data partitioned with the Data Partitioning (DP) and slicing feature of H264/AVC. A comparison between H.264/AVC and H.264/SVC data is given for relay collaboration strategies, namely Amplify-and-Forward (AF) and Decode-and-Forward (DF). The possible applications and advantages of each scheme are highlighted. A novel solution for Multiple Description Coding (MDC) is proposed based on DP and slicing features of H.264/AVC. The proposed solution is distinct because within each description created by slicing feature; provision of exploiting UEP is also available by utilizing the DP feature. Thus many possibilities emerge to selectively transmit the partitions of each description to exploit the multi path diversity.

The book describes how the scalable and error-resilience features of the video coding standards can be optimally combined with fountain codes to ensure quality of video under given channel conditions. The proposed solutions are extensible to the other video coding standards and error correction codes. The book aims to present simple and generalized solutions which can be extended to the latest video compression standards, such as HEVC. For example, the approach can be extended to other error-resilience features or importance classes within compressed video data.

ACKNOWLEDGEMENTS

The idea to write this book arose during my PhD when I developed a passion for video communications. I am thankful to the reviewers and publishers for enabling me to present my work as a book.

I am extremely thankful for the support and guidance provided by my PhD advisor Dr. Vladimir Stankovic, during my PhD. I am highly indebted to Dr. Dejan Vukobratovic for the guidance I received from him during his tenure at University of Strathclyde. I acknowledge his help for providing me with the exposure and understanding of fountain codes, and providing the code for optimizing probabilities. I am thankful to Dr. Kristian Nybom for providing the DVB error trace files and guidance for DVB-H and DVB-T2 systems. I acknowledge the support of Dr Samuel Cheng who provided the dynamic programming solution for the optimization problem for relay networks.

I am greatly indebted to my wife for sparing me the long hours that I spent writing the book. I am also thankful to my children, Maria, Salman and Haider, who showed a lot of understanding and maturity in keeping with the long hours of my PhD, and then in the writing of this book.

About the Author

Sajid Nazir received his BE degree in Electrical Engineering from NED University, Karachi, Pakistan in 1986 and M.Sc degree in Computer Engineering from National University of Sciences and Technology, Rawalpindi, Pakistan in 1996. He received his PhD degree from the University of Strathclyde, UK in Electrical Engineering with the thesis entitled, “Multimedia Communications over Mobile IP Wireless Networks”. Sajid has served as Research Fellow at University of Aberdeen from Jun 2012 to Mar 2015. He is currently a KTP Associate at London South Bank University, working as a Systems Engineer at Firstco Limited, London. His research interests are in adaptive video streaming, error resilience, video communications, and networking.

LIST OF ABBREVIATIONS

4G	4^{th} Generation
AF	Amplify-and-Forward
AL	Application Layer
AL-FEC	Application Layer Forward Error Correction
ARQ	Automatic Repeat Request
AVC	Advanced Video Coding
BER	Bit Error Rate
B-Frame	Bi-Directionally predicted Frame
BP	Belief Propagation
BPSK	Binary Phase Shift Keying
BS	Base Station
CIF	Common Interchange Format
CIP	Constrained Intra Prediction
DASH	Dynamic Adaptive Streaming over HTTP
DCT	Discrete Cosine Transform
DF	Decode-and-Forward
DP	Data Partitioning
DVB-H	Digital Video Broadcasting - Handheld
DVB-NGH	DVB - Next Generation Handheld
DVB-T	DVB - Terrestrial
DVB-T2	DVB-Second Generation Terrestrial
EEP	Equal Error Protection
ELP	Equal Loss Protection
ETSI	European Telecommunications Standards Institute
EW	Expanding Windows
EWF	Expanding Window Fountain Codes

