OPTIMIZATION OF RATELESS CODED SYSTEMS FOR WIRELESS MULTIMEDIA MULTICAST

by

Yu Cao

A thesis submitted to the
Department of Electrical and Computer Engineering
in conformity with the requirements
for the degree of Doctor of Philosophy

Queen's University
Kingston, Ontario, Canada
June 2011

Abstract

Rateless codes, also known as fountain codes, are a class of erasure error-control codes that are particularly well suited for broadcast/ multicast systems. Raptor codes, as a particularly successful implementation of digital fountain codes, have been used as the application-layer forward error correction (FEC) codes in the third generation partnership program (3GPP) Multimedia Broadcast and Multicast Services (MBMS) standard. However, the application of rateless codes to wireless multimedia broadcast/multicast communications has yet to overcome two major challenges: first, wireless multimedia communications usually has stringent delay requirements. In addition, multimedia multicast has to overcome heterogeneity. To meet these challenges, we propose a rateless code design that takes the layered nature of source traffic as well as the varying quality of transmission channels into account. A convex optimization framework for the application of unequal error protection (UEP) rateless codes to synchronous and asynchronous multimedia multicast to heterogeneous users is proposed.

A second thread of the thesis addresses the noisy, bursty and time- varying nature of wireless communication channels that challenge the assumption of erasure channels often used for the wired internet. In order to meet this challenge, the optimal combination of

application-layer rateless code and physical layer FEC code rates in time-varying fading channels is investigated. The performance of rateless codes in hybrid error-erasure channels with memory is then studied, and a cross-layer decoding method is proposed to improve decoding performance and complexity.

Acknowledgments

I am grateful to my supervisor and thesis advisor Prof. Steven D. Blostein for his guidance, technical advice, encouragement, and financial support during my Ph.D. study.

I would like to thank Prof. Geoffrey Chan for his critical and constructive feedback on part of the thesis project. My sincere thanks go to Dr. Wei Sheng for her discussions and her contributions to Lemma 4.1.

I would like to thank all my thesis committee members, Prof. Amir H. Banihashemi from Carleton University, Prof. Fady Alajaji, Prof Saeed Gazor and Prof. Michael Korenberg for their comments and suggestions which leads to the clarity of this thesis.

I would like to thank my family for their continuous support and understanding. In particular my wife Jia Ma for her support and my daughter Vanessa for the happiness she brings to my family. I would like to thank my parents for their care and love since I was born. I would also like to thank my parents-in-law for their support, especially for their help in looking after my daughter.

I thank all my labmates, past and present, for their friendship and all the good time we had together. They include, but not limit to, Lucien Benacem, Minhua Ding, Xin Guan, Luis Gurrieri, Leon Lee, Guangping Li, Patrick Li, Ye Li, Jason Liang, Hani Mehrpouyan,

Hassan Abou Saleh, Wei Sheng, Constantin Siriteanu, Edmund Tam, Neng Wang, Jinsong Wu, and Yi Zheng. I want to thank Minhua Ding for her helpful discussions, reviews and encouragements.

This work is supported in part by the Natural Sciences Engineering and Research Council of Canada Grant STPSC 356826-07, and in part by Discovery Grant 41731. Their financial support are greatly appreciated.

Contents

List of Tables

List of Figures

Acronyms

3GPP	Third Generation Partnership Program
ACK	Acknowledgement
AMC	Adaptive Modulation and Coding
ARQ	Automatic Repeat Request
AWGN	Additive White Gaussian Noise
BEC	Binary Erasure Channel
BER	Bit Error Rate
BIAWGN	Binary Input Additive White Gaussian Noise
BL	Base Layer
BP	Belief Propagation
BPSK	Binary Phase Shift Keying
BSC	Binary Symmetric Channel
BSCE	Binary Symmetric Channel with Erasures
CLD	Cross-layer Design
CLDS	Cross-layer Design with Side Information
CON	Conventional

CRC	Cyclic Redundancy Checks
CSI	Channel State Information
DF	Digital Fountain
DVB-H	Digital Video Broadcast for Mobile Handheld
EEP	Equal Error Protection
EL	Enhancement Layer
EWF	Expanding Window Fountain
FEC	Forward Error Correction
FER	Frame Error Rate
GE	Gilbert-Elliott
GOF	Group of Frames
HEEP	Hybrid Erasure-error Protocol
KKT	Karush-Kuhn-Tucker
LAN	Local Area Network
LDPC	Low Density Parity Check
LIB	Less Important Bits
LLR	Log-Likelihood-Ratio
LT	Luby Transform
MBMS	Multimedia Broadcast/Multicast Services
MDS	Maximum Distance Seperable
MIB	More Important Bits
ML	Maximum-Likelihood

NCLD	Non Cross-layer Design
PDF	Probability Density Function
PER	Packet Error Rate
PET	Priority Encoding Transmission
PSNR	Peak Signal-to-Noise Ratio
QoS	Quality of Service
RS	Reed-Solomon
RSD	Robust Soliton Distribution
SNR	Signal-to-Noise Ratio
SVC	Scalable Video Coding
UEP	Unequal Error Protection

List of Important Symbols

$(\cdot)^T$	Matrix or vector transpose
L	Total number of layers
J	Total number of user classes
K	Total number of information symbols/bits
ρ_j	Selection probability of Layer j
S_l	Number of information symbols for Layer l
ε	Transmission overhead
γ_j	Target QoS (PSNR) threshold of Class j users
g_j	Number of layers required to meet PSNR threshold γ_j
P_j	Probability threshold for class j user to meet target QoS value
δ_j	Reception capability of Class j users
q_l	PSNR value when Layers 1 to l are decoded
w_j	Weighting coefficient for Class j users
σ_j	Erasure rate of Class j users
τ_j	Outage probability threshold of Class j users
ω	Rateless code overhead for asynchronous multicast systems

α_j	Weighting coefficients of the cost function
K_l	Number of information symbols in Layer l in the PET packetization scheme
h_j	Number of layers required to meet Class j users' QoS demand
l_j	$h_j - h_{j-1}$, number of layers in Chunk j
U_j	number of symbols in Chuck j
M_0	Maximum number of transmitted packets
O_j	Outage probability of Class j users
γ	Signal-to-noise Ratio
ε_p	Packet overhead
ε_r	Average application-layer raptor code overhead
TP	Throughput
f_dT	Normalized doppler frequency
R_c	Physical layer code rate
R_m	Physical layer modulation rate
n	Index of physical layer transmission mode
$a_n, b_n, g_n, \gamma_{bn}$	Transmission mode dependent parameters for evaluating packet error rate
δ	Packet drop rate in CON scheme
λ	Packet drop rate in CLD scheme
ε	Packet corruption level
μ	Channel memory
p	Bit error probability of an unerased packet

Chapter 1

Introduction

Multimedia multicast over wireless channels has gained enormous attention due to the increasing demand for multimedia content on mobile devices. However, delivering rich multimedia content to heterogeneous users is still a challenging task. A fundamental question for reliable multimedia multicast is how to efficiently correct erased packets for multiple heterogeneous users with varying loss rates. The recently proposed fountain codes [3] provide an efficient solution to this problem.

Fountain codes are a class of rateless codes that are originally designed for binary erasure channels (BEC). Erasure channels are channels where data are either transmitted completely reliably or else erased. Besides theoretical importance, erasure channels have gained significant attention in modeling practical communication scenarios. For example, on the internet, data is transmitted in small packets. For most packets, reliability is guaranteed by appropriate protocols. The correctness of the data in a received packet is verified by a built-in check mechanism within each packet, such as checksums and cyclic redundancy checks (CRCs) [4]. Packets that are lost, corrupted, delayed or do not satisfy certain

quality of service (QoS) requirements can be discarded (erased). In wireless communications, transmitted data are subject to noise and fading. Error correcting codes are typically selected in the physical layer to correct most errors within each packet. At the packet level, the channels can also be similarly modeled as (packet) erasure channels when packets with uncorrected errors are declared as erased.

The simplest way to deal with erased packets is to retransmit them. In this method, the receiver needs to feedback the indices identifying the missing packets to the transmitter. Automatic repeat request (ARQ) is one of these protocols [4]. However, there are some drawbacks in using re-transmission schemes. First, the channel capacity of an erasure channel has been shown to be the same with and without feedback [5]. This suggests that the added complexity introduced by a feedback channel may not be necessary. Second, while re-transmission schemes may work well for unicast protocols, they are very difficult to implement in broadcast/multicast scenarios. Consider the case where a server is distributing the same information to multiple users. Since each user has different channel conditions resulting in different missing packets, re-transmission according to requests from each individual user can quickly overwhelm the server.

Another way to deal with erased packets is to deploy traditional block codes on the source packets before transmission to efficiently correct missing packets at the receiver. By efficiently, we mean codes that perform at near Shannon limits while maintaining low complexity as is suitable for practical deployment. The most widely used classical erasure codes are Reed-Solomon (RS) codes [4] [6]. RS codes have the desirable property that they are maximum distance separable (MDS) [4]. A MDS code is a code whereby

the original K source symbols can be recovered by any K successfully received symbols. However, the encoding and decoding complexities limit practical RS codes to small code lengths. In addition, as with other traditional block codes, the rate of RS codes has to be predetermined based on the erasure probability. In heterogeneous multicast scenarios, different receivers may have different erasure probabilities. In such scenarios, the transmitter would be forced to select the rate according to the worst user, which is very inefficient. One way to overcome the inefficiency is to regenerate and transmit a new block of RS codes when some receiver fails to decode the source block. Due to the lack of an "on-the-fly" generation feature of RS codes, the receiver may receive plenty of duplicate packets before receiving useful information to finish decoding. Other researchers propose to use hybrid approaches which combine traditional erasure codes with ARQ schemes [7]. Although these approaches reduce re-transmission rate they still do not perform well enough for multicast scenarios.

A class of recently developed forward error corrections codes, named fountain codes [3], are ideal solutions to the above problems. A well implemented fountain code has the following properties: (i) fountain codes have near optimal performance for erasure channels, i.e., from only slightly more than K successfully received encoding symbols, the receiver is able to recover the original K information symbols with high probability; (ii) fountain codes can be universal, meaning that the same fountain code can achieve near optimal performance regardless of the erasure probability; and (iii) fountain codes are effectively rateless. The source can generate a potentially infinite number of symbols and

only stops transmission when it receives an acknowledgement (ACK) from the receiver indicating successful decoding. Therefore, the rate of a fountain code is determined at the receiver on-the-fly rather than predetermined during encoding.

The above properties make fountain codes self-adaptive to channel conditions without requiring feedback. The concept of fountain codes is first proposed in [8], where the approach is named a Digital Fountain (DF). The name comes from the following analogy. The encoder is the "fountain", encoded symbols/packets are water drops; the "fountain" creates potentially an infinite number of water drops during encoding. The decoder is the "bucket", which has capacity equal to the number of information symbols. The "bucket" collects water drops sprayed from the fountain. When the "bucket" is full, it is able to recover the original information independently of which water drops have been collected. The first practical implementation of fountain codes is the Luby Transform (LT) code [9], which has a very simple linear structure. In LT codes, a degree d is first chosen according to some degree distribution, then d binary bits are randomly chosen among K source bits and are linearly combined in the binary field to produce an output symbol. The decoding of LT codes is similar to belief propagation (BP) decoding of low density parity check (LDPC) codes over erasure channels. With carefully designed degree distributions, LT codes are able to perform nearly as well as an ideal fountain code with a decoding complexity proportional to $K \ln K$. Later on, Shokrollahi [10] [11] extended the idea of LT codes to raptor codes, which have even lower decoding complexity. The basic idea of raptor codes is to pre-code the source bits with a block code before the inner LT code.

LT and raptor codes have been found to have many applications in internet and wireless communications, such as peer-to-peer communications, parallel downloading [12] and multicasting [13]. For example, in 3GPP Multimedia Broadcast/Mutlicast Services (MBMS), raptor codes have been chosen as the forward error correction (FEC) code in the application layer for file downloading services [14]. This is due to the great attributes of raptor codes: linear encoding complexity, linear decoding complexity and nearly optimal performance independent of channel conditions.

1.1 Motivation and Thesis Overview

There are generally two types of wireless multicast services: file downloading and multimedia streaming. File download delivery services are usually associated with transmitting a chunk of data to a variable number of users. File download services require high accuracy but can tolerate long delays. On the other hand, multimedia streaming services are real time applications. For users with a limited amount of buffering, the service is usually delay sensitive but can tolerate a small amount of errors.

While rateless codes have been shown to be very promising for transmitting large chunks of data, the application of rateless codes to wireless multimedia multicast applications is still a challenging task. The two major challenges for wireless multimedia multicast are user heterogeneities and time varying wireless channels. In this thesis, we mainly focus on addressing these two major challenges in the application of rateless codes to wireless multimedia multicast.

The first challenge in the application of rateless codes for wireless multimedia multicast is user heterogeneity. Although rateless codes can provide extra flexibility in code design as well as better efficiency compared to fixed-rate codes for multimedia multicast applications, the original design of raptor codes, despite being highly efficient for broadcasting bulk data, has very poor progressive decoding performance. On the other hand, multimedia content often has a scalable structure in which certain source bits have higher priorities than others. Therefore, an efficient fountain code designed for multimedia streaming applications should provide unequal error protection (UEP) of different source symbols. Code optimization for UEP has been proposed for fixed-rate FEC codes in [15] [16] [17]. However, these approaches do not consider guaranteed QoS for heterogeneous users and do not take rateless code performance into account. Recently, UEP rateless code designs have been proposed in [18] [19]. However, these involve complex encoding and decoding and use computationally complex exhaustive search to optimize configuration parameters.

The second challenge of the application of raptor codes in multimedia multicast is the time-varying wireless channels. They are generally not efficiently modeled as erasure channels. On the other hand, most applications of rateless codes to date assume a perfect erasure channel [11] [20], where received fountain encoded packets are either received error free or completely lost. It has been shown in [11] that raptor codes are "universally" optimal for erasure channels, meaning that the same code parameters can achieve optimal performance regardless of the channel erasure probability. Despite all the advantages of fountain codes, there are some difficulties involved in applying fountain codes to wireless broadcast/multicast channels. In wireless communications, transmission data is subject to noise

and fading, and errors can also be bursty. Powerful channel codes, such as turbo codes [20], are adopted in the physical layer to correct most errors within each packet. In this case, the channel at the packet level can be modelled as a (packet) erasure channel when packets with uncorrected errors are declared as erased.

As system simulation results of a raptor-coded cellular broadcast system in [20] and [21] suggest, an optimal balance between physical and application-layer code rates exists. In many cases, a higher packet error rate (PER) that is corrected by the application-layer raptor code can be more efficient than traditional designs. However, the conclusions provided in [20] are from system simulations, and do not provide sufficient insight and quantification of the optimal rate combination.

There are also approaches that study the performance of raptor codes over additive white Gaussian noise (AWGN) or fading channels using soft decoding [22] [23]. It has also been shown in [22] that raptor codes no longer posses the property of universal optimality over those noisy channels. In addition, performing soft decision decoding increases the decoding complexity significantly compared to that of erasure decoding.

In this thesis, we address the above two major challenges from a few different aspects that have not been investigated in the literature. To address the first challenge, we propose a convex optimization framework for UEP raptor code design for synchronous and asynchronous multimedia multicasting to heterogeneous users. As existing simulation results suggest (e.g. [20] [24]), a non cross-layer approach to the design of rateless codes can result in low system efficiency in noisy channels with low SNR even though raptor codes have high efficiency at the application layer. Therefore, to address the second challenge,

we propose two cross-layer approaches for rateless coded wireless communication systems with memory to improve the overall system efficiency. In the first cross-layer approach, we propose to use traditional erasure decoding of an application-layer raptor code to maintain low decoding complexity. However, we jointly optimize the physical-layer code rate and application-layer raptor code rate to maximize system throughput in both slow and fast fading channels. In the second cross-layer approach, a cross-layer hybrid erasure-error decoding scheme is proposed, which utilizes soft information from physical-layer decoder as well as side information from the build-in CRCs. The receivers are flexible in adopting the proposed hybrid erasure-soft decoding scheme to achieve desirable performance and complexity. The rest of the thesis is organized as follows:

In Chapter 2, background on rateless codes and their performances are first reviewed. The fundamentals behind the encoder and decoder design of LT codes and raptor codes are described. The performance of LT codes and raptor codes in both erasure channels and noisy channels are also simulated. The systematic version of raptor codes, that has been standardized is also introduced.

In Chapter 3, an optimization framework for UEP rateless code design for multimedia multicast to heterogeneous users with different QoS requirements is presented. Optimization problems with objectives that either provide guaranteed QoS or provide best-effort QoS for heterogeneous users are formulated. A random interleaved rateless encoder design is proposed. Unlike previous designs, existing standardized raptor codes can be directly applied to this design without requiring re-optimization. In solving the optimization problem for guaranteed QoS, the original problem is transformed to a convex optimization problem,

and then optimal configuration parameters of the proposed UEP rateless code are obtained.

In Chapter 4, a different UEP rateless code optimization problem that focuses on asynchronous multimedia multicast is addressed. A similar optimization problem with a different UEP rateless code design was first formulated in [25] and [26]. The problem is solved using an iterative search method in [25] and [26] which has high complexity. In Chapter 4, we show that under certain assumptions and under the relaxation of integer constraints, this problem can be formulated in terms of convex optimization. An analytical solution is then found for the asynchronous multicast optimization problem without outage constraints [25]. For the more general formulation with outage constraints [26], an analytical solution is found for the special case where there are two user classes. Numerical solutions using convex optimization software [27] [28] are required to solve the more general cases.

In Chapter 5, the optimal rate combination of the application-layer rateless codes and physical-layer codes in wireless fading channels is investigated. Both slow and fast fading channel conditions are considered. The optimal physical-layer modulation and code rate pair is analyzed in order to maximize overall system throughput. For slow fading systems, cross-layer adaptive modulation and coding design is also proposed to maximize system throughput. The performance of the proposed cross-layer design is then compared to traditional designs via numerical examples.

In Chapter 6, a hybrid erasure-error decoder is proposed to obtain desirable decoder complexity and performance. The performance of raptor codes in hybrid error-erasure channels and Gilbert-Elliott channels are first presented, which does not appear in the existing literature. A hybrid erasure-error decoder is then proposed as a cross-layer approach

for the decoding of raptor codes. A hierarchical Markov model is proposed for modeling the cross-layer hybrid error-erasure channels with memory. The performance of raptor codes under both cross-layer and non cross-layer decoding schemes are compared using the proposed channel model.

Chapter 7 concludes this thesis and suggests future work.

1.2 Thesis Contributions

The primary contributions of this thesis are briefly summarized as follows.

- A random interleaved raptor code with unequal error protection (UEP) properties for multimedia multicast is proposed and an optimization framework of UEP rateless code design for multimedia multicasting to heterogeneous users is presented. The problem of minimizing the transmission overhead while providing heterogeneous users different levels of QoS guarantees is solved. Numerical results using and-or tree analysis are presented that demonstrate that the proposed random interleaved UEP rateless code outperforms non-optimized rateless codes as well as recently proposed UEP rateless codes that use the same LT degree distributions. The guaranteed-QoS problem using standardized raptor codes is transformed to a convex optimization problem. The convexity of the problem is proven and a simplified method to solve the convex optimization problem analytically is proposed. The optimization problem of providing best-effort QoS under a fixed transmission rate employing the proposed UEP rateless code is also formulated. Numerical results show that with the ability to

adopt superior standardized raptor codes, the proposed UEP rateless code can provide a significant additional gain compared to existing rateless code designs for both the guaranteed QoS and best-effort QoS problems.

- In optimizing unequal error protection (UEP) rateless codes for asynchronous multimedia multicast to heterogeneous users, we prove that under certain nonrestrictive assumptions and with the relaxation of integer constraints, this problem can be transformed into a convex optimization problem. An analytical solution for the asynchronous multicast optimization problem without outage constraints is then found. An analytical solution for the more general formulation with outage constraints is found for the case of two user classes. For the general problem with outage constraints and more than two user classes, standard convex optimization software is used to solve the problem. Numerical examples for the cases of two-user classes and four-user classes are presented, which illustrate the new techniques.

- The system throughput of a two-layer rateless coded system applied to Rayleigh fading channels is determined as a function of channel SNR and the choice of physical-layer modulation scheme and code rate. The optimal choices of physical-layer modulation and code rate that maximize overall system throughput in both slow and fast fading channels are computed. A cross-layer adaptive modulation and coding (AMC) scheme that maximizes overall system throughput for slow fading channels is proposed. Numerical results show that the proposed cross-layer adaptive and non-adaptive schemes outperform traditional non cross-layer adaptive and non-adaptive

schemes. Insight is also provided on the balance between packet error correction and error correction via application-layer erasure codes.

- The performance of raptor codes over binary symmetric channels with erasures (BSCE) is analyzed and simulated. The property that the average overhead of raptor codes over BSCE is the same as that of corresponding binary symmetric channels (BSC) is proven. The performance of fixed rate raptor codes over Gilbert-Elliott (GE) channels with and without channel state information (CSI) at the receiver is simulated. Sensitivity to different channel parameters is also investigated.

- A hybrid erasure-soft decoding algorithm as a cross-layer protocol for decoding application-layer raptor codes is proposed. A hierarchical Markov model is applied to model the correlated fading channel that the raptor codes experience. The performance and throughput of different protocols using the hierarchical Markov model are compared, and the advantage of the cross-layer protocol employing the proposed hybrid erasure-soft decoder is quantified.

Chapter 2

Background

In this chapter, rateless codes and their performance in erasure channels are first reviewed. The performances of Luby Transform (LT) and raptor codes in noisy channels are then described. After that, application of rateless codes to wireless broadcast/multicast channels are highlighted. Finally, standardized raptor codes are introduced.

2.1 LT and raptor codes in erasure channels

LT codes, proposed by Luby [9], are the first practical realization of fountain codes. LT codes encode K information symbols $(x_1, x_2, ..., x_K)$ into a potentially infinite number of output symbols $(z_1, z_2, z_3, ...)$. The input information symbol can be a one-bit binary symbol, a general l-bit symbol, or a data packet. The encoding process is done as follows:

1. A degree d is chosen by sampling from a given distribution D, which is called the "degree distribution". Let $(\Omega_1, \Omega_2, ..., \Omega_K)$ be a distribution on $\{1, 2, ..., K\}$ such that Ω_i represents the probability that value i is chosen. The degree distribution D is

denoted by the polynomial $\Omega(x) = \Sigma_{i=1}^{K} \Omega(i) x^i$.

2. d input symbols are chosen uniformly at random from the K input symbols. The value of the output symbol is the exclusive-or (XOR) of the d chosen input symbols, i.e., $z = \Sigma_{i=1}^{K} v_i x_i$, where the coefficients $v_i \in \{0,1\}$ for $i = 1,2,...,K$, $\Sigma_{i=1}^{K} v_i = d$ and modulo 2 bit-wise sum is used.

The assumption is made that the receiver knows the coefficients v_i. In practice, this can be achieved in many ways. For example, the degree and the list of neighbors can be included in a packet header, or a key may be associated with each encoded symbol and both encoder and decoder use the same function of the key to compute the degree and set of neighbors. The encoding process of LT codes can be represented by the Tanner graph shown in Fig. 2.1. In the graph, input variable nodes and output variable nodes represent information symbols and encoded output symbols, respectively. The check nodes represent the XOR calculation. A direct way to decode a LT code in erasure channels is to use Gaussian

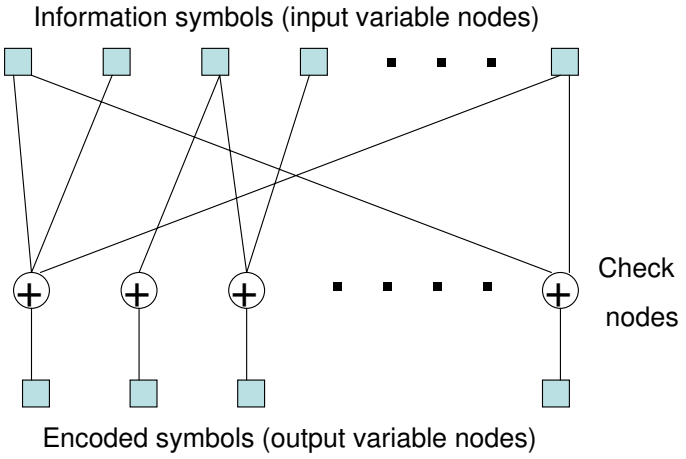

Figure 2.1. Shown is the Tanner graph of a LT code.

elimination to solve the set of linear equations. However, Gaussian elimination generally has computational complexity of $O(K^3)$ [29], which is very high. A practical decoding scheme is to use the iterative belief propagation (BP) algorithm, in a very similar manner to the BP decoding of low density parity check (LDPC) codes. The BP decoding is much simpler in erasure channels than in noisy channels since all the symbols received are either completely correct or completely uncertain. The decoding is summarized briefly as follows:

1. Find an encoded symbol z_n that is only connected to one information symbol x_i. Then decode the information symbol, i.e., determine x_i from z_n. If no such encoded symbol exists, the decoding process stops.

2. Add the value of the decoded information symbol x_i to all the other encoded symbols that are connected to x_i. Then remove all edges that are connected to x_i.

3. Repeat Steps 1 and 2 until all information symbols have been decoded successfully.

The design of the degree distribution plays a central role in the performance and complexity of LT codes. The most crucial value is the average degree, because both encoding and decoding complexity scale linearly with the total number of degrees. However, to ensure that every information symbol is covered with high probability, the total degree of the output symbols must be at least of order $K \ln K$. Therefore, $\ln K$ serves as a lower bound for the average degree of an ideal LT code. In addition, there should be at least one encoded symbols with degree 1 to initiate the decoding process. Based on these requirements, a

robust soliton distribution (RSD) has been proposed by Luby [9]. The RSD μ is given by

$$\mu(d) = \frac{\rho(d) + \tau(d)}{\beta}, \quad d = 1, 2, ..., K, \tag{2.1}$$

where $\beta = \Sigma_{d=1}^{K}(\rho(d) + \tau(d))$, $\rho(d)$ is the ideal soliton distribution,

$$\begin{cases} \rho(1) = 1/K \\ \rho(d) = 1/d(d-1), \quad d = 2, 3, ..., K \end{cases} \tag{2.2}$$

and $\tau(d)$ is defined as

$$\tau(d) = \begin{cases} R/(dK) & \text{for} \quad d = 1, 2, ..., \lceil K/R \rceil - 1 \\ R\, ln(R/\delta)/K & \text{for} \quad d = \lceil K/R \rceil \\ 0 & \text{for} \quad d > \lceil K/R \rceil, \end{cases} \tag{2.3}$$

where $R = c\ \ln(K/\delta)\sqrt{K}$ is the average number of degree one encoded symbols, and c and δ are parameters. The RSD in fact achieves the average degree lower bound of an ideal LT code, with an average degree that scales linearly with $\ln K$, and ensures that the receiver can recover the original K information symbols with probability $1 - \delta$ when βK output symbols are successfully received [9].

In practice, a constant average degree which results in linear decoding complexity is desirable. Shokrollahi [10] fulfilled this requirement by extending the idea of LT codes to raptor codes. Raptor codes are simply constructed by concatenating a block code with a weakened LT code. The weakened LT codes have a constant average degree $\bar{d}$. Using a weakened LT code, the fraction of information symbols that are not covered by any K output symbols are approximately $e^{-\bar{d}}$. Hence, if the pre-code is able to correct this fraction

of erasures, all the K information symbols will be recovered. The drawback of raptor codes is that they are bounded away from the capacity when the inner LT codes and the precode are decoded separately, with a gap equal to the rate of the pre-code. The design of the pre-code and the parameters of raptor codes are presented in detail in [11]. One of the preferred pre-codes are high rate LDPC codes, due to their sparse Tanner graphs, resulting in low decoding complexity using the BP algorithm. We next illustrate the raptor code from [11] for simulation and analysis. The pre-code of this raptor code is a left regular and right Poisson LDPC code with rate 0.95. The variable nodes of this LDPC code have constant degree 4 and check nodes that are connected to each variable node are chosen uniformly at random. The Tanner graph of the LDPC pre-code is shown in Fig. 2.2. The inner LT codes use the degree distribution,

$$\Omega_r(x) = 0.007969x + 0.493570x^2 + 0.1666220x^3 + 0.072646x^4 + 0.082558x^5$$
$$+0.056058x^8 + 0.037229x^9 + 0.055590x^{19} + 0.025023x^{65} + 0.003135x^{66}. \quad (2.4)$$

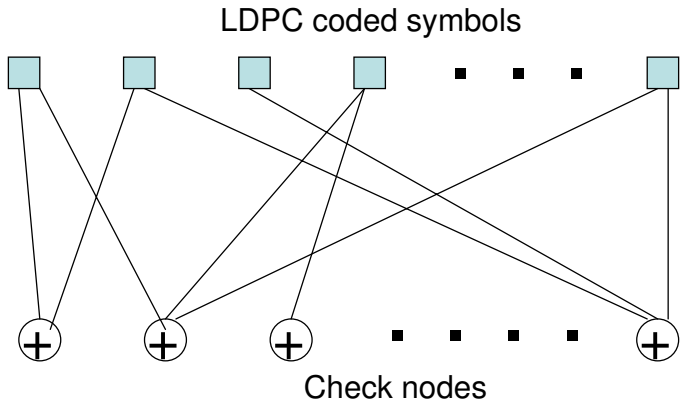

Figure 2.2. Tanner graph of the LDPC code.

The average number of output degrees per information node of the above weakened LT codes is independent of code dimension K, and therefore has the advantage of low decoding complexity compared to RSD-based LT codes. The performance of the inner weakened LT codes with a degree distribution given by (2.4) and the performance of raptor codes with the same inner LT codes for a BEC with zero erasure rate are simulated in Figs. 2.3 and 2.4. Once a new symbol is received, we attempt to decode both the inner LT codes and the LDPC codes using iterative BP algorithms until the decoding is completely successful or stuck. The transmitter continues to send more encoded symbols until all the information symbols are successfully decoded. Figs. 2.3 and 2.4 show histograms of the numbers of received symbols required to successfully decode the weakened LT codes and the raptor codes. The raptor codes require very little overhead while the LT codes require a large overhead to successfully decode the message. When the weakened LT codes are used, the number of decoded information symbols versus the number of received encoded symbols for three different randomly chosen realizations are shown in Fig. 2.5. From the intermediate performance of the inner LT codes shown in Fig. 2.5, it can be seen that after collecting slightly more than 10000 encoded symbols, only a very small portion of source symbols are not covered, which can be easily decoded by the pre-code.

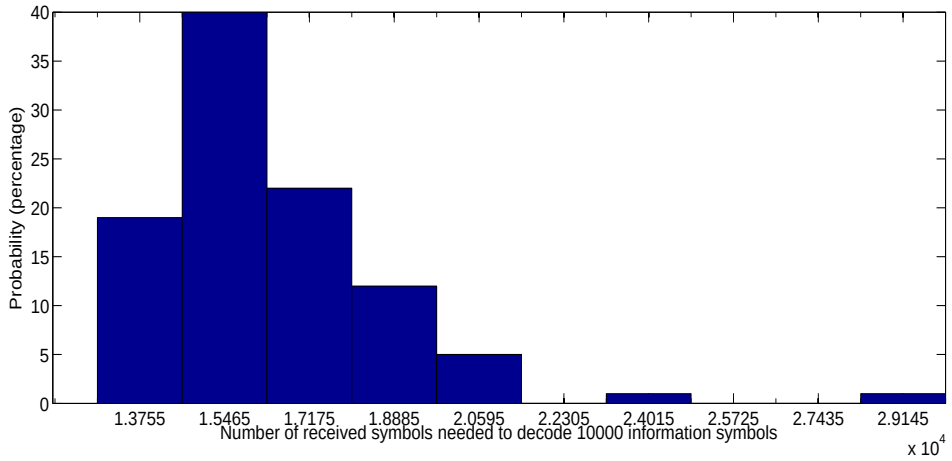

Figure 2.3. Histogram of the number of received symbols for successfully decoding the inner LT codes with code dimension $K = 10000$.

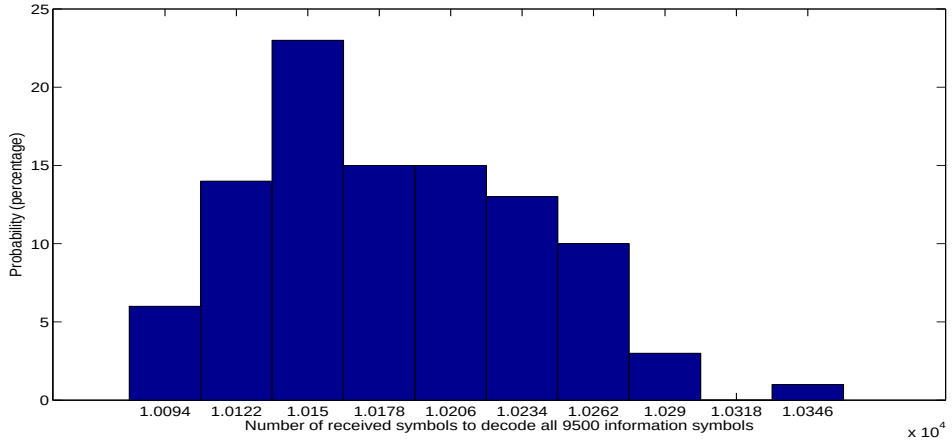

Figure 2.4. Histogram of the number of received symbols for successfully decoding the raptor codes with code dimension $K = 9500$ and a rate-0.95 LDPC code as the precode.

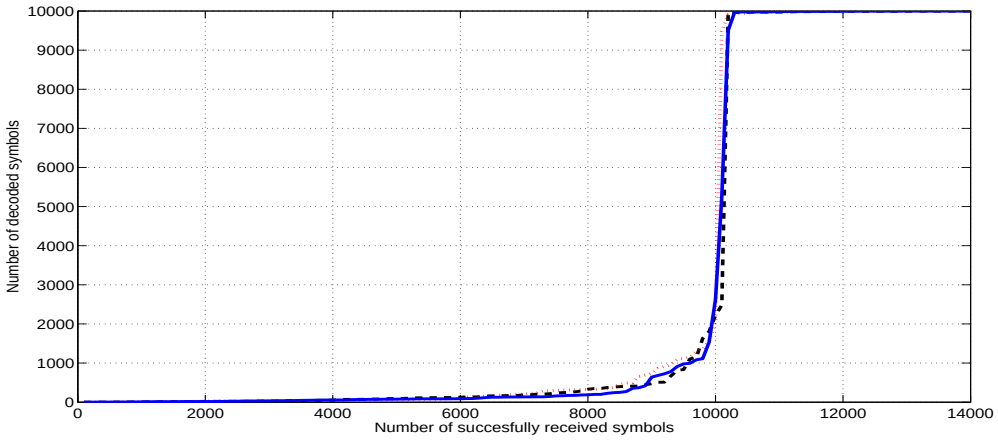

Figure 2.5. Intermediate performance of the weakened LT codes with code dimension $K = 10000$ and degree distribution $\Omega(x) = \Omega_r(x)$.

2.2 Raptor codes in noisy channels

The success of raptor codes in binary erasure channels (BEC) suggests that the fountain concept may be extendable to more general types of channels, such as additive white Gaussian noise (AWGN) channels. The use of fountain codes in noisy channels can be described in the following way: the receiver collects output symbols and calculates the channel capacity. Once channel capacity exceeds the realized code rate, the receiver starts to attempt to decode. The performances of raptor codes over binary symmetric channels (BSC) and Binary input AWGN (BIAWGN) channels are simulated in [23] [22]. Decoding of raptor codes is normally performed iteratively using BP algorithms, which have been widely used in decoding linear codes such as LDPC codes [30] and turbo codes [31] [32]. By merging the Tanner graph of LT codes (Fig. 2.1) and the Tanner graph of the LDPC pre-codes

(Fig. 2.2) into one Tanner graph, the message passing rule of the BP decoding over the Tanner graph of raptor codes is very similar to that of LDPC codes. For a LT output bit z_i, the initial log-likelihood ratio (LLR), defined as $LLR(z_i) = \ln(P(z_i = 0|y_i)/P(z_i = 1|y_i))$ where y_i is the corresponding received symbol, can be calculated based on the channel information [30]. For BEC,

$$LLR(z_i) = \begin{cases} +\infty, & y_i = 0 \\ -\infty, & y_i = 1 \\ 0, & y_i = erasure. \end{cases} \tag{2.5}$$

For BSC with symbol-error probability ε,

$$LLR(z_i) = (-1)^{y_i} \ln((1-\varepsilon)/\varepsilon). \tag{2.6}$$

For BIAWGN channels with noise variance σ^2,

$$LLR(z_i) = 2y_i/\sigma^2. \tag{2.7}$$

The initial LLRs of the LT input bits can be set to zero. The LLRs are iteratively updated by passing messages between the variable nodes and check nodes of the Tanner graph. After a pre-determined number of iterations, a hard decision is made based on the final LLRs. In practice, the receiver has to attempt to decode at some point and relies on a cyclic redundancy check (CRC) to determine whether the decoding results are correct. A good starting point is when the realized code rate is below the channel capacity. When decoding fails, the receiver waits for a certain number of symbols and attempts to decode again [33]. The waiting time has to be carefully chosen as too little waiting time will result in high complexity and too much waiting time will result in lower performance [34] [35].

In [23] and [22], the performances of LT and raptor codes are simulated based on bit error rate (BER) for a given code rate. We simulate the performances of LT and raptor codes in Fig. 2.6, where the LT codes have a code dimension $K = 10000$ and the raptor

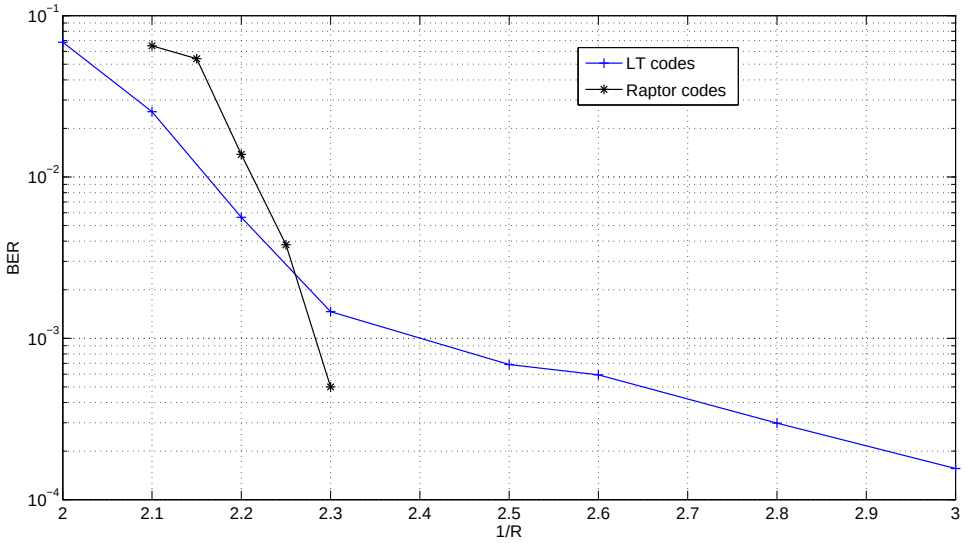

Figure 2.6. LT codes and Raptor codes over BSC channels with $\varepsilon = 0.11$

codes have a code dimension $K = 9500$ with a rate-0.95 LDPC code as the precode and both use the degree distribution $\Omega_r(x)$ given by Eq. (2.4). It can be seen that raptor codes easily outperform LT codes of the same rate. Also, LT codes have error floors even at a low rate. This is not only because there are some information symbols not connected to output symbols, but also because some symbols are not protected enough. The LT codes using RSD have similar error floors. In addition, the performance of raptor codes in BIAWGN channels has been simulated in [23]. The performance is very close to the Shannon limit, with an overhead of less than 10 percent. The above simulation results indicate that a raptor

code designed for a BEC channel can perform quite well for other types of channels. Similar simulation studies on raptor codes over fading channels have been carried out in [33]. Some further detailed analysis of raptor codes over BSC and BIAWGN channels appears in [22]. Besides simulation results, the authors in [22] use a Gaussian approximation (GA) method to derive criteria for the output degree distribution to produce good raptor codes. A lower bound of the fraction of output symbols of degree 2 is derived, which depends on the noise level in the BIAWGN and BSC channels. This observation suggests that a universal raptor code which can perform arbitrarily close to Shannon limits for any noise level of AWGN channel may not exist. However, [22] also proves that asymptotically, a universal raptor code designed for BECs needs at most $\log_2 e$ overhead when applied over BSCs using the BP decoding algorithm.