EW-RLC	Expanding Window – Random Linear Codes
FEC	Forward Error Correction
FMO	Flexible Macroblock Ordering
FPS	Frame Per Second
FSMC	Finite State Markov Chain
GE	Gaussian Elimination
GF	Galois Field
GOF	Group of Frames
GOP	Group of Pictures
H.264/AVC	H.264/Advanced Video Coding
H.264/SVC	H.264/Scalable Video Coding
HARQ	Hybrid Automatic Repeat Request
HDTV	High Definition Television
HEVC	High Efficiency Video Coding
HPC	High Priority Class
HPL	High Priority Layer
HSDPA	High-Speed Packet Downlink Access
HTTP	Hypertext Transfer Protocol
IDR	Instantaneous Decoder Refresh
I-Frame	Intra Coded Frame
IP	Internet Protocol
ITU	International Telecommunication Union
LDPC	Low Density Parity Check
LPC	Low Priority Class
LPL	Low Priority Layer
LT	Luby Transform
LTE-A	Long Term Evolution- Advanced
MB	Macroblock
MBIU	Macroblock Intra Update
MBMS	Multimedia Broadcast Multicast Service

MC	Motion Compensation
MCP	Motion Compensated Prediction
MD	Multiple Description
MDC	Multiple Description Coding
ME	Motion Estimation
MPE	Multi Protocol Encapsulation
MPEG	Moving Picture Experts Group
NAL	Network Abstraction Layer
NC	Network Coding
NOW	Non-overlapping Windows
OFDMA	Orthogonal Frequency Division Multiple Access
PE	Probability of Error
PET	Priority Encoded Transmission
P-Frame	Predictively Coded Frame
PLP	Physical Layer Pipes
PLR	Packet Loss Rate
PS	Probability of Selection
PSNR	Peak Signal-to-Noise Ratio
QCIF	Quarter Common Intermediate Format
QoS	Quality of Service
RASSA	Rate Adaptive Selective Segment Assignment
RCPC	Rate Compatible Punctured Convolutional Codes
RGB	Red-Green-Blue
RLC	Random Linear Codes
RNG	Random Number Generator
RS	Reed Solomon
RTP	Real Time Transport Protocol
RTC	Raptor Code
SNR	Signal-to-Noise Ratio
SVC	Scalable Video Coding

TCP	Transmission Control Protocol
TDM	Time Division Multiplexing
UEP	Unequal Error Protection
ULP	Unequal Loss Protection
UDP	User Datagram Protocol
VCL	Video Coding Layer
VOD	Video-on-Demand
WiMAX	Worldwide Interoperability for Microwave Access
XOR	Exclusive OR
YUV	Brightness/Luminance/Chrominance Space

CHAPTER ONE

INTRODUCTION TO VIDEO COMMUNICATIONS AND STREAMING

1.1 Introduction

Video is the most important medium for communications. Video communication is unimaginable without compression due to the large size of video data. The larger video data file sizes necessitate a high degree of compression for video streaming data for its storage and transmission. Modern video coding standards employ motion-compensated prediction (MCP) between consecutive video frames and a high degree of compression. The video data also has delivery time constraints different to those of non-real time flows like file transfers. Video communication is now mostly over Internet protocol (IP) based packet networks. IP networks are best-effort and provide no guarantees of bandwidth, delay, and losses [1].

The compressed bitstream becomes highly susceptible to any data losses as even a small loss can result in losing synchronization between the sender and receiver resulting in error propagation to subsequent frames. Thus there is a need to devise schemes to offset the effect of data loss.

With video coding standards like H.264/AVC or SVC it is possible to encode the video stream so that its encoded content can be adapted to cope with the variations in bandwidth. Such adaptations are possible because of video scalability, that is, a subset of the encoded video data can be extracted which is a low quality representation of the complete data. H.264/SVC inherently provides scalability, that is, the encoded video data can be arranged in a hierarchy of importance starting at base layer and one or more enhancement layers. Similarly, H.264/AVC data can be prioritized into most important data (analogous to base layer), and other (one or more enhancement) layers of decreasing importance through data partitioning, slicing and other features. The most important encoded data is the base layer which can provide an acceptable quality. Any loss from the base layer will render the received video data useless. Each enhancement layer successively adds to the video quality, and with all the enhancement layers the best quality can be achieved. After encoding, it is possible to extract subsets of video data which can suit heterogeneous devices with different displays, memory and processing power. Although, it is possible to drop data from the enhancement layers but care must be taken as the enhancement layers also have dependencies. It can therefore be advantageous to design a drop policy taking into account the layer dependencies and to provide better loss protection for prioritised transmission to the important layers.