Despite the relatively good performance of raptor codes over noisy channels, the complexity of the BP decoding of raptor codes in AWGN and fading channels is much larger than in erasure channels. Reducing decoding complexity is therefore a useful area of research. The results in [35] and [34] show that it is possible to utilize the output of the previous decoding attempt to initialize the current decoding attempt. Such an approach would reduce the total number of iterations required for a successful decoding when more than one decoding attempt is used.

LT and raptor codes are not the only implementable codes that are used to approximate a digital fountain (DF) over wireless communications channels. In [36], the author proposes to use turbo codes to approximate a DF over AWGN and fading channels. Turbo codes are among the most powerful codes that perform very close to Shannon capacity in AWGN

and fading channels. In [36], the information symbols are first passed to a rate $1/3$ turbo encoder. The output bits of the turbo encoder are then applied to a potentially limitless number of random interleavers to produce a rateless code. This *turbo-fountain* technique performs quite well as indicated by simulations. The simulation results show that the turbo-fountain code closely approximates a DF code in noisy and fading channels. However, there is still a gap between the performance of turbo-fountain codes and the theoretical capacity limits throughout the SNR range, suggesting that future improvements in the codes are possible. The turbo fountain codes are, in fact, a special type of raptor code, known as pre-code only (PCO) raptor codes [11]. The turbo code used in the turbo-fountain code forms the pre-code. The turbo-fountain code does not include an inner LT code, but uses a potentially endless set of random interleavers instead. The efficiency of such PCO raptor codes over erasure channels is analyzed in [37].

2.3 Application of fountain codes to wireless multicast

As described in the introduction, fountain codes are particularly useful in wireless multicast scenarios because of their superior performance and channel adaptivity without feedback. The advantage of using fountain codes for asynchronous multicast has long been justified since the invention of LT codes [13]. Later on, fountain codes have also been proposed for use over wireless channels. Raptor codes have been standardized in the 3GPP Multimedia Broadcast/Multicast Services (MBMS) system [14]. The MBMS are mobile multimedia services over GSM-based 3rd generation (3G) cellular networks. In MBMS, the raptor

codes are used at the application layer to provide packet-level protection as a complement to the FEC codes at the physical layer. The receiver drops the packets where uncorrected errors occur within the packet and uses raptor codes to recover the source packets at the packet-level. Since at the packet level, the channel is well approximated by a virtual erasure channel, all the advantages of raptor codes over erasure channels can be exploited.

In [24], the concatenation of LT codes and bit-interleaved coded modulation (BICM) is used in a wireless internet application, modeled as an erasure channel contaminated by AWGN. The performance of this scheme is very good when the channel SNR exceeds a threshold, but rather poor in low SNR. This is because in low SNR, the inner BICM fails to correct some errors in most packets and results in very low packet throughput. In [38], this problem is resolved by utilizing the soft information to perform iterative decoding of LT codes. This requires the use of CRC to verify whether decoding is successful. The decoding algorithms are the same as discussed in Section 2.2. This decoding process is considered as an information collection process. Universal optimal fountain codes over AWGN and fading channels are defined here as rateless codes such that as long as the mutual information the receiver collects exceeds the entropy of the source information, the receiver is able to recover the information. Assuming that universal optimal fountain code is used, the probability that the receiver cannot successfully decode the K information symbols at time N is given by

$$p_o(K,N) = P_r\{B\sum_{i=1}^{N} R_I(h_i) \leq K\} \tag{2.8}$$

where B is the bandwidth, h_i is the particular channel realization at time slot i and $R_I(h)$ is the mutual information supported by the channel. Similar definitions of ideal fountain

codes can also be found elsewhere such as in [33]. Simulation results show that the performance gain by utilizing soft information is very large in low SNR. Also, it is shown that when soft decoding of fountain codes is utilized, there is no need to use another channel code to protect the packets. This is because a single fountain code with soft decoding can provide better performance than that using both physical layer FEC and erasure fountain codes.

There are generally two types of wireless multicast services: download services and media streaming services. Download services require very high accuracy but can tolerate long delays. On the other hand, media streaming services are delay sensitive but can usually tolerate some small amounts of errors. Streaming is usually accomplished in such a way that the receiver can begin play-out by using the early received data while receiving new data. The design of delay sensitive asynchronous media delivery using fountain codes has been discussed in [39]. The original media data are divided into M segments, $\{S_1, S_2, ..., S_M\}$, and transmitted in parallel using fountain codes. S_1 is chosen to have the highest rate to minimize initial play out delay. Segmentation is a useful technique to balance the initial play out delay and bandwidth expansion. If $t_{DTS}(S_i)$ represents the latest time the receiver has to reconstruct segment S_i after play-out begins, for a given initial play-out delay δ, the outage probability can be expressed as

$$p_{out}(\delta) = 1 - \prod_{i=1}^{M} (1 - p_o(|S_i|, t_{DTS}(S_i) + \delta)) \tag{2.9}$$

where $p_o()$ is given by (2.8). Simulation results using universal optimal fountain codes defined in [38] show that the required initial play-out delay decreases when the average SNR increases. The results also show that such fountain codes outperform erasure fountain

codes by a large margin.

2.4 Systematic standardized raptor codes

Since the introduction of raptor codes [10], a fully specified version of raptor codes has been approved [40]. The raptor codes that have been specified in the 3GPP MBMS standard [14], known as standardized raptor codes, is a systematic version of raptor codes. The standardized raptor codes have refined the original raptor code design to ensure low encoding and decoding complexity as well as low overhead. The standardized raptor codes use a very efficient implementation of a maximum likelihood (ML) decoder as opposed to iterative decoder and therefore allow very fast encoding and decoding. Compared to Reed-Solomon codes, the computational complexity of standardized raptor codes is orders of magnitude less [21]. For details about the pre-code, the degree distribution and the parameters used for standardized raptor codes, readers can refer to [14] (Annex B).

The reception overhead performance of standardized raptor codes can be expressed by the decoding failure probability $P_e^r(M,K)$, which is the probability that the receiver fails to fully recover K source symbols after M symbols are successfully received. Extensive simulations have been conducted to evaluate the performance of standardized raptor codes. It has been shown in [20] [21] that for $K > 200$, the decoding failure probability of standardized raptor codes can be well modeled by the empirically determined equation,

$$P_e^r(M,K) = \begin{cases} 1 & \text{if } M \leq K \\ ab^{M-K} & \text{if } M > K \end{cases} \tag{2.10}$$

where a and b are constants given by $a = 0.85$, $b = 0.567$. The above results indicate that the decoding failure probability decreases exponentially once $M \geq K$. For a typical size of code dimension K, the overhead of standardized raptor codes is very small.

Chapter 3

Optimization of unequal error protection rateless codes for multimedia multicast

3.1 Introduction

Multimedia communications has gained an increasing amount of attention due to high demand for high-definition video content. In multimedia communications, while multicast solutions usually provide better bandwidth efficiency than multiple uni-cast solutions, multicast over lossy packet networks is still a challenging task due to heterogeneous user networks as well as stringent delay requirements. On the source coding side, scalable video coding (SVC) has been developed to allow for progressive reconstruction of multimedia content at the receiver. However, on the channel coding side, the design of forward error correction (FEC) codes for multimedia multicast faces the problem of heterogeneity. For example, Reed-Solomon (RS) codes are usually customized for specific loss rates. For heterogenous users with differing loss rates, traditional FEC design accommodates the

worst-case scenario, which is very inefficient.

Recently proposed rateless codes, also known as fountain codes [41], provide efficient and flexible FECs for broadcasting or multicasting over erasure channels. Fountain codes are rateless in the sense that the transmitter may generate, as needed, a potentially infinite number of encoded symbols. The receiver may successfully recover all information symbols by collecting any subset of the encoded symbols as long as the number of received symbols is slightly greater than the number of information symbols. Raptor codes [11], representing the most successful implementation of digital fountain codes, have been used as application layer FEC codes in the third generation partnership program (3GPP) Multimedia Broadcast/Multicast Services (MBMS) standard [14]. Fountain codes, highly efficient for broadcasting bulk data, can potentially improve flexibility and efficiency for multicast applications. However, the original raptor-code design [11] has very poor progressive decoding performance. On the other hand, multimedia content may have a scalable structure in which certain source symbols have higher priorities than others. Therefore, an efficient fountain code is sought for multimedia streaming applications that also provides unequal error protection (UEP) of source symbols.

The idea of priority encoding transmission (PET) was pioneered in [42] over a decade ago. Since then, many approaches to provide UEP for multimedia have emerged. For example, in [15], Mohr et al. propose a PET-based packetization scheme for transmitting compressed images over noisy channels. In [16] and references therein, Mohr's scheme is optimized to minimize end-to-end distortion. Similar optimizations of receiver-driven

networks have been investigated in [43]. A summary of these and other approaches to rate-distortion-based optimization can be found in [17]. Most prior work employs fixed-rate codes, such as RS codes. Because RS codes are maximum distance separable (MDS), code performance is usually not factored into the optimization. Rateless codes, on the other hand, are based on random linear codes, with overhead that varies for different designs and implementations. In this thesis, code performance is taken into account in the UEP rateless code optimization.

Recently, techniques have been proposed for UEP design of rateless codes. In [1], message symbols are encoded by non-uniform selection of source symbols, where the performance of more important symbols is improved at the expense of slightly decreased overall performance. The design in [1] has been applied to MPEG-II video transmission in [18]. In [2], expanding window fountain (EWF) codes are proposed, where source symbols are arranged to lie inside a sequence of windows that are nested and expanding in size such that the larger windows contain all the symbols of the smaller windows. In [19], the design of EWF codes in [2] is applied to scalable video multicasting. However, the UEP raptor code schemes in [18] [19] do not provide user QoS guarantees. In addition, the designs in [1] and [2] have inherent disadvantages: since both alter the overall Luby Transform (LT) code [9] degree distribution, they significantly change code behavior. As it is well known that both performance and complexity are sensitive to degree distribution, the above designs may worsen code behavior unless the degree distribution is jointly-optimized, which would be very complex.

Previous approaches to UEP optimization for both fixed-rate FEC codes [15] [16] [17]

and rateless codes [18] [19] focus almost exclusively on providing best-effort QoS, i.e., maximization of an average fidelity measure of video/image quality of end users for a given transmission rate. However, because rateless codes have no pre-determined rate at the transmitter, guaranteed QoS may be achieved by simply transmitting enough coded symbols to meet users' QoS demands. Therefore, It is important to minimize the number of transmitted coded symbols to meet user QoS demands.

In this chapter, we address the open problem of guaranteed QoS optimization, as well as address best-effort QoS optimization of UEP rateless code design for multimedia multicasting to heterogeneous users, i.e., users with differing reception capabilities and QoS requirements. We propose a simpler and more modular UEP rateless code design which encodes different layers separately and retains desirable properties of rateless codes. An important advantage over [19] and others is that the proposed UEP design allow for direct application of existing high-performance raptor codes, such as those used in the 3GPP standard [14], known as standardized raptor codes. In addition, the receiver is able to decode each layer separately to significantly reduce decoding complexity. Section 3.2 describes the system setup and proposed UEP rateless code design; Section 3.3 presents problem formulations for different scenarios; Section 3.4 provides the solution for guaranteed QoS. In Section 3.4, for the proposed UEP raptor codes, the original problem formulation for guaranteed QoS is transformed into a convex optimization problem where optimal selection probabilities are obtained analytically. In contrast, the algorithm in [19] requires a numerical search to find the optimum, which is a complex task when the number of user classes is large. Comparisons with the UEP rateless coding schemes in [1] and [19] for

both guaranteed and best-effort QoS problems are provided in Section 3.5.

Related work is presented in [25] and [26], which investigates optimal UEP rateless code design using a packetization structure similar to that in [15] in a different asynchronous multimedia multicast system setup and uses computationally expensive iterative search to find the optimal allocation parameters. Unlike the proposed approach, the rateless code used in both [25] and [26] is assumed to have fixed overhead and hence code performance is not factored in.

3.2 System setup and proposed design

3.2.1 System setup

A multimedia server transmits multimedia content simultaneously to multiple users, which may include streaming with strict delay requirements. Therefore, it is not feasible to use a large-dimension fountain code to protect the entire multimedia source. Instead, the multimedia content is divided into multiple coded blocks. The server first compresses each source block using a pre-defined source coder and then encodes the source information using a UEP rateless code. The encoded symbols are multicast over a wireless lossy packet network.

User subscribers are categorized into different classes due to different reception capabilities. Assume that there are J classes of users. For Class j users, the reception capability δ_j is defined as the proportion of symbols that the receiver can successfully receive compared to the number of transmitted symbols, where $1 \leq j \leq J$. Therefore, in each transmission

session, the number of successfully received encoded symbols for each user in Class j is $\delta_j M$, where M is the number of symbols transmitted [1]. Without loss of generality, we assume $0 < \delta_1 \leq \delta_2 ... \leq \delta_J \leq 1$. The reception capabilities are determined by channel quality and bandwidth between server and receiver. For example, a Class 1 user may represent a mobile cell phone with limited reception quality due to size and power restrictions, while a Class 2 user may represent an automobile equipped with a larger antenna and better reception quality. Users in different classes also have different QoS requirements. An *outage QoS guarantee* refers to the ability of users to recover a given portion of source data with an achieved target probability. Without loss of generality, peak signal-to-noise ratio (PSNR), a common measure of visual quality, is used in the proposed QoS formulation. The proposed formulation is also applicable to QoS requirements defined using other measures of media quality.

Let K represent the number of information symbols in a raptor-coded source block. Assume the server is transmitting $M = (1+\varepsilon)K$ encoded symbols in order to meet all the subscribed users' QoS demands, where ε is the transmission overhead. The rateless coded source block is divided into L layers in order of importance, where Layer 1 contains the most important source symbols and Layer L contains the least important source symbols. For example, for a compressed video or image file, Layer 1 typically represents the base layer (BL), and Layer 2 represents the first enhancement layer (EL) etc. The number of

[1] For analytical simplicity, the number of received symbols for each user class is modeled as being equal to δ_j multiplied by the transmission volume, which is the same as in [19]. In Appendix A or [44], an alternative model that takes into account the randomness of the number of received symbols is evaluated, and found to produce similar transmission overheads.

source symbols in Layer l is given by S_l. Therefore, $K = \sum_{l=1}^{L} S_l$.

3.2.2 Proposed UEP rateless code

We propose a random interleaved UEP rateless encoder to serve as FEC for multimedia multicast, whose structure is illustrated in Fig. 3.1. The encoder assumes that source

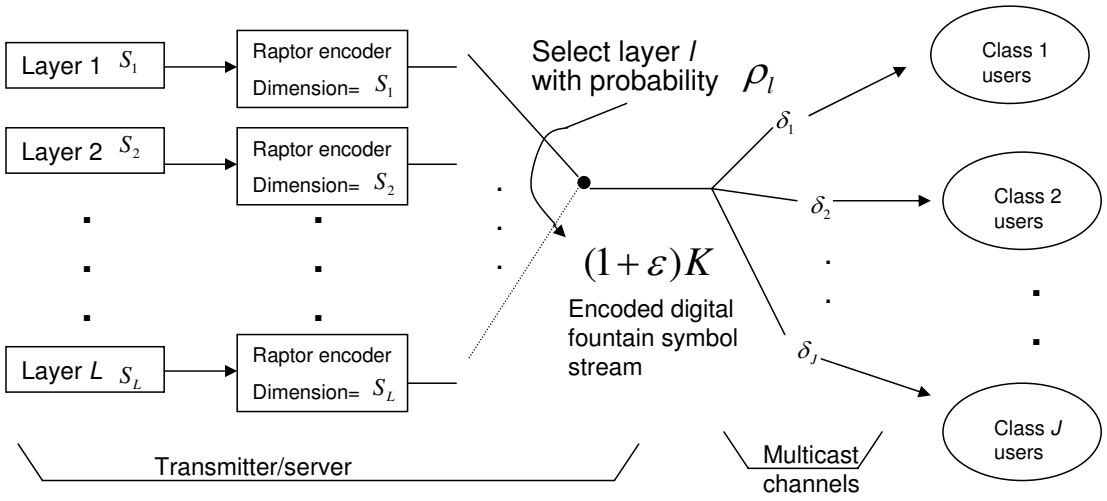

Figure 3.1. The proposed random interleaved UEP raptor coding method.

symbols have been allocated to the L layers prior to the encoding process. Encoding is performed in the following way: a layer is randomly selected where the probability of selecting layer l is ρ_l for $l = 1, 2, ..., L$ and $\sum_{l=1}^{L} \rho_l = 1$. Then, the encoded output symbol is generated by the raptor encoder for Layer l with code dimension S_l, degree distribution $\Omega_l(x)$ and precode C_l. Therefore, the overall encoded data stream is an interleaved stream of raptor encoded symbols from the L encoders. The proposed UEP rateless coded scheme extends the fountain code principle to random interleaving, which has origins in the design of turbo fountain codes [45] and EWF codes [19]. This approach is able to control the rate

of different layers using selection probabilities while maintaining the rateless property. In [45], the output turbo-coded symbols are randomly selected to create a potentially infinite number of rateless coded symbols. However, in [45], there is no probability distribution associated with the selection to create the UEP property. The probabilistic based encoding structure of the proposed UEP rateless codes is more similar to that of [19], except that independent coding layers are used rather than an over-lapping window approach. As a result, the raptor encoder and selection probabilities of the interleaver can be optimized separately. On the other hand, in [19], the degree distributions and selection probabilities need to be optimized jointly for best performance, a complex task that is not carried out. In practice, one can alter the ordering of the output symbols from the random interleaved UEP raptor coder by using scheduling algorithms to meet desirable goals while maintaining the priority of each layer. Some investigations on re-ordering rateless code output symbols have been recently proposed in [46] and [47]. However, these approaches only apply to limited situations where the ordering of the encoded symbols matters.

Note that unlike [25] [26], the proposed design does not specify a packetization structure and thus retains generality. Without loss of generality, the result in this thesis may be applied to data packets rather than symbols, creating packet erasure channels.