The rest of the Chapter is organised as follows. The video compression standards are discussed in Section 1.2. Section 1.3 covers the error correction. Multiple description coding is described in Section 1.4. IP and Network protocols are covered in Section 1.5 and 1.6. Multimedia communication standards are covered in Section 1.7. Section 1.8 and 1.9 briefly describe video streaming and some of its applications. Finally Section 1.10 concludes the chapter.

1.2 Video Compression Standards

Video is a succession of still images. An image represents light intensity in spatial coordinates most of which remains constant or varies little from one image to the next except in areas where there is motion or changes in ambient light.

The size of video data is many orders of magnitude greater than text or scalar data. Sophisticated video compression algorithms have been developed so that video data can be easily stored and communicated.

For colour images, a colour space conversion is first applied to convert the Red-Green-Blue (RGB) image into a YUV (brightness/luminance/ chrominance space) where these components can be assigned different weights depending on the human visual perception, which is strongest for brightness.

Spatial redundancy is exploited by employing signal transformation, such as, Discrete Cosine Transform (DCT) which works on adjoining pixels and removes the redundancy. In order to remove temporal redundancy, the similarity between the frames which are closer together in time can be exploited. Instead of coding each frame in isolation, the similarity between frames is exploited by, first predicting it based on a previously coded frame, and then coding the difference in this prediction. In order to reduce the difference between frames, the motion is estimated by a process termed motion estimation (ME).

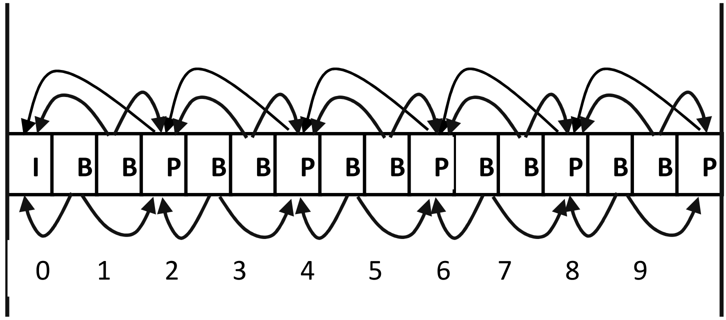

Figure 1-1: GOP structure of 16 frames.

A frame coded independently of other frames is termed as intra-frame or I-frame. Such frames typically exist at the start of video sequence or a group-of-pictures (GOP) and generally have larger sizes. A frame based on a prediction to a previous frame only is termed as a predictively-coded or P-frame. The prediction however could be done based on a previous and future frame as well, which gives even better compression, such frames are termed as bi-directionally predicted frames or B-frames. The different coded frames and their dependencies are shown in Figure 1-1. The selection of prediction dependencies between frames can have a significant effect on video streaming performance, e.g. in terms of compression efficiency and error resilience.

Video compression is normally lossy, that is after compression of the video data, its reconstruction back to the original may not be exact. However, such loss is acceptable and its nature may be governed by a given application. The latest video compression standards achieve compression by applying the same basic principles as described above. The colour space redundancy is exploited by a colour space conversion. Then the temporal redundancy is removed by applying Motion Compensation (MC)-prediction, and the spatial redundancy is removed by applying the DCT. The resulting DCT coefficients are subsequently processed to generate the compressed bit stream. However, as this compressed bit stream is highly susceptible to quality degradation even by bit losses hence the modern video coding standards employ quite advanced techniques to contain the adverse effect of such losses.

1.2.1 H.264/Advanced Video Coding (AVC)

H.264/AVC [2] is a recent state of the art video compression standard. Similar to the prior video coding standards, it is based on Motion Compensated Prediction (MCP), which requires maintaining synchronisation between the encoding and decoding operations.

H.264/AVC has achieved significant improvement in rate-distortion efficiency relative to existing standards [3]. H.264/AVC has been adopted by various application standards and is increasingly used in most video applications.

1.2.2 H.264/Scalable Video Coding (SVC)

SVC is an extension of H.264/AVC standard and adds scalability features to it. A scalable video encoder compresses a raw video sequence into multiple layers. Scalable video means that it is possible to extract a subset of compressed video data which can be used to playback the video at a lower quality. Scalability makes it possible to encode the video only once in different layers or classes, some of which (the least significant ones) can be removed to fit a bandwidth, processing power or display characteristics of the target device. Modern video transmission is over mobile heterogeneous devices and the video resolution and quality must adapt to the characteristics of the device. SVC is a highly attractive solution to the problems posed by the characteristics of modern video transmission systems [4].

The most important layer is the base layer, without which decoding fails but which provides a coarse quality at a much reduced data rate. Additional compressed layers are enhancement layers that provide additional quality to the received video stream. Enhancement layers can be decoded only in conjunction with the base layer. On a Quality of Service (QoS)-enabled IP network it would even be possible to send the base layer with a higher priority than the other layers.

The networks which support prioritization can make use of scalability by assigning a higher priority to the layers according to their importance.

The base layer could thus be transported with the highest priority. However, Internet does not provide any such prioritization and all packets are equally likely to be lost. In such networks, the scalability alone does not bring any advantage for video transport; however, channel coding can be used to make the base layer more tolerant to errors [6]. The layers can be combined to adapt to different frame rates, spatial resolutions, or quality, of the video content, giving rise to temporal, spatial, and quality scalability.

1.2.3 High Efficiency Video Coding (HEVC)

HEVC is the latest video coding standard which provides 50% reduction in bandwidth for the same quality as compared to H.264 standard. The high coding efficiency will enable video communication over low-power and bandwidth-constrained mobile devices.

HEVC is block based and utilises many of the coding concepts from H.264/AVC standard in innovative ways to achieve a high degree of compression.

1.3 Error Correction

The Internet is prone to packet losses which can result from packet drops due to congestion or bit errors. Such losses can degrade the user experience and therefore measures must be taken to recover from losses and provide a graceful degradation in video quality.

There are different means available to protect the video data and to increase the quality of the user experience at the receiver. This could be accomplished by providing:

- Error resilience
- Error protection
- Error concealment

The data transmission based on IP is in the form of packets where each packet may be independently routed. Some of the packets may be lost en-route e.g., due to buffer overflows or may be delayed beyond their display deadline. The traditional solutions for data delivery with re-transmission of lost packets as in Transmission control protocol (TCP) does not work well for real-time video transmission because of the tight delay constraints of each packet. As TCP is based on an acknowledgement of packets from the receiver so it will give poor performance for transmission over large distance. One of the solutions is to use channel coding techniques which could recover the original data despite losses. The solutions based on schemes like Reed Solomon (RS) codes are inflexible because the code rate has to be fixed in advance. Moreover, the encoding and decoding operations are quite complex especially for large Galois Field. For such codes the error characteristics of the channel must be known in advance in order to adjust the code rate to it. This solution does not extend well to multiple receivers as then only a worst-case erasure channel can be assumed for all receivers.

1.3.1 Error Resilience Features

The latest video standards are aimed at packet-based networks and thus cope mainly with packet losses instead of bit errors [2]. There are many error-resilience features built into the latest coding standard aimed at curtailing the video degradation at the receiver if some information is lost during transmission.

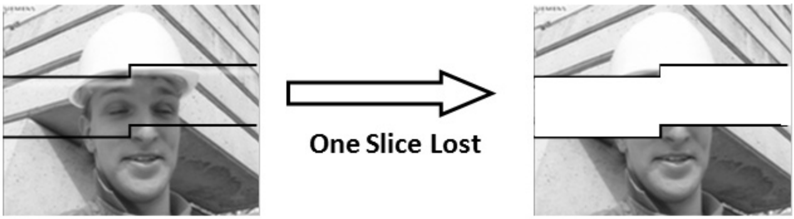

Figure 1-2: Effect of a lost slice is restricted to smaller area.

1.3.1.1 Slice Structured Coding

A frame may be divided into suitably sized portions or slices to provide protection against degrading the whole frame during transmission. Each picture may be split into one or several slices as shown in Figure 1-2. The macroblocks will be organized in slices and this has the desirable effect of making transmitted packets smaller; making the overall recovered video frames better tolerant of errors. Slice structured coding also introduces slice headers to act as resynchronisation points to localize the errors and prevent error propagation to subsequent frames. Each slice can be correctly decoded without the use of data from other slices provided in the same frame. Each slice is encapsulated in a separate packet by H264/AVC encoder. Slices are self-contained, thus, if a coded slice is available to the decoder, all the MBs in that slice can be decoded.