3.3 Problem formulations with QoS constraints

3.3.1 Guaranteed QoS formulation

In the system under consideration, we consider established users that require play-back media at a quality no lower than their own QoS requirement. In this scenario, since the transmitter has to provide guaranteed QoS for all user classes before the start of transmission of the next source block, the throughput of the system for each source block is determined by the maximum number of transmitted symbols required to satisfy the QoS of each individual user class. Therefore, the objective of the UEP raptor code design problem is to provide different levels of QoS guarantees according to users' requirements while minimizing the maximum transmission overhead ε:

Problem 3.10 (Guaranteed QoS):

$$\min_{\rho_1,\ldots,\rho_L} \quad \varepsilon \tag{3.1}$$

$$s.t.\ \text{Prob}(PSNR_j \geq \gamma_j) \geq P_j, \quad j = 1,2,...,J, \tag{3.2}$$

where $PSNR_j$ represents the PSNR of the successfully recovered source data of the class j user given $M = (1+\varepsilon)K$ transmitted symbols, and γ_j and $1-P_j$ denote the target PSNR threshold and target outage probability for the Class j user, respectively. The aim is to allocate coding rates across layers through optimization of the probabilities ρ_l, $1 \leq l \leq L$.

We require the source (e.g., video, image) coder to be progressive, so that the reconstruction media quality is determined mainly by the symbol errors in the lowest layer encountered in the recovery process. Let $q_l, l = 1,2,...,L$, represent the PSNR that is achieved

when Layers 1 to l are successfully recovered in the raptor decoding process. For a given source coder, if the source PSNR is represented as a non-decreasing function of the total number of source symbols decoded by the receiver, denoted by $f(.)$, then we have $q_l = f(\sum_1^l S_l)$ and $q_1 \leq q_2 ... \leq q_L$. For each class j, $1 \leq j \leq J$, let $g_j \in \{1,2,...,L\}$ be the minimum index that satisfies $q_{g_j} \geq \gamma_j$. In order to satisfy $PSNR_j \geq \gamma_j$, users in Class j require the raptor decoder to successfully decode at least Layers 1 to g_j. For a given UEP raptor code design, let $P_e(l, j)$ represent the probability that the Class j decoder fails to fully decode Layer l given transmission overhead ε and reception quality δ_j. In the most stringent case when decoding errors across layers are independent, the QoS requirements of end users can be simplified to

$$\prod_{l=1}^{g_j}(1 - P_e(l, j)) \geq P_j \quad j = 1, 2, ..., J. \tag{3.3}$$

3.3.2 Best-effort QoS formulations

While above formulation focuses on minimizing transmission overhead subject to satisfying users' guaranteed QoS, this section considers transmission overhead that is upper bounded due to delay constraints or cost. For this scenario, given a maximum transmission overhead, ε_0, the service provider attempts to provide users of different classes with the best possible QoS. While this scenario is similar to that considered in [19], the following best-effort QoS problem extends that in [19] by 1) considering both constrained and unconstrained cases, 2) allowing for allocating different weighting factors to different user classes as well as 3) possessing the advantages of the proposed random interleaved UEP raptor codes.

The average PSNR of users in Class j, which serves as a measure of the best-effort QoS, can be evaluated as $\mathrm{E}(PSNR_j) = \sum_{l=1}^{L} p_{l,j} q_l$, where q_l is the PSNR achieved when Layers 1 up to l are successfully recovered. The quantity $p_{l,j}$ represents the probability that a Class j user successfully recovers Layers 1 to l but fails to recover Layer $l+1$. The optimization of UEP rateless codes has to balance users of different classes with different channel qualities. If we assign a weighting coefficient w_j for Class j where $0 \leq w_j \leq 1$, $\sum_{j=1}^{J} w_j = 1$ and the choice of w_j depends on both the importance of the user class as well as the number of users in each user class, then it is reasonable to consider the weighted average PSNR over all user classes as the objective function. The problem can be formulated as:

Problem 3.20: (Best-effort QoS)

$$\max_{\rho_1,\rho_2,\ldots,\rho_L} \sum_{j=1}^{J} w_j \cdot \Big(\sum_{l=1}^{L} p_{l,j} \cdot q_l\Big) \tag{3.4}$$

$$\text{subject to} \quad \varepsilon \leq \varepsilon_0 \tag{3.5}$$

where

$$p_{l,j} = \begin{cases} \prod_{i=1}^{l}(1 - P_e(i,j)) \times P_e(l+1,j) & l = 1,2,\ldots,L-1 \\ \prod_{i=1}^{L}(1 - P_e(i,j)) & l = L, \end{cases} \tag{3.6}$$

and $P_e(i,j)$ is probability that Class j users fail to decode Layer i.

In Problem 3.20, no guaranteed minimum QoS is provided. To address this concern, for a given maximum transmission overhead, the service provider may instead aim to provide best-effort QoS to multiple user classes, but under the additional constraint that a minimum QoS guarantee for each user class is met. This problem can be formulated as:

Problem 3.30: (Best-effort QoS with constraints on individual classes)

$$\max_{\rho_1,\rho_2,\ldots,\rho_L} \sum_{j=1}^{J} w_j \cdot \Big(\sum_{l=1}^{L} p_{l,j} \cdot q_l\Big) \tag{3.7}$$

$$\text{subject to} \quad \varepsilon \leq \varepsilon_0 \tag{3.8}$$

$$\text{and} \ \prod_{l=1}^{g_j}(1-P_e(l,j)) \geq P_j \ \ j=1,2,\ldots,J \tag{3.9}$$

where $p_{l,j}$ is given by (3.6) and $g_j \in \{1,2,\ldots,L\}$ $(j=1,2,\ldots J)$ is the minimum layer index that satisfies $q_{g_j} \geq \gamma_j$. Problem 3.20 can be viewed as a special case of Problem 3.30 without the service-fulfillment probability constraints.

In the next section, we show that Problem 3.10 can be transformed to an equivalent convex optimization problem when standardized raptor codes are employed. Unfortunately, Problems 3.20 and 3.30 cannot be transformed to a convex optimization problem in the same way as Problem 3.10 due to the form of the $p_{l,j}$ expressions. However, Problems 3.20 and 3.30 can still be solved numerically by searching the $(L-1)$-dimensional parameter space of $\{\rho_1,\rho_2,\ldots,\rho_{L-1}\}$, checking the constraints (3.9) and the resulting average PSNR (3.7). When $L=2$, the numerical method is significantly simplified as it only needs to find the optimal $\rho_1 \in [0,1]$ that gives the maximum average PSNR. Numerical results comparing the proposed scheme with EWF codes [19] for Problems 3.20 and 3.30 are shown later in Fig. 3.8.

3.4 Solving the guaranteed QoS problem

3.4.1 Evaluation of decoding failure probability

An advantage of the proposed design is that existing high-performance standardized raptor codes can be directly applied. Standardized raptor codes are refined systematic raptor codes designed to ensure low encoding/decoding complexity and overhead. Details about the pre-code, degree distribution and construction of standardized raptor codes can be found in [14], (Annex B). When standardized raptor codes with maximum likelihood (ML) decoding are used, for code dimension greater than 200, the decoding failure probability, i.e., failure to decode k source symbols after m symbols are successfully received, can be accurately modeled by the empirically determined equation [20],

$$P_e^r(m,k) = \begin{cases} 1 & \text{if } m \leq k \\ ab^{m-k} & \text{if } m > k \end{cases} \tag{3.10}$$

where $a = 0.85$, $b = 0.567$ are constants. Note that for $k < 200$, Eq. (3.10) underestimates the error probability due to short block length. One way to improve code performance for layers with fewer symbols is to merge source layers with similar optimized selection probabilities ρ_j into larger layers. We note that, for high-rate media such as video, the condition $k < 200$ is unlikely to occur.

When more general LT or raptor codes using iterative decoding are employed, the decoding failure probability $P_e(l,j)$ can be approximated by

$$P_e(l,j) = 1 - (1 - e_{l,j})^{S_l}, \tag{3.11}$$

where $e_{l,j}$ is the symbol error probability of a Class j user decoding Layer l. This approximation is based on the assumption that symbol errors in iterative decoding are mutually independent. The symbol error probability $e_{l,j}$ of an iterative decoder can be analytically determined by "and-or" tree analysis [48] [1]. Applying the and-or tree technique, Eq. (3.11) can be evaluated using

$$e_{l,j}^{n} = \begin{cases} 1 & n = 0 \\ \exp(-(1+\varepsilon)K\rho_l\delta_j\Omega'(1-e_{l,j}^{n-1})/S_l) & n \geq 1 \end{cases} \tag{3.12}$$

where $\Omega(.)$ is the LT code degree distribution, $\Omega'(x)$ denotes derivative with respect to x, and n is the number of decoding iterations. The asymptotic symbol error probability of iterative decoding $e_{l,j} = \lim_{n\to\infty} e_{l,j}^{n}$ can be estimated by choosing a large value n in Eq. (3.12) [48].

3.4.2 Convexity analysis

For a given transmission overhead ε, the total average number of encoded symbols transmitted for Layer l in each transmission block is defined by $t_l = (1+\varepsilon)K\rho_l$, and satisfies $\sum_{l=1}^{L} t_l = (1+\varepsilon)K$. When standardized raptor codes are used, substituting $m = t_l\delta_j$ and $k = S_l$ into Eqs. (3.10) and (3.3), and taking the logarithm of the constraints described by (3.3), Problem 3.10 is transformed to:

Problem 3.11:

$$\min_{t_1,\ldots,t_L} \quad \Sigma_{i=1}^{L} t_i \tag{3.13}$$

$$s.t. -\Sigma_{l=1}^{g_j} \log[1-c_l\alpha_j^{t_l}] + \log P_j \leq 0, \quad j = 1,2,\ldots,J, \tag{3.14}$$

where $c_l = ab^{-S_l}$, $\alpha_j = b^{\delta_j}$ and $g_j \in \{1,2,...,L\}$. The constraints that $t_l\delta_j \geq S_l, j = 1,2,...,J, l = 1,2,...,g_j$ are implicitly guaranteed by the $\log(.)$ function. To ensure an integer solution, we compute $t_1,t_2,...,t_l$ as if real-valued, then round to the nearest larger integer. Although the above transformation uses the decoding failure probability evaluation of standardized raptor codes given by Eq. (3.10), a similar method can be applied to other decoding failure probability models that can be approximated by an exponential function.

To solve Problem 3.11, we first prove convexity. As the objective function is linear, we only need to prove that the constraint functions are convex with respect to $t_i, i = 1,2,...,L$. It can be shown that for $l = 1,2,...,L$, the second derivatives of $-\log(1-c_l\alpha_l^{t_l})$ with respect to t_l satisfy

$$\frac{\partial^2[-\log(1-c_l\alpha_l^{t_l})]}{\partial t_l^2} = \frac{c_l\alpha_l^{t_l}(\log\alpha_l)^2}{(1-c_l\alpha_l^{t_l})^2} > 0 \quad j = 1,2,...,J. \tag{3.15}$$

According to the second order condition of convex functions [27], $-\log[1-c_l\alpha_l^{t_l}]$ is a convex function of t_l. Since nonnegative weighted sums preserve convexity [27], the constraint functions (3.14) are convex functions of the vector $\mathbf{t} = [t_1,t_2,...,t_L]^T$. Problem 3.11 can therefore be solved numerically by available convex optimization algorithms [27].

Let $\mathbf{t} = [t_1,t_2,...,t_L]^T$ and $\lambda = [\lambda_1,\lambda_2,...,\lambda_J]^T$ be the variable vectors of the primal and dual problems of Problem 3.11, respectively. If $\mathbf{t}^* = [t_1^*,t_2^*,...,t_L^*]^T$ and $\lambda^* = [\lambda_1^*,\lambda_2^*,...,\lambda_J^*]^T$ represent sets of primal and dual optimal points, they must satisfy the Karush-Kuhn-Tucker (KKT) optimality conditions for objective functions $f_0(.)$ and constraint functions $f_j(.)$:

$$f_j(\mathbf{t}^*) \leq 0, \quad j = 1,2,...,J \tag{3.16}$$

$$\lambda_j^* \geq 0, \quad j = 1,2,...,J \tag{3.17}$$

$$\lambda_j^* f_j(\mathbf{t}^*) = 0, \quad j = 1,2,...,J \tag{3.18}$$

$$\nabla f_0(\mathbf{t}^*) + \sum_{j=1}^{J} \lambda_j^* \nabla f_j(\mathbf{t}^*) = 0 \tag{3.19}$$

where here $f_0(\mathbf{t}) = \Sigma_{i=1}^{L} t_i$ and $f_j(\mathbf{t}) = -\Sigma_{l=1}^{g_j} \log[1 - c_l \alpha_j^{t_l}] + \log P_j, j = 1,2,...,J$. Since the original Problem 3.11 is convex and satisfies Slater's condition, the above KKT optimality conditions provide the necessary and sufficient conditions for optimality [27]. In general, solving the KKT condition is not straightforward. However, if we can identify a set of inequality constraints that are most likely to be active, i.e., achieve equality at the optimal solution, then we can obtain a corresponding set of primal and dual solution points and verify the optimality with KKT condition.

A simplification to the problem arises if we have a one-to-one mapping between user classes and channel coding layers, i.e., $g_j = j$ for $j = 1,2,..,J$ and $L = J$, which is the assumption used in the formulation of [19], and if all the inequality constraints are active. Using the above assumption, the solution to Problem 3.11 can be obtained by finding t_1 using the constraint for Class 1 in Eq. (3.14) and substituting the solution of t_1 into the next constraint, solving for t_2 with the constraint for Class 2 in Eq. (3.14) etc. until all of the variables $t_1, t_2, ..., t_L$ are determined. However, since this simplification has not been proven to be equivalent to Problem 3.11 in general, the solution obtained in this manner has to be verified using the KKT optimality conditions. If all the inequality constraints are active, Eqs. (3.16) and (3.18) are automatically satisfied. Therefore, if we obtain a solution $\mathbf{t}^*$ of Problem 3.11 by solving $f_j(\mathbf{t}^*) = 0, j = 1,2,...,J$, we can substitute the value of $\mathbf{t}^*$ into Eq.

(3.19) and obtain λ^*. If λ^* satisfies Eq. (3.17), i.e., $\lambda_j^* \geq 0, j = 1,2,...,J$, then we have proven that the value of $\mathbf{t}^*$ we obtained is indeed an optimal solution of Problem 3.11. If the KKT optimality condition is not satisfied, then numerical method can be used to solve this convex optimization problem.

In the following, we propose an algorithm to transform a general guaranteed QoS problem into a problem with one-to-one mapping between user classes and channel coding layers. The idea is to reduce the dimensionality of the problem by removing redundant user constraints and merging source coding layers. The process is explained in the following algorithm:

3.4.3 Class-to-layer mapping algorithm

Algorithm 3.1: (Class-to-layer mapping algorithm)

Step 1 (User class amalgamation): Repeat the following class amalgamation operation until $g_i < g_k$ for every $i < k$, where $1 \leq i \leq J, 1 \leq k \leq J$: for any pair of user class indices i and k where $i < k$ (hence $\delta_i < \delta_k$), if Class i users have the same or higher target PSNR threshold than Class k users (i.e., $\gamma_i \geq \gamma_k$ or $g_i \geq g_k$), we absorb Class k into Class i.

Step 2 (Source layer merging): Repeat until for every layer $1 \leq l \leq L$, there exists a class $j, 1 \leq j \leq J$ such that $g_j = l$: if there exists a source layer l where there is no corresponding user class (i.e., no j exists such that $g_j = l$), Layers l and $l+1$ are merged to form a new source layer l' with code dimension $S_{l'} = S_l + S_{l+1}$.

Step 1 finds a set of the most demanding user classes with respect to their channel conditions; Step 2 reduces the number of channel coding layers to the minimum without

compromising the performance. After performing Algorithm 3.1, we can show the following fact:

Lemma 3.1: After performing Algorithm 3.1, $L = J$ and $g_j = j$ for $j = 1, 2, ..., J$. If for every Class k that has been absorbed into Class i in Step 1, $P_i \geq P_k$ is also satisfied, then the new optimization problem after performing Algorithm 3.1 is equivalent to Problem 3.11. In addition, any further partitioning of layers cannot reduce the minimum transmission overhead required to achieve the QoS requirements.

Proof: First we show that any QoS constraint dropped from Step 1 (user class amalgamation) is irrelevant. Suppose the QoS constraint of Class i users is satisfied, i.e., $\prod_{l=1}^{g_i}(1 - P_e(l,i)) \geq P_i$. Since $i < k$, we have $\delta_i < \delta_k$. Hence, Class k users receive more coded symbols than Class i users. Therefore, the decoding failure probability $P_e(l,i) > P_e(l,k)$ for all $1 \leq l \leq L$. Then, because $g_i \geq g_k$, from the assumption of Lemma 3.1, $P_i \geq P_k$, and

$$\prod_{l=1}^{g_k}(1 - P_e(l,k)) > \prod_{l=1}^{g_k}(1 - P_e(l,i)) \geq \prod_{l=1}^{g_i}(1 - P_e(l,i)) \geq P_i \geq P_k. \tag{3.20}$$

Hence, the QoS constraint for Class k users is also satisfied.

Next we show that after performing Algorithm 3.1, the number of source layers L and the number of user classes J are equal. The class amalgamation procedure ensures that the set g_j, $j = 1, 2, ..., J$ is monotonically increasing with j. This fact does not change after performing the source layer merging procedure. Since $g_j \in \{1, 2, ..., L\}$, we have $L \geq J$. On the other hand, source layer merging ensures that for any $l = 1, 2, ...L$, there exists an integer $j \in \{1, 2, ..., J\}$ such that $g_j = l$. Therefore, we also have $L \leq J$. Thus, $L = J$. Together with the fact that g_j is monotonically increasing with j, we can conclude that

$g_j = j$ for $j = 1, 2, ..., J$.

Finally, to complete the proof, we need to show that any further partitioning of layers cannot reduce the required minimum transmission overhead. The details can be found in Appendix B. **QED**.

Remark 1: The condition that for every Class k that has been absorbed into Class i, $P_i \geq P_k$, is a sufficient condition for Lemma 3.1 but not a necessary condition. Even if this condition is not satisfied, it is possible that the transformed problem due to Algorithm 3.1 results in the optimal solution. In addition, if this condition is violated, to ensure that the optimal solution of the transformed problem is the optimal solution of the original problem, we can always verify if the obtained solution satisfies all the constraints of the user classes that have been amalgamated in Step 1. If not, we can resort to solving the convex problem (Problem 3.11) numerically. This is further illustrated in Section 3.4.4.

Remark 2: For best-effort QoS Problem 3.30, the transformation described by Algorithm 3.1 may not be optimal. Since optimal solution also depends on the fidelity measure of the multimedia source.

3.4.4 A numerical example on video multicasting

We now illustrate the mapping and solutions to the guaranteed QoS problem in an example where the transmitter is multicasting a H.264 SVC [49] coded stream. We use a H.264 SVC video sequence which contains a total of 15 layers with one base layer (BL) and fourteen enhancement layers (ELs). Since the focus of the paper is on optimizing the channel coder for a given source coder, the number of information symbols and the corresponding PSNR

Table 3.1. Example of user classes and their QoS requirements

User class index (j)	1	2	3	4
User reception capability δ_j	0.4	0.5	0.6	1
User QoS requirement (PSNR threshold γ_j(dB))	25.79	29	27.25	40.28
number of decoded symbols to achieve QoS	400	1155	700	3800
number of decoded source layers required (g_j)	1	4	2	15
probability threshold P_l	0.8	0.9	0.85	0.95

values of this H.264 SVC video sequence are taken from Table I of [19]. We consider the same source symbol size as [19], where each source symbol represents 400 source bits. The UEP rateless encoders and decoders are operated at the symbol level. We assume there are four classes of users with reception capabilities and QoS requirements shown in Table 3.1.

In order to simplify the mapping between classes and layers, we observe that $g_2 > g_3$ while $\delta_2 < \delta_3$, which means that Class 3 users have better reception capabilities than Class 2 users, while at the same time, have a lower PSNR requirement. Therefore, the QoS constraint from Class 3 users can be dropped. In addition, since the number of layers required by the three classes are 1, 4 and 15, after the layer-merging procedure of Algorithm 3.1, we obtain a new set of channel layers with Layer 1 comprising the BL, Layer 2 consisting of the first 3 ELs, and Layer 3 consisting of the fourth to fourteenth ELs. Since $P_2 > P_3$ in Table 3.1, from Lemma 3.1, the new problem after mapping is equivalent to the original problem. The parameters of the new problem after the mapping are shown in Table 3.2.

Table 3.2. User classes and QoS requirements after the mapping

Combined class-layer index (j or l)	1	2	3
reception capability δ_j	0.4	0.5	1
PSNR threshold γ_j(dB)	25.79	29	40.28
number of decoded symbols to achieve QoS	400	1155	3800
number of decoded layers required $g_j = j$	1	2	3
probability threshold P_j	0.8	0.9	0.95
number of symbols S_l in each layer	400	755	2645

When standardized raptor codes are applied to the three new layers, the problem is to find ρ_1, ρ_2 and ρ_3 to minimize $t_1+t_2+t_3$ such that

$$\begin{cases} (1-ab^{(t_1\delta_1-S_1)}) \geq P_1 \\ (1-ab^{(t_1\delta_2-S_1)})(1-ab^{(t_2\delta_2-S_2)}) \geq P_2 \\ (1-ab^{(t_1\delta_3-S_1)})(1-ab^{(t_2\delta_3-S_2)})(1-ab^{(t_3\delta_3-S_3)}) \geq P_3. \end{cases} \tag{3.21}$$

Assuming all the inequality constraints are active, we obtain a minimum overhead $\varepsilon_{min} = 36.2\%$, which is achieved when $\rho_1 = 0.1946$, $\rho_2 = 0.2933$ and $\rho_3 = 0.5121$. This solution is then verified using KKT conditions and we find that this solution is indeed optimal. In contrast, an equal error protection (EEP) allocation requires a minimum overhead of 152%, a factor of four higher than the optimal UEP solution.

With the optimal selection parameters, we find that Class 3 users of the original problem (Table 3.1) can successfully decode the base layer and one enhancement layer with a probability higher than 99.9%. This means that even if the target probability threshold

$P_3 = 99\%$ in Table 3.1, which violates the assumption of Lemma 3.1, the problem transformed by Algorithm 3.1 still has the same optimal solution as the original problem. As a further remark, suppose the conditions of Lemma 3.1 are violated, we assume the extreme case of $P_3 = 99\%$ and vary the value of δ_3 within the range $0.5 = \delta_2 < \delta_3 \leq 1 = \delta_4$. We find that only when $0.5 < \delta_3 < 0.503$, our obtained solution does not satisfy the QoS constraint of Class 3 users; we expect that, in practice, two distinct classes to have greater reception capability difference than $\frac{0.003}{0.500} \times 100\% = 0.6\%$.

3.5 Numerical and simulation results

To demonstrate the advantages of optimization, the optimized UEP raptor code scheme is first compared to EEP raptor coding. For simplicity, only two layers are considered. In these comparisons shown in Figs. 3.2 to 3.4, the proposed random interleaved UEP design employing standardized raptor codes is used. The code dimension of the standardized raptor code used in layer l is S_l. The small inefficiencies incurred by the standardized raptor codes are characterized by the decoding failure probability $P_e^r(.)$ in Eq. (3.10). The optimal selection probability ρ_1 and minimum overhead for the UEP scheme are obtained by the simplified method described in Section 3.4 for solving Problem 3.11, i.e., by assuming that all inequality constraints are active. All the results shown in Figs. 3.2 to 3.4 satisfy the KKT optimality conditions after verification. The EEP scheme allocates each encoded symbol such that every information symbol has the same priority. Therefore, for EEP, the ratio ρ_1/ρ_2 is fixed to S_1/S_2. For all the results shown in this section, the parameters chosen

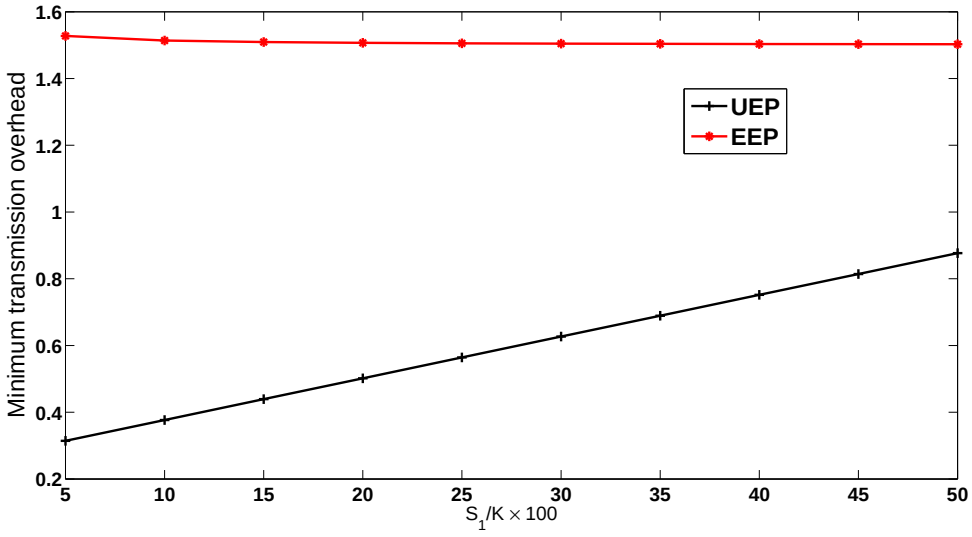

Figure 3.2. UEP versus EEP for varying S_1/S_2, $K = 9000; \delta = [0.4, 0.8]; P = [0.95, 0.8]$.

are described in the caption of each corresponding figure. Fig. 3.2 shows the minimum transmission overhead requirement for optimized UEP and EEP raptor codes as the ratio between the numbers of bits in the two layers, S_1/S_2, is varied. Fig. 3.3 shows the same comparison as the channel reception quality δ_1 of the first user class varies. It can be seen that UEP has a significant advantage over EEP in almost all cases. When the channel reception qualities of the two classes approach each other, EEP approaches the optimized UEP in performance. Fig. 3.4 shows the minimum overhead when different values of ρ_1 are used. It can be seen that the minimum required transmission overhead is very sensitive to the choice of ρ_1, and an arbitrary non-optimized allocation scheme may perform much worse than both the optimal allocation scheme and EEP.

Next, to investigate the performance of the proposed random interleaved UEP raptor

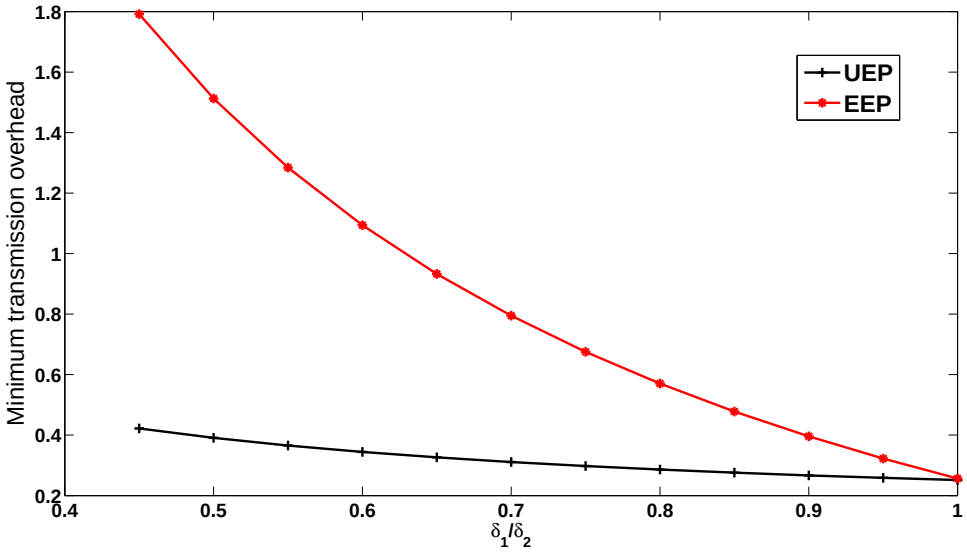

Figure 3.3. UEP versus EEP for varying δ_1/δ_2, $S = [1000, 8000]; \delta_2 = 0.8; P = [0.95, 0.8]$.

coding, we compare its performance to a recent UEP raptor code design proposed by Rahnavard *et al.* [1] as well as EWF codes proposed by Vukobratovic *et al.* [19] in Figs. 3.5 and 3.6. We remark that [1] and [19] cannot deploy existing standardized raptor codes directly. Therefore, for a fair comparison, instead of using standardized raptor codes, we use LT codes with iterative decoding and degree distribution

$$\begin{aligned}\Omega_r(x) = 0.007969x + 0.493570x^2 + 0.166622x^3 \\ +0.072646x^4 + 0.082558x^5 + 0.056058x^8 + 0.037229x^9 \\ +0.055590x^{19} + 0.025023x^{65} + 0.003135x^{66}\end{aligned} \tag{3.22}$$

for all the layers of our proposed random interleaved UEP code. It should be stressed that the advantage of being able to utilize high performance standardized raptor codes is not shown in these comparisons, which would further favor the proposed design. The degree

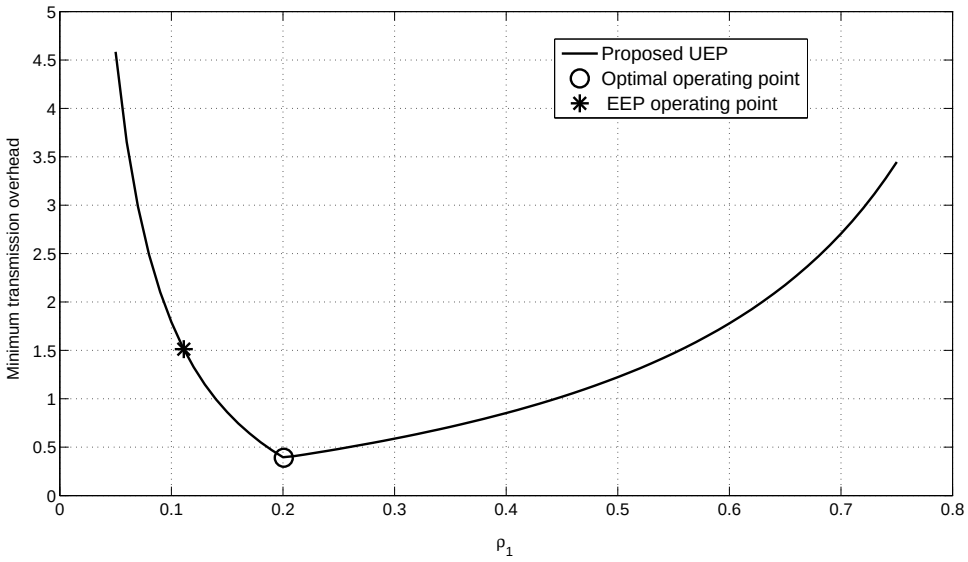

Figure 3.4. The effect of layer allocation probability ρ_1, standardized raptor codes, $L = 2; K = 9000; S = [1000, 8000]; \delta = [0.4, 0.8]; P = [0.95, 0.8]$.

distribution, (3.22), originally from [11], has been adopted by [1]. In Figs 3.5 and 3.6, the same degree distribution $\Omega_r(x)$ is applied to Rahnavard's UEP raptor code and all the windows of the EWF code. For analytical simplicity, no pre-code is used for all three schemes. The decoding failure probability P_e on the left side of the constraint functions in Eq. (3.3) is evaluated as follows: the symbol error probability of each layer in Rahnavard's scheme and the EWF code can both be estimated by the "and-or" tree technique [1] [19]. The symbol error probability e_l of Layer l for Rahnavard's scheme, the EWF code, and the proposed UEP scheme are obtained using Eqs. (6) and (7) in [1], Eq. (7) in [2], and Eq. (3.12) in this paper, respectively. The failure probability of decoding each layer can then be estimated as $P_e(l) = 1 - (1 - e_l)^{S_l}$.

Fig. 3.5 shows the minimum transmission overhead required to satisfy all user constraints of the proposed UEP scheme and Rahnavard's UEP scheme with different values for parameter k_M [1], which governs the degree of non-uniformity of input symbol selections. It can be seen that the optimized proposed UEP scheme outperforms Rahnavard's UEP scheme even if k_M is optimized. Fig. 3.6 shows a similar comparison between the proposed scheme and the EWF code. The size of the first window in the EWF code is fixed to the number of symbols in Layer 1 (S_1). Parameter Γ_1 is the probability of choosing the first window (the more important layer) during encoding (see [19]). It can be observed that the proposed scheme, when optimized, also performs better than the EWF code with optimized Γ_1. Note that there are two local minima in Fig. 3.6 because the evaluated symbol error rates of the more important bits are not monotonically decreasing as Γ_1 increases (see Fig. 1 in [19]).

As pointed out in [2], one of the advantages of EWF codes over Rahnavard's UEP scheme is flexibility in using different degree distributions applied to different windows. Therefore, in order to isolate the performance gains due to different code structures, degree distributions and decoder efficiencies, we use different degree distributions applied to different windows of EWF codes as well as to different layers of the proposed UEP scheme. Fig. 3.7 shows the performance of the three different UEP schemes using LT codes after optimization over their respective configuration parameters. The performance curves depict minimum transmission overhead for different sets of parameters and variable numbers of symbols in the first layer/window. The degree distributions chosen for the more important bits (MIB) and less important bits (LIB) are denoted as $\Omega_1(x)$ and $\Omega_2(x)$, respectively.

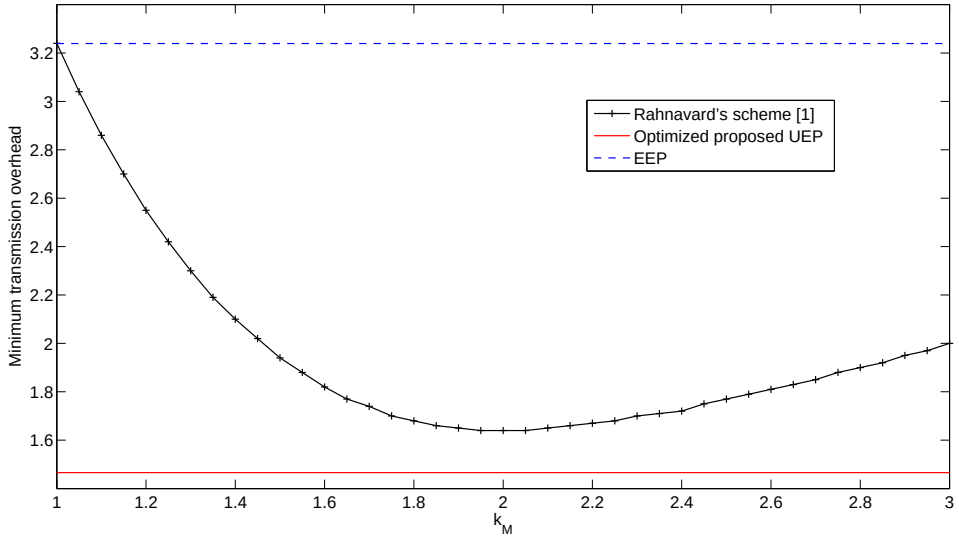

Figure 3.5. Proposed scheme versus [1]. LT codes with iterative decoding, $L = 2; K = 9000; S = [1000, 8000]; \delta = [0.4, 0.8]; P = [0.95, 0.8]$.

Apart from using degree distribution $\Omega_r(x)$ described by Eq. (3.22), we also show the performance when a truncated robust soliton distribution (RSD) $\Omega_{rs}(k_{rs}, \delta, c)$, where k_{rs} is the maximum degree, is applied to the MIB for EWF codes and the proposed random interleaved UEP scheme. The truncated RSD is a stronger degree distribution compared to $\Omega_r(x)$ at the cost of higher decoding complexity. It can be seen from Fig. 3.7 that using the stronger truncated RSD for the MIB provides a significant performance boost for both the EWF codes and the proposed UEP codes. Nevertheless, when both schemes use the truncated RSD distribution for the MIB, the minimum required transmission overhead for the proposed UEP scheme is still lower. A more important advantage of the proposed scheme is easy adoption of standardized raptor codes. As shown in Fig. 3.4, the minimum overhead

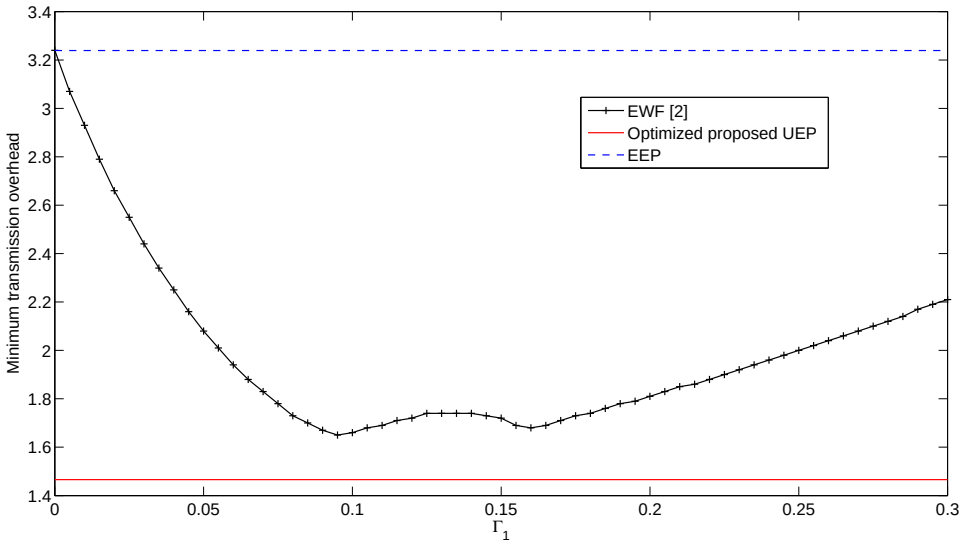

Figure 3.6. Proposed scheme versus the EWF code [2]. LT codes with iterative decoding, $L = 2; K = 9000; S = [1000, 8000]; \delta = [0.4, 0.8]; P = [0.95, 0.8]$.

for $S_1 = 1000$ using a standardized raptor code is only around 40%, which is significantly lower than any curve shown in Fig. 3.7. This difference comes from the superiority of standardized raptor codes, which includes the use of a high performance pre-code as well as more efficient maximum likelihood (ML) decoding in contrast to iterative decoders used for Fig. 3.7.

Fig. 3.8 shows the PSNR performance of the proposed random interleaved UEP scheme and the EWF scheme for the best-effort QoS formulation described by Problems 3.20 and 3.30. We evaluate transmission of the H.264 SVC coded CIF *Stefan* video sequence [19], where we consider the case of two layers, with the base layer as the first layer, containing $S_1 = 400$ symbols and all enhancement layers as the second layer with $S_2 = 3400$ symbols.