1.3.1.2 Data Partitioning

DP is a feature available in the extended profile which supports the partitioning of a frame/slice in up to three partitions, based on the importance of the encoded video syntax elements for video reconstruction [92], [93].

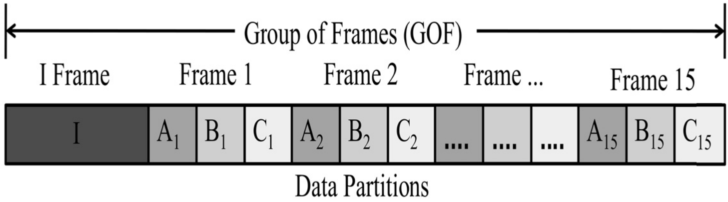

Figure 1-3: Relative position of DP A, B and C in a GOF of 16 frames.

A GOF with each frame split in three partitions is shown in Figure 1-3. DP A contains the most important data comprising slice header, quantization parameters, and motion vectors. DP B contains the intra-coded macroblocks (MB) residual data, and DP C contains inter-coded MB residual data [92], [89] © 2012 IEEE, [94] © 2013 IEEE. The decoding of DP A is always independent of DP B and C [91], [90] © 2011 IEEE. At the decoder, if the Type B or Type C partition is missing, the header information in the Type A partition may be used in order to aid in error concealment.

However, if DP A is lost the remaining partitions cannot be utilized. The decoding of DP B is possible without DP C, but not other way around. To make DP B independent of DP C, Constrained Intra Prediction (CIP) parameter in H.264/AVC encoder must be set [51], [91], [92], [93], [58], [90] © 2011 IEEE, [89] © 2012 IEEE.

1.3.2 Fountain Codes

Fountain codes are rateless codes which can potentially generate an unlimited number of encoded packets from a limited number of source packets. Each encoded packet is based on a combination of the source packets according to some distribution. It is thus not necessary as to which packets are received but rather that they are received in sufficient quantity. The decoding is possible if the number of received encoded packets at the receiver is just a little more than the source packets. The encoding thus eliminates the effect of independent losses at different receivers, and also there is little requirement to send feedback to the sender.

For the application of Fountain codes to video data, a Group of pictures (GOP) could be treated as a source block. The encoded packets for a particular GOP are generated based on the available bandwidth. The solutions based on Fountain codes adapt well to the varying bandwidth as the receiver or sender can terminate reception/transmission depending on the available bandwidth. Such codes are ideally suited for use in multicast scenarios because there is no requirement to target a particular receiver. It is thus imperative to design Forward Error Correction (FEC) solutions which are adaptive to the varying wireless channel conditions, i.e., the bandwidth and the packet loss rates. This adaptive behaviour becomes even more important for heterogeneous receivers.

The FEC protection could be provided at different layers of the network protocol stack. However, providing the FEC solution at the application layer (AL) makes it more flexible. It can also be easily implemented in software. In addition, a video application knows best how to handle each packet and therefore it is better to leave such decisions to be taken at the application layer. Error resilience features and FEC at the AL are thus often used to protect the data against losses.

The Third Generation Partnership Project (3GPP) recommends the use of FEC for Multimedia Broadcast and Multicast Services (MBMS) and, more specifically, adopts the use of Raptor FEC code in the AL. Digital Video Broadcasting-Handheld (DVB-H) uses Raptor codes and similar AL-FEC schemes are proposed for DVB-Next Generation Handheld (NGH) and Long Term Evolution-Advanced (LTE-A) at the AL. Due to backward compatibility with DVB-Terrestrial (T), DVB-H uses the same first generation FEC schemes with convolutional coding and Reed-Solomon (RS) block coding in addition to AL Raptor codes.

Other than Raptor codes, a class of rateless codes that is gaining increased popularity for applications in wireless broadcast/cellular networks are Random Linear Codes (RLC) [68] © 2011 IEEE. With multi-hop and cooperative communications becoming popular in emerging wireless