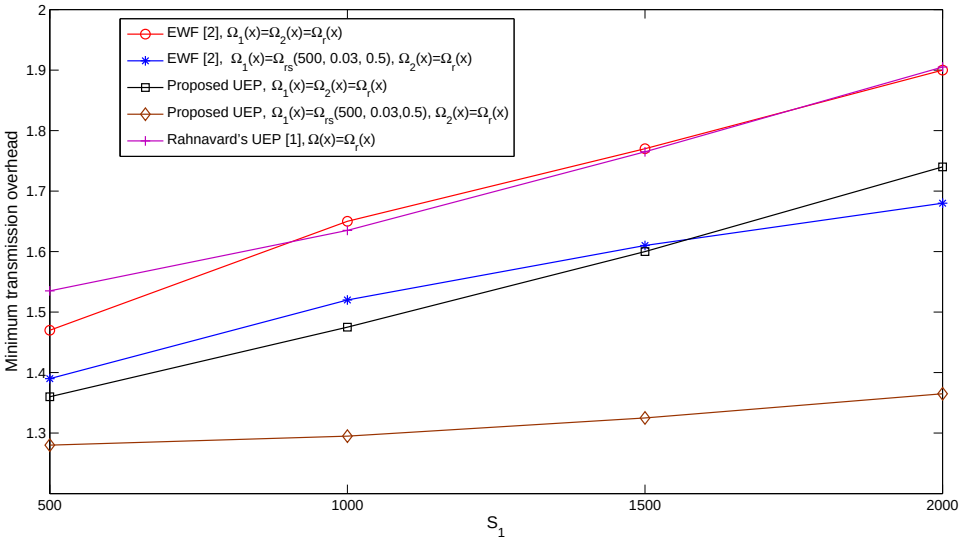

Figure 3.7. Performance comparisons using LT codes with different degree distributions. $L = 2; K = 9000; \delta = [0.4, 0.8]; P = [0.95, 0.8]$.

Successfully decoding the first layer provides a PSNR of 25.79 dB while decoding both the first and second layers provides a PSNR of 40.28 dB. The performance is shown as the average PSNR versus the selection probability ρ_1 for the proposed random interleaved UEP scheme and the first window selection probability Γ_1 of the EWF code. Given ρ_1 or Γ_1, the average PSNR is obtained numerically by setting $\varepsilon = \varepsilon_0$ and substituting the corresponding decoding failure probabilities $p_e(i, j)$ into (3.6) and (3.7). The selection probabilities for the two different schemes have different meanings and are not comparable. However, in order to compare the maximum achievable PSNR of the schemes, all the performance curves are shown in one figure. For the cross-marked and star-marked curves, we have used the LT code with an iterative decoder and degree distribution $\Omega_r(x)$ applied to all windows and

layers. For these parameters, the proposed random interleaved UEP scheme provides a maximum average PSNR of 32.42 dB when optimized while the EWF scheme provides 32.36 dB when optimized. For problem 3.30, the feasible regions of selection probabilities ρ_1 and Γ_1 are obtained by checking the constraints (3.9). We note that for Problem 3.30, the maximum achievable average PSNRs remain the same since both the optimal operating points of the proposed UEP scheme and the EWF code are inside the feasible regions. The diamond-marked curve shows the results when standardized raptor codes are employed for the proposed random interleaved UEP scheme. A maximum average PSNR of 40.28 dB can be achieved for $0.11 \leq \rho_1 \leq 0.18$, which, as expected, is significantly higher than the other two LT coded curves. We can also observe from Fig. 3.8 that different choices of ρ_1 result in significant differences in average PSNR, showing that the optimization process is necessary.

The above results for LT codes with iterative decoding are obtained using "and-or" tree analysis which assume infinite block length. To verify the results in practice, we test the performance of the proposed random interleaved encoder via Monte Carlo simulations. We consider a guaranteed QoS problem with two source layers and two users classes. The number of source symbols in each layer and the reception overhead of each user class are given in the caption of Fig. 3.9. The rateless encoders that are used to encode each layer are both LT codes constructed using degree distribution $\Omega_r(x)$. By using and-or tree analysis via Eq. (3.11) and (3.12) to determine the constraints in Problem 3.10, we obtain the optimal selection probability $\rho_1 = 0.19$ and minimum overhead $\varepsilon_{min} = 1.475$. Using the

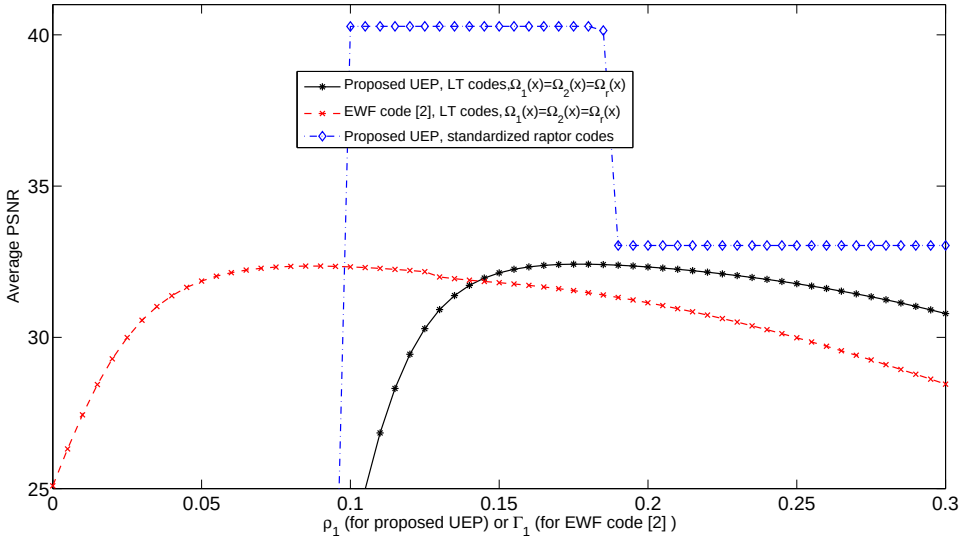

Figure 3.8. Average PSNR performance of two UEP schemes. LT codes with iterative decoding, $L = 2; S = [400, 3400]; \delta = [0.55, 1]; P = [0.95, 0.8]; \varepsilon_0 = 1; w = [0.5, 0.5]$.

optimal selection probability $\rho_1 = 0.19$, we verify the probabilities of successfully achieving the target PSNRs of the two user classes via Monte Carlo simulations for different given transmission overheads in Fig. 3.9. The iterative decoding is performed using the belief propagation (BP) algorithm as in [9] and [11]. The horizontal axis shows the transmission overheads chosen for simulation, which covers the minimum overhead we achieved with our analysis ($\varepsilon_{min} = 1.475$), as well as 5% (1.525) and 10% (1.575) greater than the minimum. The resulting PSNR for each user class is computed for each realization. The vertical axis shows the relative frequency that the PSNR is larger than desired threshold (Prob$(PSNR_j \geq \gamma_j)$) for each user class. Using the minimum overhead $\varepsilon_{min} = 1.475$ obtained from and-or tree analysis, it can be seen from Fig. 3.9 that the simulation results

closely match the theoretical analysis. The probability of reaching the target PSNR γ_j in simulation is the same as or above the desired probability threshold P_j. With an extra 5% overhead at $\varepsilon = 1.525$, higher probabilities in reaching target QoS can be obtained.

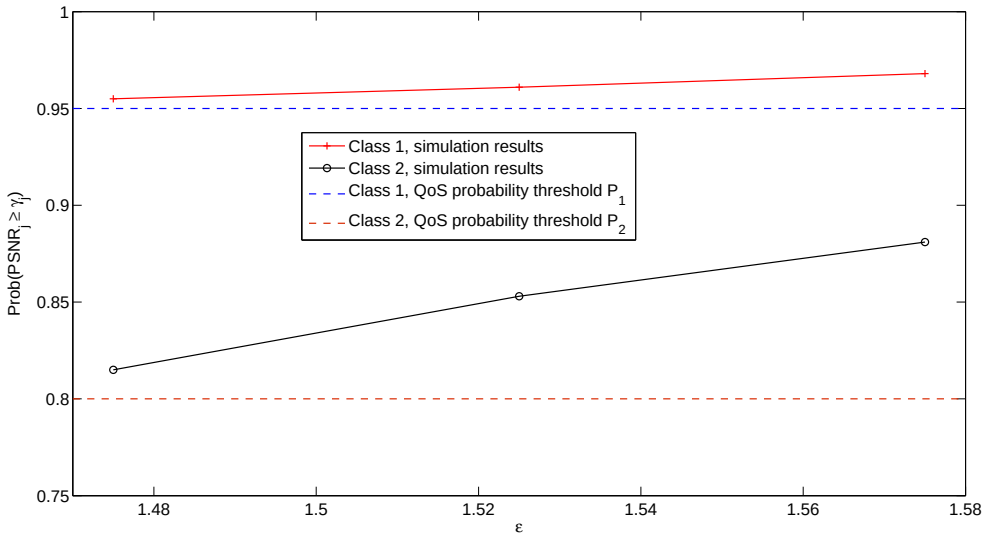

Figure 3.9. Outage probability comparison, simulation results versus desired threshold, $L = 2; S = [1000, 8000]; \delta = [0.4, 0.8]; P = [0.95, 0.8]; \rho_1 = 0.19$.

3.6 Conclusions

Two general problems, guaranteed and best-effort QoS, are formulated for optimizing UEP rateless codes for scalable multimedia multicasting systems with heterogeneous users. A random interleaved UEP raptor code design is proposed. The guaranteed QoS problem is converted into a convex optimization problem, which can be solved analytically in many

practical scenarios. Numerical results show that, for the same system parameters, the minimum transmission overhead required for the optimized proposed UEP rateless codes can be fewer than a quarter of that for an EEP design and more than 10% lower than that for optimized EWF codes [19] and non-uniform-selection UEP rateless codes [1]. Significant additional gains for the proposed UEP scheme can be obtained by employing superior standardized raptor codes. For example, in the best-effort QoS example shown in Fig. 3.8, the maximum achievable average PSNR using the proposed design employing standardized raptor codes is around 8 dB higher than that of the proposed design and EWF codes employing LT codes with an iterative decoder.

Chapter 4

Rateless coded asynchronous multicast

4.1 Introduction

In Chapter 3, optimization of UEP rateless codes for scalable multimedia multicast is investigated. In this chapter, we consider a different scenario where the users can asynchronously tune-in and out of the multicast session at any time. In this asynchronous multicast scenario, users who receive enough transmitted symbols to achieve a desired quality will either leave the multicast session or use the rest of the transmission time frame to perform other valuable tasks. For that reason, the system cost is mainly the time that the users spend to collect the rateless coded symbols to achieve their own QoS requirements, which are proportional to the numbers of transmitted coded symbols required for each user class given a fixed symbol rate. Therefore, for a multicast system with different user classes, the cost function for this asynchronous multicast scenario is an average of transmitted symbols required to achieve the different QoS weighted by the different user classes.

The asynchronous multicast optimization problem for a different UEP rateless code

design is first investigated in [25]. In [25], a priority encoding transmission (PET) based packetization scheme [15] using a rateless code with fixed overhead is proposed. The formulation in [26] generalizes the optimization problem in [25] by adding the constraint of a transmission deadline. However, in solving the optimization problem of asynchronous multimedia multicast, both [25] and [26] use an iterative search method which has high complexity.

In this chapter, we investigate the asynchronous multicast optimization problem and find a systematic and low complexity solution. The PET packetization scheme that is proposed in [15] and [25] are used as opposed to the random interleaved UEP scheme proposed in Chapter 3. We show that under certain assumptions and under the relaxation of integer solution constraints, this problem can be transformed into a convex optimization problem. Furthermore, an analytical solution is found for the asynchronous multicast optimization problem [25] without the outage constraints. For the more general formulation with outage constraints [26], an analytical solution is found for the special case where there are two user classes, and we resort to numerical solution using convex optimization software [27] [28] to solve the more general cases.

The rest of the chapter is organized as follows: Section 4.2 describes the system setup and transmission scheme proposed in [25] and [26]; Section 4.3 presents the optimization problem formulation; Section 4.4 transforms the formulation in Section 4.3 to an equivalent but simplified problem by reducing the number of parameters and then the convexity of the transformed problem is proven; Section 4.5 presents analytical solutions to the convex optimization problem formulated in Section 4.4; Section 4.6 provides numerical results to

verify our findings; Section 4.7 provides the conclusions of this chapter.

4.2 Asynchronous multicast system and packetization scheme

4.2.1 System setup

Similar to the system described in Chapter 3, a server is multicasting a scalable multimedia source, e.g., image or video, to a total of J class of users. We assume the indices of the J classes of users are ordered in terms of their QoS requirements (as opposed to the channel conditions used by Chapter 3), i.e., $\gamma_1 \leq \gamma_2 ... \leq \gamma_J$, where γ_j is the target peak-signal-to-noise-ratio (PSNR) for Class j users. In the following formulation, the PSNR can also be replaced by other fidelity measures without loss of generality. Each user in Class j is assumed to experience a packet erasure channel with erasure rate σ_j, which is assumed to be fixed. For a more general setup where σ_j is a random variable, readers can refer to Appendix C.

In asynchronous systems without transmission deadlines, users can join or leave the multicast session at any time [25]. Nevertheless, if the multimedia content is partitioned into multiple groups of frames (GOF) before multicasting, there is usually a deadline for the transmission of a particular frame due to delay requirements. Therefore, it is assumed as in [26] that the server can transmit at most M_0 packets for all users in each multicast session without causing an outage. It is assumed that all the users access the multicast system at the same time as the start of the transmission. However, the results in this chapter can also apply to the case where users access the multicast system at a random time with

known statistics, a discussion of which can be found in detail in Appendix C.

4.2.2 The priority encoding transmission (PET) packetization scheme

In contrast to Chapter 3 that use a random interleaved UEP scheme, in this chapter, a PET based packetization scheme is used. The PET packetization scheme is first proposed with Reed-Solomon (RS) codes in [15]. Later, [25] proposes a transmission scheme to combine the PET packetization scheme with a rateless code. The source symbols are first loaded into the PET rateless coded packetization structure before transmission. The detailed structure of the PET packetization with rateless coding is shown in Fig. 4.1. The source symbol stream is first partitioned into L vertical layers such that source symbols with a lower layer index are more important than those with a higher layer index, where L is the total number of layers, which is equal to the packet length. Note that the layers here are defined differently from the layers used in Chapter 3. The number of source symbols inside Layer l, $l \in \{1,2,...,L\}$ is denoted by K_l. The K_l source symbols are then encoded by a rateless encoder to produce arbitrary number of rateless encoded symbols $d_{l1}, d_{l2}, ..., d_{li}, ...$, where d_{li} represents the $i-th$ rateless encoded symbol in the $l-th$ layer. Each data packet is formed by selecting one corresponding encoded symbol from each layer, i.e., Packet i consists of symbols $d_{1i}, d_{2i}, d_{3i}, ..., d_{Li}$. The PET packetization design objective is to ensure that the information symbols in a layer with smaller index are better protected against packet loss compared to information symbols in a layer with a larger index. Therefore, the following constraints have to be satisfied:

$$K_1 \leq K_2 \leq \cdots \leq K_L. \tag{4.1}$$

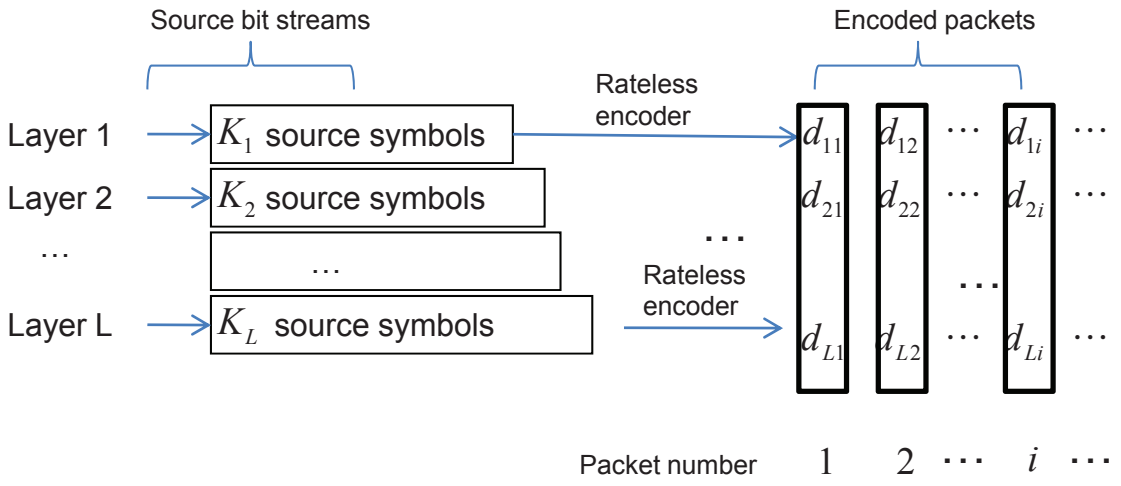

Figure 4.1. The transmission scheme using PET packetization combined with rateless coding.

4.3 The asynchronous multicast optimization problem

4.3.1 Users' QoS requirements

As with the scenario described in Chapter 3, for a given source coder, the PSNR or equivalent fidelity measures are a function of the number of decoded symbols, where the symbols are in decreasing order of importance. The QoS requirement of a Class j user is that the obtained PSNR value should be greater than or equal to QoS threshold γ_j, which can be expressed as:

$$PSNR_j(K_1, K_2, ..., K_L) \geq \gamma_j, \tag{4.2}$$

where $PSNR_j(K_1, K_2, ..., K_L)$ represents the PSNR of Class j users given the allocated set of $K_1, K_2, ..., K_L$. Let $f(.)$ represent the PSNR value as a non-decreasing function of the total number of symbols decoded by the receiver, h_j denote the number of layers (starting from the first layer) that a Class j user needs to decode to achieve the target PSNR threshold γ_j for a given set $\{K_1, K_2, ..., K_L\}$. Then the QoS requirements can be expressed as:

$$f(\sum_{l=1}^{h_j} K_l) \geq \gamma_j \tag{4.3}$$

or, due to the monotone one to one function $f(.)$, as

$$\sum_{l=1}^{h_j} K_l \geq f^{-1}(\gamma_j). \tag{4.4}$$

where f^{-1} denotes the inverse function of f. Note that unlike the QoS constraints described in Chapter 3, there is no outage probability involved in the QoS constraints. Rather, in this asynchronous multicast scenario, the receiver keeps collecting rateless coded packets until its QoS requirement is met. In the case where the receiver cannot reach its own QoS requirement even if the maximum number of packets M_0 has been transmitted by the server, an outage occurs. However, this outage probability considered as a separate type of outage constraint as described in the following subsection.

4.3.2 Outage probabilities and constraints

In this section, we briefly describe the outage probability constraints. It is assumed that in order for the receiver to decode Layer l with K_l information symbols, the receiver needs to collect at least $\lceil K_l(1+\omega) \rceil$ encoded symbols, where $\lceil x \rceil$ denotes the smallest integer that is larger than x and ω is the overhead of the rateless code, which is assumed to be

fixed for different realizations. For a Class j user with erasure rate σ_j, the failure probability of decoding Layer h_j is equivalent to the situation where the receiver collects fewer than $\lceil K_{h_j}(1+\omega)\rceil$ encoded packets after M_0 packets have been transmitted, which can be expressed as:

$$O_j = \sum_{i=0}^{\lceil K_{h_j}(1+\omega)\rceil-1} \begin{pmatrix} M_0 \\ i \end{pmatrix} \left(1-\sigma_j\right)^i \sigma_j^{M_0-i}, \tag{4.5}$$

where O_j is the outage probability of Class j users. Users in different classes may have different outage tolerances. Therefore, the outage constraints can be expressed as:

$$O_j \leq \tau_j, \qquad j = 1,2,...,J, \tag{4.6}$$

where τ_j is the outage probability threshold for Class j users. Note that we have used the assumption that the outage probability is mainly limited by the decoding failure probability of Layer h_j. This assumption is reasonable as the performance of a well designed rateless code has very steep error curves. Therefore, given the constraint of (4.1), the probability that a user fails to decode Layer l where $l < h_j$ given that Layer h_j is successfully decoded can be neglected compared to the probability of failure in decoding Layer h_j.

4.3.3 Cost function

In the asynchronous multicast problem, the outage probability is mainly bottle-necked by users that require the largest number of transmitted packets to meet their QoS constraints. As described in Section 4.1, a reasonable objective is to minimize the average time that all users spend collecting rateless coded packets until their QoS requirements have been

met, which is proportional to the average number of transmitted packets required to meet each user's QoS constraint, or equivalently, to successfully decode Layer h_j. If there are no limits on the maximum number of transmitted packets, the average number of transmitted packets required to meet the QoS constraint of a Class j user is given by:

$$E(M_j) = \frac{K_{h_j}(1+\omega)}{1-\sigma_j}, \tag{4.7}$$

where M_j is the number of transmitted packets required to meet Class j user's QoS requirement, $E(.)$ denotes expectation, h_j is the number of decoded layers required to achieve Class j users' QoS requirements and σ_j is the erasure rate of Class j users. For analytical simplicity, the ceiling operator on $K_{h_j}(1+\omega)$ is omitted in the cost function. Eq. (4.7) can be intuitively interpreted as having the additional compensation factor $\frac{1}{1-\sigma_j}$ in order to compensate a packet loss rate of σ_j compared to that of no packet loss. For a more detailed justification of (4.7), users can refer to Appendix C. When there is a limit M_0 on the maximum number of transmitted packets, Eq. (4.7) still serves as an upper bound on the average number of required transmitted symbols for all the users. As the probability that a user is in outage is usually well within the outage constraints, Eq. (4.7) is still an appropriate cost function for this PET rateless code design with a transmission deadline.

The total overall cost function averaged over all the user classes is thus given by

$$M_{av} = \sum_{j=1}^{J} w_j E(M_j), \tag{4.8}$$

where w_j is a given weighting coefficient for Class j users which takes the number of users and the relative significance of users in each user class into account, where for $1, 2, ..., J$, $0 \leq w_j \leq 1$ and $\sum_1^J w_j = 1$.

4.3.4 Optimization problem formulation

In summary, the asynchronous multicast optimization problem is to find the best non-decreasing set $\{K_1, K_2, ..., K_J\}$ to minimize Eq. (4.8), such that users' target PSNR is met, under the outage constraints of each individual user class. This problem can be summarized as:

Problem 4.10:

$$\min_{K_1,...,K_L} \quad M_{av} \tag{4.9}$$

subject to

$$K_1 \leq K_2 \leq \cdots \leq K_L \tag{4.10}$$

and

$$O_j \leq \tau_j, \qquad j = 1, 2, ..., J. \tag{4.11}$$

On the other hand, if there is no transmission deadline as the scenario discussed in [25], the problem can be formulated as

Problem 4.20:

$$\min_{K_1,...,K_L} \quad M_{av} \tag{4.12}$$

subject to

$$K_1 \leq K_2 \leq \cdots \leq K_L. \tag{4.13}$$

We will later determine an analytical solution for Problem 4.20.

4.3.5 Greedy search algorithm

A direct way to solve Problems 4.10 and 4.20 is to list all possible sets of source symbol layer allocations $K_1, K_2, ..., K_L$, test the constraints and compare the resultant cost. The optimal solution is the set $\{K_1, K_2, ..., K_L\}$ which provides the minimal cost among those that satisfy the constraints. However, the complexity of such a brute-force search algorithm can be prohibitively high as it depends on the total number of source symbols $f^{-1}(\gamma_J)$ as well as L. A suboptimal greedy search algorithm is proposed in [25] and [26]. The basic idea of this algorithm is to first find all possible sets of $[h_1, h_2, ..., h_J]$, which represents the number of layers required to decode for each user class. For each predetermined allocation of $[h_1, h_2, ..., h_J]$, the greedy search algorithm allocates $\lceil f^{-1}(\gamma_j) \rceil - \lceil f^{-1}(\gamma_{j-1}) \rceil$ source symbols to layers $h_{j-1}+1$ to h_j *nearly equally*, where $\lceil f^{-1}(\gamma_0) \rceil = 0$. By *nearly equally* we mean the difference among the number of source symbols in Layers $h_{j-1}+1$ to h_j, i.e., $K_{h_{j-1}+1}, K_{h_{j-1}+2}, ..., K_{h_j}$, should be at most 1. In the end, the best possible allocation is chosen by selecting the one with minimal cost that satisfies the outage constraints.

The greedy search algorithm proposed in [25] has a complexity of roughly $JL^{J-1}/2$ elementary operations. Although the complexity is lower compared to brute force search over all possible sets of $\{K_1, K_2, ..., K_L\}$, it is still very high for large L and J. In addition, the optimality of the greedy search algorithm has not been proven. This motivates a lower complexity and analytical solution to this problem.

4.4 Transformation to equivalent and simplified convex optimization problem

4.4.1 Grouping layers to chucks

To find an analytical solution to Problems 4.10 and 4.20, we first divide the source symbols in each block into J chunks, where the number of symbols in each chunk is given by

$$U_j = \begin{cases} \lceil f^{-1}(\gamma_j) \rceil - \lceil f^{-1}(\gamma_{j-1}) \rceil & j = 2, ..., J \\ \lceil f^{-1}(\gamma_1) \rceil & j = 1, \end{cases} \tag{4.14}$$

where γ_j is the target PSNR threshold for Class j users and $f^{-1}(x)$ is the minimum number of decoded symbols required to reach the target PSNR threshold x. Define

$$l_j = h_j - h_{j-1} \qquad j = 2, ..., J \tag{4.15}$$

where h_j is the number of required decoded layers to reach target QoS of Class j users and $h_0 = 0$. Therefore, l_j represents the number of layers allocated to Chunk j, which should satisfy

$$\sum_{j=1}^{J} l_j = h_J = L. \tag{4.16}$$

In addition, we have

$$\sum_{i=h_{j-1}+1}^{h_j} K_i = U_j \qquad j = 1, 2, ..., J. \tag{4.17}$$

4.4.2 Simplification of the cost function

Substituting Eq. (4.7) into (4.8), the cost function can be expressed as

$$\begin{aligned} M_{av} &= \sum_{j=1}^{J} w_j E(M_j) \\ &= \sum_{j=1}^{J} w_j \frac{K_{h_j}(1+\omega)}{1-\sigma_j} \\ &= \sum_{j=1}^{J} \eta_j K_{h_j}, \end{aligned} \tag{4.18}$$

where $\eta_j \equiv \frac{w_j(1+\omega)}{1-\sigma_j}$ are the combined cost coefficients of Class j users, which combine the importance weighting coefficient w_j, code-efficiency factor $(1+\omega)$ and reception capability $1/(1-\sigma_j)$.

4.4.3 Reducing the number of parameters

Problem 4.10 contains L variables. In order to obtain an equivalent problem with fewer variables, we first prove the following fact.

Lemma 4.1 : If we relax the integer constraints on $K_i, i = 1,2,...,L$, then the optimal solution to Problem 4.10 should satisfy $K_{h_{j-1}+1} = K_{h_{j-1}+2} = ... = K_{h_j}$ for $j = 1,2,...,J$, i.e., the U_j source symbols of Chunk j should be allocated equally among layers $h_{j-1}+1$ to h_j.

To prove this, assume there exists allocation Scheme C, where there is at least one chunk (index denoted by j) within which the U_j source symbols are not equally allocated among layers $h_{j-1}+1$ to h_j. Then, because K_j is a non-decreasing set and $\sum_{i=h_{j-1}+1}^{h_j} K_i = U_j$, we have, for allocation Scheme C, $K^C_{h_j} > U_j/(h_j - h_{j-1}) = U_j/l_j$, where $K^C_l, l = 1,2,...,L$ represents the source symbol to layer allocation for Scheme C. We next construct allocation

Scheme D where $K^D_{h_{j-1}+1} = K^D_{h_{j-1}+2} = ... = K^D_{h_j} = U_j/l_j$ with all the other chunks allocated identically to Scheme C, where $K^D_l, l = 1,2,...,L$ represents the source symbols to layer allocation for Scheme D. Because $K^C_{h_j} > U_j/l_j = K^D_{h_j}$ and $K^C_{h_i} = K^D_{h_i}$ for all $i \neq j$, it is obvious from the cost function expression (4.18), $M^C_{av} = \sum_{i=1}^{J} \eta_i K^C_{h_i} > \sum_{i=1}^{J} \eta_i K^D_{h_i} = M^D_{av}$, where M^C_{av} and M^D_{av} are the overall costs of Scheme C and Scheme D, respectively. Hence, allocation Scheme C does not minimize the cost function. Therefore, an optimal solution that minimizes M_{av} should allocate U_j symbols among layers $h_{j-1}+1$ to h_j equally for $j = 1,2,..,J$, i.e., $K_1 = K_2 = ... = K_{h_1} \leq K_{h_1+1} = ... = K_{h_2} \leq ... \leq K_{h_{J-1}+1} = ... = K_{h_J} = K_L$.

QED

Therefore, we can reduce the number of parameters by using one parameter K_{h_j} to represent the $K_l, h_{j-1}+1 \leq l \leq h_j$ values of Chunk j, i.e., $K_{h_{j-1}+1} = K_{h_{j-1}+2} = ... = K_{h_j} = U_j/l_j$. Using the above, the cost function can be further simplified as

$$M_{av} = \sum_{j=1}^{J} \eta_j K_{h_j} = \sum_{j=1}^{J} \eta_j U_j/l_j = \sum_{j=1}^{J} \alpha_j/l_j, \tag{4.19}$$

where

$$\alpha_j \equiv \eta_j U_j = \frac{w_j U_j(1+\omega)}{1-\sigma_j}. \tag{4.20}$$

Also, using the constraint that K_j is a non-decreasing set, Eq. (4.1), can be transformed to $K_{h_1} \leq K_{h_2} ... \leq K_{h_j} = K_J$ or $U_1/l_1 \leq U_2/l_2 ... \leq U_J/l_J$.

In addition, observe that since the average outage probability (4.5) is monotonically increasing with K_{h_j}, the constraints are equivalent to K_{h_j} being less than a threshold, or alternatively, equivalent to l_j being greater than a constant for $j = 1,2,...,J$.

4.4.4 Transformation to a convex optimization problem

In summary, if we relax the integer constraints on $K_i, i = 1,2,...,L$, the allocation problem described by Problem 4.10 can be transformed to the following equivalent problem,

Problem 4.11:

$$\min_{l_1,...,l_J} \quad \sum_{i=1}^{J} \alpha_i / l_i \tag{4.21}$$

subject to

$$\sum_{i=1}^{J} l_i = L \tag{4.22}$$

$$0 \leq l_J/U_J \leq l_{J-1}/U_{J-1} ... \leq l_1/U_1, \tag{4.23}$$

$$l_j \geq \theta_j, \qquad j = 1,2,...,J, \tag{4.24}$$

where $\theta_j,\ j = 1,2,...,J$ are constants derived from average outage probability constraints given by Eq. (4.6).

Similar to the transformation of Problem 4.10 to Problem 4.11, Problem 4.20, which does not have outage constraints, can be transformed to:

Problem 4.21:

$$\min_{l_1,...,l_J} \quad \sum_{i=1}^{J} \alpha_i / l_i \tag{4.25}$$

subject to

$$\sum_{i=1}^{J} l_i = L \tag{4.26}$$

$$0 \leq l_J/U_J \leq l_{J-1}/U_{J-1} ... \leq l_1/U_1. \tag{4.27}$$

Note that in Problem 4.11 and 4.21, we have assumed that $l_i, i = 1, 2, .., J$ are continuous variables rather than integers. In the following section, we solve alternative tractable versions of problems to 4.11 and 4.21 that are equivalent to the original allocation problems without integer constraints. We next show that Problems 4.11 and 4.21 are both convex optimization problems. This can be proven by showing that the objective function is convex and all the constraint functions are linear.

4.5 Analytical solutions

4.5.1 Dual problem

The standard way to proceed is to simplify Problem 4.11 by solving the dual problem, because the optimal solutions of Problem 4.11 and its dual are the same due to convexity [27]. The Lagrange function

$$L(l,\lambda,v) = \sum_{i=1}^{J} \frac{\alpha_i}{l_i} + v(\sum_{i=1}^{J} l_i - L) + \sum_{i=1}^{J-1} \lambda_i (\frac{l_{i+1}}{U_{i+1}} - \frac{l_i}{U_i}) - \lambda_J \frac{l_J}{U_J} + \sum_{i=1}^{J} \lambda_{i+J}(\theta_i - l_i), \quad (4.28)$$

where $l = (l_1, l_2, ..., l_J)$ are primary variables, $\lambda = (\lambda_1, \lambda_2, ..., \lambda_{2J})$ and v are dual variables. The Lagrange dual function is

$$\begin{aligned} g(\lambda, v) = & \quad \inf_l L(l,\lambda,v) \\ = & \quad \inf_l \left(\sum_{i=1}^{J} \left(\frac{\alpha_i}{l_i} + l_i (v - \frac{\lambda_i}{U_i} + \frac{\lambda_{i-1}}{U_{i-1}} - \lambda_{i+J}) + \lambda_{i+J}\theta_i \right) \right) \end{aligned} \quad (4.29)$$

where $\frac{\lambda_0}{U_0} = 0$ and the dual function have meaningful solutions only if $v - \frac{\lambda_i}{U_i} + \frac{\lambda_{i-1}}{U_{i-1}} - \lambda_{i+J} > 0$ for $i = 1, 2, ..., J$. By solving $\frac{\partial L(l,\lambda,v)}{\partial l_i} = 0$ and substituting the resulting $l_1, l_2, ..., l_J$ back

into $L(l,\lambda,v)$ to obtain $g(\lambda,v) = \inf_l L(l,\lambda,v)$, the dual problem can be written as

$$\max_{v,\lambda} \qquad g(\lambda,v) = \sum_{i=1}^{J} \left(2\sqrt{\alpha_i(v - \frac{\lambda_i}{U_i} + \frac{\lambda_{i-1}}{U_{i-1}} - \lambda_{i+J})} + \lambda_{i+J}\theta_i \right) \tag{4.30}$$

subject to

$$\lambda_j \geq 0, \qquad j = 1,2,...,2J. \tag{4.31}$$

Unfortunately, it is not straightforward to find an analytical solution to the dual problem.

4.5.2 Analytical optimization solution without outage constraints

We next show that the optimal solution of Problem 4.21, the problem without outage constraints, can be obtained analytically. As a first step, we solve Problem 4.21 without the inequality constraints provided by (4.27), which is easily done by using Lagrange's method. The Lagrange L function in this case would be

$$\begin{aligned} L(l,v) &= \sum_{i=1}^{J} \alpha_i/l_i + v(\sum_{i=1}^{J} l_i - L) \\ &= \sum_{i=1}^{J} (\alpha_i/l_i + vl_i) - vL. \end{aligned} \tag{4.32}$$

The optimal solutions l_i for Problem 4.21 without inequality constraints (4.27) minimize $\alpha_i/l_i + vl_i$ for all $i = 1,2,...,J$. By setting the derivative of $\alpha_i/l_i + vl_i$ with respect to l_i to zero, the optimal l_i can be obtained as

$$l_i = \frac{L\sqrt{\alpha_i}}{\sum_{i=1}^{J}\sqrt{\alpha_i}}, \qquad j = 1,2,...,J. \tag{4.33}$$

Now if the above solution satisfies the inequality constraints provided by (4.27), then we are done. In order to solve the problem when (4.33) does not satisfy inequality constraints (4.27), we deploy a similar method that used in [50]. First, we prove the following fact:

Lemma 4.2: For any $j \in (1,2,...,J-1)$, if $\frac{\sqrt{\alpha_{j+1}}}{U_{j+1}} > \frac{\sqrt{\alpha_j}}{U_j}$, then the optimal solution of Problem 4.21 satisfies $\frac{l_j}{U_j} = \frac{l_{j+1}}{U_{j+1}}$.

Proof: Assume $\frac{\sqrt{\alpha_{j+1}}}{U_{j+1}} > \frac{\sqrt{\alpha_j}}{U_j}$, $j \in (1,2,...,J-1)$ and that there exists a set of optimal solutions $X = \{l_i^X, i = 1,2,...,J, \frac{l_j^X}{U_j} \neq \frac{l_{j+1}^X}{U_{j+1}}\}$. Now we create a new set of solutions $Y = \{l_i^Y, i = 1,2,...,J\}$ by perturbing l_j^X and l_{j+1}^x by a very small amount $\Delta l > 0$ such that the constraints (4.23) are still valid, i.e., for $i = 1,2,..,J$,

$$l_i^Y = \begin{cases} l_i^X - \triangle\, l & \text{if } i = j \\ l_i^X + \triangle\, l & \text{if } i = j+1 \\ l_i^X & \text{if } i \neq j, j+1. \end{cases} \tag{4.34}$$

Now the difference in the cost functions due to (4.36) is

$$\begin{aligned} M_{av}^Y - M_{av}^X &= \frac{\alpha_j}{l_j^X - \Delta l} + \frac{\alpha_{j+1}}{l_{j+1}^X + \Delta l} - \frac{\alpha_j}{l_j^X} - \frac{\alpha_{j+1}}{l_{j+1}^X} \\ &= \Delta l \left(\frac{\alpha_j}{l_j^X (l_j^X - \Delta l)} - \frac{\alpha_{j+1}}{l_{j+1}^X (l_{j+1}^X + \Delta l)} \right). \end{aligned} \tag{4.35}$$

As $\frac{\sqrt{\alpha_{j+1}}}{U_{j+1}} > \frac{\sqrt{\alpha_j}}{U_j}$ and $\frac{l_j^X}{U_j} > \frac{l_{j+1}^X}{U_{j+1}}$, we have

$$\frac{\alpha_{j+1}}{\alpha_j} > \frac{U_{j+1}^2}{U_j^2} > \frac{(l_{j+1}^X)^2}{(l_j^X)^2}. \tag{4.36}$$

Therefore, if $\Delta l > 0$ is chosen to be small enough, then we certainly have $\frac{\alpha_{j+1}}{\alpha_j} > \frac{l_{j+1}^X (l_{j+1}^X + \Delta l)}{l_j^X (l_j^X - \Delta l)}$. Hence $M_{av}^Y - M_{av}^X < 0$, which contradicts the assumption that the set of solutions X is optimum. Therefore, the optimal solution must satisfy $\frac{l_j}{U_j} = \frac{l_{j+1}}{U_{j+1}}$ if $\frac{\sqrt{\alpha_{j+1}}}{U_{j+1}} > \frac{\sqrt{\alpha_j}}{U_j}$ for any $j = 1,2,...,J-1$. **QED**

With the above lemma, if $\frac{\sqrt{\alpha_{j+1}}}{U_{j+1}} > \frac{\sqrt{\alpha_j}}{U_j}$, we have $K_{h_j} = \frac{U_j}{l_j} = \frac{U_{j+1}}{l_{j+1}} = K_{h_{j+1}}$. Therefore, we can reduce the number of parameters further by grouping Chunk j and Chunk $j+1$ into a

new chunk, i.e, $U_j^{new} = U_j + U_{j+1}$, $l_j^{new} = l_j + l_{j+1}$, $\alpha_j^{new} = \left(\alpha_j/l_j + \alpha_{j+1}/l_{j+1}\right)\left(l_j + l_{j+1}\right) = \left(\alpha_j/U_j + \alpha_{j+1}/U_{j+1}\right)\left(U_j + U_{j+1}\right)$, $J_{new} = J-1$ and decrease the indices of all parameters with index larger than $j+1$ by one. This procedure is repeated until no $j \in \{1,2,...,J-1\}$ with $\frac{\sqrt{\alpha_{j+1}}}{U_{j+1}} > \frac{\sqrt{\alpha_j}}{U_j}$ exists, i.e., $\frac{\sqrt{\alpha_{j+1}}}{U_{j+1}} \leq \frac{\sqrt{\alpha_j}}{U_j}$ for all $j = 1,2,...,J-1$. Denote $\hat{x}$ as the new value of variable x after the chunk grouping process, Problem 4.21 is then transformed to:

Problem 4.22:

$$\min_{\hat{l}_1,...,\hat{l}_{\hat{J}}} \quad \sum_{i=1}^{\hat{J}} \hat{\alpha}_i/\hat{l}_i \tag{4.37}$$

subject to

$$\sum_{i=1}^{\hat{J}} \hat{l}_i = L \tag{4.38}$$

$$0 \leq \hat{l}_{\hat{J}}/\hat{U}_{\hat{J}} \leq \hat{l}_{\hat{J}-1}/\hat{U}_{\hat{J}-1} ... \leq \hat{l}_1/\hat{U}_1, \tag{4.39}$$

where $\frac{\sqrt{\hat{\alpha}_j}}{\hat{U}_j}$ is a non-increasing set for $j = 1,2,...,\hat{J}$. Following the same approach as in (4.33), the optimal solution without constraint (4.39) is:

$$\hat{l}_j^0 = \frac{L\sqrt{\hat{\alpha}_j}}{\sum_{i=1}^{\hat{J}} \sqrt{\hat{\alpha}_i}}, \qquad j = 1,2,...,\hat{J}. \tag{4.40}$$

Since $\frac{\hat{l}_j^0}{\hat{U}_j} = \frac{\sqrt{\hat{\alpha}_j}}{\hat{U}_j}\frac{L}{\sum_{i=1}^{\hat{J}}\sqrt{\hat{\alpha}_i}} \geq \frac{\sqrt{\hat{\alpha}_{j+1}}}{\hat{U}_{j+1}}\frac{L}{\sum_{i=1}^{\hat{J}}\sqrt{\hat{\alpha}_i}} = \frac{\hat{l}_{j+1}^0}{\hat{U}_{j+1}}$, the inequality constraints (4.39) are satisfied. Therefore, the solutions $\hat{l}_j^0, j = 1,2,...,\hat{J}$ given by (4.40) form an optimal set of solutions for Problem 4.22. We have thus found an analytical solution to Problem 4.21, which is also the problem described in [25].

The analytical solution only requires the computation of J coefficients $\frac{\sqrt{\alpha_j}}{U_j}$, and perform at most $J-1$ chunk grouping processes, and compute J variables according to (4.40). In

the worst case, the analytical solution requires roughly $14J$ elementary operations, which is linear in J and independent of L. The complexity of the analytical solution is significantly lower than that of the greedy search algorithm proposed in [25], which requires roughly $JL^{J-1}/2$ elementary operations. For example, if the packet size $L = 1000$, with $J = 4$ user classes, the greedy search algorithm requires on the order of 10^9 elementary operations, while the analytical solution requires less than 100 elementary operations.

Given the solution of Problem 4.21, we now proceed to find the solution of the more general problem with the outage constraints, i.e., Problem 4.11. First, if the outage thresholds satisfy $\sum_{i=1}^{J} \theta_j > L$, then there is no feasible solution for Problem 4.11. This corresponds to the case when the transmission deadline M_0 is too small for the outage probability constraints to be met. In the case when $\sum_{i=1}^{J} \theta_j = L$, the problem is trivial as there is only one feasible solution $l_j = \theta_j$ for $j = 1, 2, .., J$. On the other hand, if the optimal solution we found for Problem 4.21 satisfies all the outage probability constraints described by (4.24), then the optimal solution of Problem 4.21 is also the optimal solution of Problem 4.11. This corresponds to the case when M_0 is large enough such that all the outage constraints are loose and do not play any role in the optimization problem. When $\sum_{i=1}^{J} \theta_j < L$ and at least one outage probability constraint is active, solving the problem is not straightforward in general. However, given that the optimization problem is convex with linear constraints, we resort to standard convex optimization software to numerically solve the problem, e.g., `CVX` [28] [27]. Because Problem 4.11 has many fewer parameters than the formulation in [26], the complexity of the numerical method is much simpler.

4.5.3 Analytical solution for two user classes

For the special case where there are only two classes of users, we have the following simplification to the solution of Problem 4.11. Assume the set of optimal solutions for Problem 4.21 obtained using the method described above is given by $l_1 = l_1^0$ and $l_2 = l_2^0$ where $l_1^0 + l_2^0 = L$ and that the outage constraints can be simplified to $l_1 \geq \theta_1$ and $l_2 \geq \theta_2$. The optimal solution set for Problem 4.11 is given by:

$$(l_1, l_2) = \begin{cases} (l_1^0, l_2^0) & \text{if } \theta_1 \leq l_1^0 \text{ and } \theta_2 \leq l_2^0 \\ \text{No feasible solution} & \text{if } \theta_1 + \theta_2 > L \\ (\theta_1, L - \theta_1) & \text{if } \theta_1 > l_1^0 \text{ and } \theta_2 \leq l_2^0 \\ (L - \theta_2, \theta_2) & \text{if } \theta_1 \leq l_1^0 \text{ and } \theta_2 > l_2^0. \end{cases} \tag{4.41}$$

The first two cases of Eq. (4.41) follow straightforwardly from the previous discussion in Section 4.5.2. The last two cases in Eq. (4.41) can be proven using *complementary slackness*. As shown before, Problem 4.11 is a convex optimization problem. Given that $\sum_{i=1}^{J} \theta_j < L$ and there are points that are strictly feasible, Slaters constraint qualification condition holds [27]. Therefore, according to Slater's Theorem, strong duality holds for Problem 4.11 and its dual problem described by (4.30) and (4.31) [27]. Strong duality means the optimal duality gap is zero. Also, complementary slackness holds if strong duality holds [27]. The condition of complementary slackness [1] means that the *i*-th optimal Lagrange multiplier λ_i is zero unless the *i*-th constraint is active at the optimum, which in

[1] Complementary slackness is also part of the KKT optimality conditions, which hold true if strong duality holds.

turn means that the optimal solution for the problem is the same as the optimal solution without this constraint. For Case 3 of (4.41), namely where $\theta_1 > l_1^0$ and $\theta_2 \leq l_2^0$, if the optimal Lagrange multiplier for the constraint $l_1 \geq \theta_1$ is zero, then the optimal solution should be the same as the case without the constraint $l_1 \geq \theta_1$, which is given by $(l_1, l_2) = (l_1^0, l_2^0)$. This solution, however, is not feasible because it does not satisfy the constraints. Therefore, for Case 3, we conclude that constraint $l_1 \geq \theta_1$ is active, i.e., $l_1 = \theta_1$. Similarly, we can prove that the solution for the case $\theta_1 \leq l_1^0$ and $\theta_2 > l_2^0$ of (4.41) is optimal.

Finally, assume that we have reached a set of optimal solutions l_i^0, $i = 1, 2, ..., J$ for Problem 4.11 or 4.21. We can proceed to find a close-to-optimal solution for the original problem with integer constraints: $l_1 = \lfloor l_1^0 \rceil$ and $l_j = \lfloor \sum_{i=1}^{j} l_i^0 - \sum_{i=1}^{j-1} l_i \rceil$ for $j = 2, ..., J$,, where $\lfloor x \rceil$ is the nearest integer to x. Each chunk of U_j source symbols is allocated nearly equally in the sense that there is maximum of one symbol difference among $K_{h_j-1}, ...$ and K_{h_j}.

4.6 Numerical results

To illustrate the above process of solving Problem 4.10, the following numerical example is considered. The same parameters as the simulation setup in [25] is used for verification purpose. In the setup, the server is multicasting a scalable image or video sequence to two user classes. The total number of layers in the packetization scheme is fixed at $L = 47$. The two user classes have a target QoS (PSNR) $\gamma_1 = 27$ dB and $\gamma_2 = 30$ dB,

respectively. From the PSNR curve of the image, the minimum number of source symbols needed to provide the corresponding QoS requirements are $\lceil f^{-1}(\gamma_1)\rceil = 11072$ and $\lceil f^{-1}(\gamma_2)\rceil = 24728$, respectively [25]. Therefore the chunk sizes $U_1 = \lceil f^{-1}(\gamma_1)\rceil = 11072$ and $U_2 = \lceil f^{-1}(\gamma_2)\rceil - \lceil f^{-1}(\gamma_1)\rceil = 13656$. For each user class, the channel erasure rate is $\sigma_1 = \sigma_2 = 0.0549$. [2] The overhead of the rateless code is assumed to be $\omega = 5\%$. The weighting coefficient for Class 1 users w_1 can be varied to obtain different results.

In order to find the best allocation scheme based on the packetization structure for Problem 4.20, the following relevant coefficients are first computed. From (4.20), $\alpha_j = \eta_j U_j = \frac{U_j w_j(1+\omega)}{1-\sigma_j} = 1.111 U_j w_j$ for $j = 1,2$ and weighting coefficient $w_2 = 1 - w_1$. We now determine a sufficient condition for the optimal solution to satisfy $K_{h_1} = K_{h_2}$. According to Lemma 4.2, if $\frac{\sqrt{\alpha_2}}{U_2} > \frac{\sqrt{\alpha_1}}{U_1}$, the optimal solution of Problem 4.21 should satisfy $\frac{l_1}{U_1} = \frac{l_2}{U_2}$. The above condition is equivalent to $\frac{\sqrt{w_2 U_2(1+\omega)/(1-\sigma_2)}}{U_2} > \frac{\sqrt{w_1 U_1(1+\omega)/(1-\sigma_1)}}{U_1}$. Since $\sigma_1 = \sigma_2$ and $w_2 = 1 - w_1$, this condition is equivalent to $\frac{1-w_1}{w_1} > \frac{U_2}{U_1}$ or $w_1 < 0.448$. Therefore, for $w_1 < 0.448$, the optimal solution is to allocate the number of source symbols equally among the $L = 47$ layers, i.e., the optimal allocation scheme is an equal error protection (EEP) scheme. Hence if $w_1 < 0.448$, each layer should have $\lceil f^{-1}(\gamma_2)\rceil / L = 24728/47 = 526.13$ source symbols, To satisfy the integer constraints, we should then allocate 526 source symbols for the first $47 \times 527 - 24728 = 41$ layers and allocate 527 symbols for the last 6 layers. Generalizing from the above process, we uncover the following interesting corollary as a special case of Lemma 4.2:

[2] The channel erasure rates are chosen such that they are equivalent to the parameterizations used in [25] in the sense that they satisfy Eq. (C.7).

Corollary 4.2: In Problem 4.21, if there are only two user classes with the same channel conditions and the weighting coefficient of class 1 users satisfies $w_1 < U_1/(U_1 + U_2)$, then the optimal allocation scheme is an EEP scheme.

Now if $w_1 > 0.448$, the problem is transformed to Problem 4.22 and the optimal solution is given by (4.40), i.e., $l_1^0 = \frac{L\sqrt{\alpha_1}}{\sum_{i=1}^{2}\sqrt{\alpha_i}} = L\frac{\sqrt{w_1 U_1}}{\sqrt{w_1 U_1}+\sqrt{w_2 U_2}}$. Therefore, if $w_1 = 0.6$, then $l_1^0 = 24.65$. With integer constraints, we have $l_1 = \lfloor l_1^0 \rceil = 25$, $l_2 = 22$. To allocate $U_1 = 11072$ symbols nearly equally among the first 25 layers, we obtain $K_1 = K_2 = K_3 = 442$ and $K_4 = K_5 = ... = K_{25} = 443$. Similarly, to allocate $U_2 = 13656$ symbols among the remaining 22 layers, we obtain $K_{26} = K_{27} = ... = K_{31} = 620$ and $K_{32} = K_{33} = ... = K_{47} = 621$. If $w_1 = 0.8$, then $l_1^0 = L\frac{\sqrt{w_1 U_1}}{\sqrt{w_1 U_1}+\sqrt{w_2 U_2}} = 30.22$ and $l_1 = \lfloor l_1^0 \rceil = 30$. If $w_1 = 1$, then $l_1 = L = 47$. Therefore, the optimal solution is to allocate $U_1 = 11072$ symbols among all 47 layers. The number of source symbols allocated for each layer when $w_1 = 0.4, 0.6, 0.8$ and 1 are shown in Figs. 4.2 to 4.5, respectively. By comparing Figs. 4.2 to 4.5 to Fig. 2 of [25] obtained by the greedy search method, we can see that all the results match each other accurately. Readers can refer to [25] for comparisons of the rateless-UEP scheme used in this paper to the UEP scheme in [15] using Monte Carlo simulations, where the advantages of rateless codes over fixed-rate RS codes are clearly demonstrated.

Now assume there are outage constraints $l_1 \geq 25$ and $l_2 \geq 17$, i.e, $\theta_1 = 25$ and $\theta_2 = 17$. The optimal solution for Problem 4.11 can thus be computed according to Eq. (4.41). When $w_1 \leq 0.6$, we always have $\theta_1 > l_1^0$ and $\theta_2 \leq l_2^0$. Therefore, the optimal solution is changed to $l_1 = \theta_1 = 25$, $l_2 = L - \theta_1 = 22$. When $w_1 \geq 0.8$, we always have $\theta_1 \leq l_1^0$ and $\theta_2 \geq l_2^0$, the optimal solution is $l_1 = L - \theta_2 = 30$ and $l_2 = \theta_2 = 17$. When $0.6 < w_1 < 0.8$,

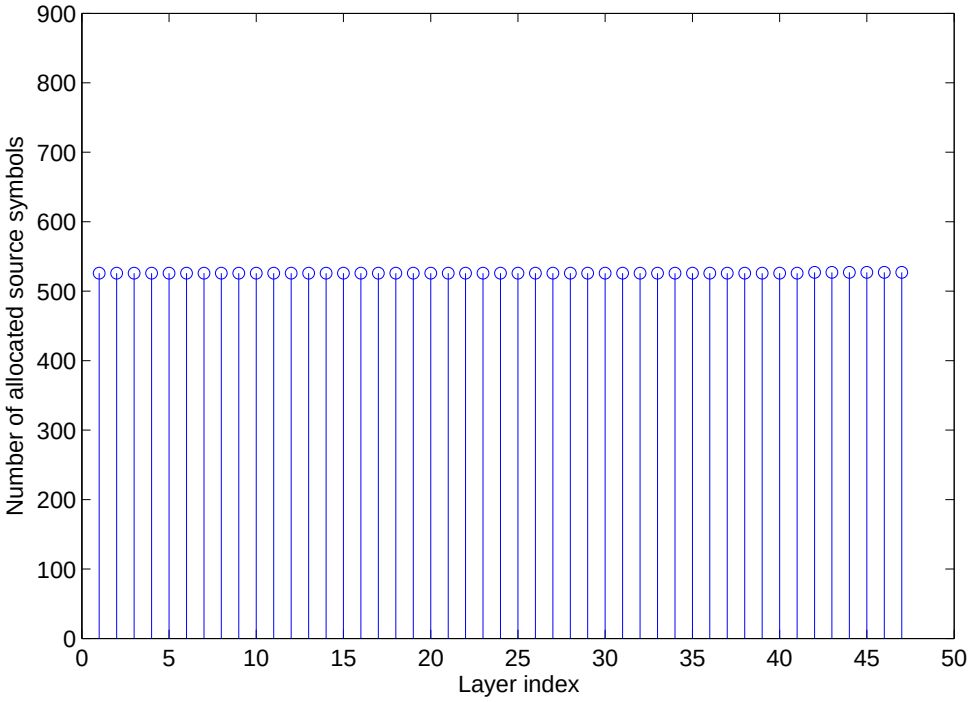

Figure 4.2. Allocation of source symbols for an asynchronous multimedia multicast system with two user classes and no transmission deadline with parameter $w_1 = 0.4$.

we always have $\theta_1 \leq l_1^0$ and $\theta_2 \leq l_2^0$. Therefore, the optimal solution is the same as the solution for Problem 4.21, i.e., $l_1 = l_1^0$ and $l_2 = l_2^0$.

The number of source symbols allocated for each layer for the cases of $0 \leq w_1 \leq 0.6$ and $0.8 \leq w_1 \leq 1$ are shown in Figs. 4.6 and 4.7, respectively. Comparing the results of Fig. 4.6 and 4.7 to Figs. 4.2 and 4.5, the differences due to the outage constraints for the cases of $w_1 = 0.4$ and $w_1 = 1$ can be easily observed.

To demonstrate the solution of Problem 4.10 with more than two classes, we consider an example where the server is multicasting a scalable video stream to four user classes. The number of symbols and the PSNR values of the video sequence are taken from Table

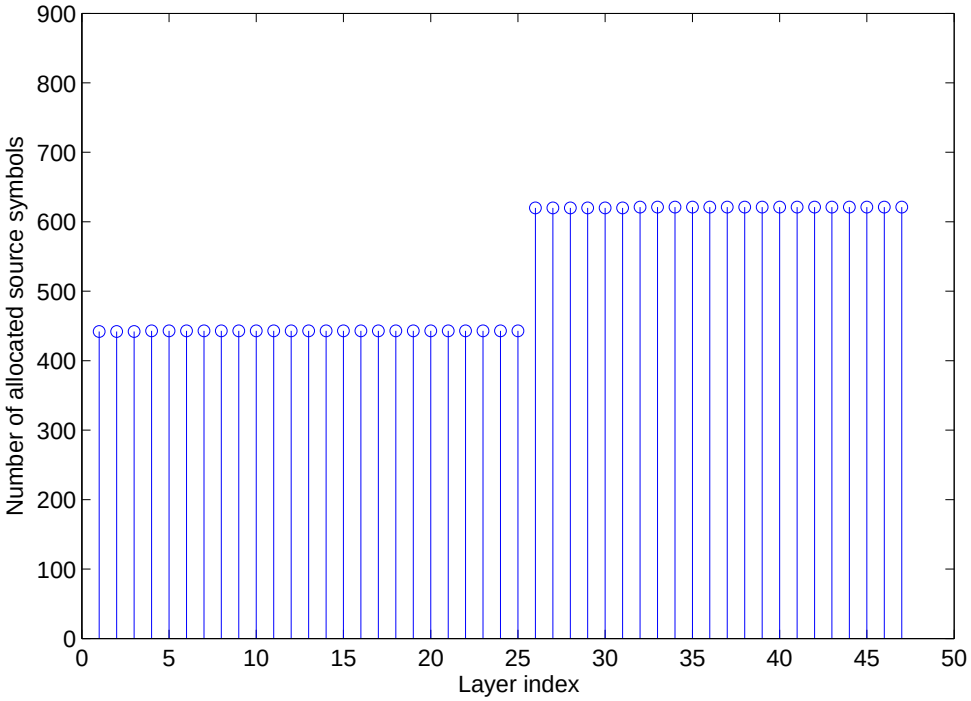

Figure 4.3. Allocation of source symbols for an asynchronous multimedia multicast system with two user classes and no transmission deadline with parameter $w_1 = 0.6$.

I of [19]. The video sequence consists of one base layer (BL) and fourteen enhancement layers (EL) with a total of 3800 symbols where each symbol represents 50 bytes. The channel erasure rates σ_j, PSNR thresholds γ_j and weighting coefficients w_j of the user classes is shown in Table 4.1 . The total number of layers $L = 47$ and the rateless code's overhead is assumed again to be $\omega = 5\%$. The number of symbols required to reach the PSNR threshold $f^{-1}(\gamma_j)$, U_j, $\alpha_j = w_j U_j (1+w)/(1-\sigma_j)$ and $\sqrt{\alpha_j}/U_j$ is computed based on the above given parameters.

First we consider the case where there is no transmission deadline, i.e., no outage constraint. The optimization problem is thus in the form of Problem 4.21. Following

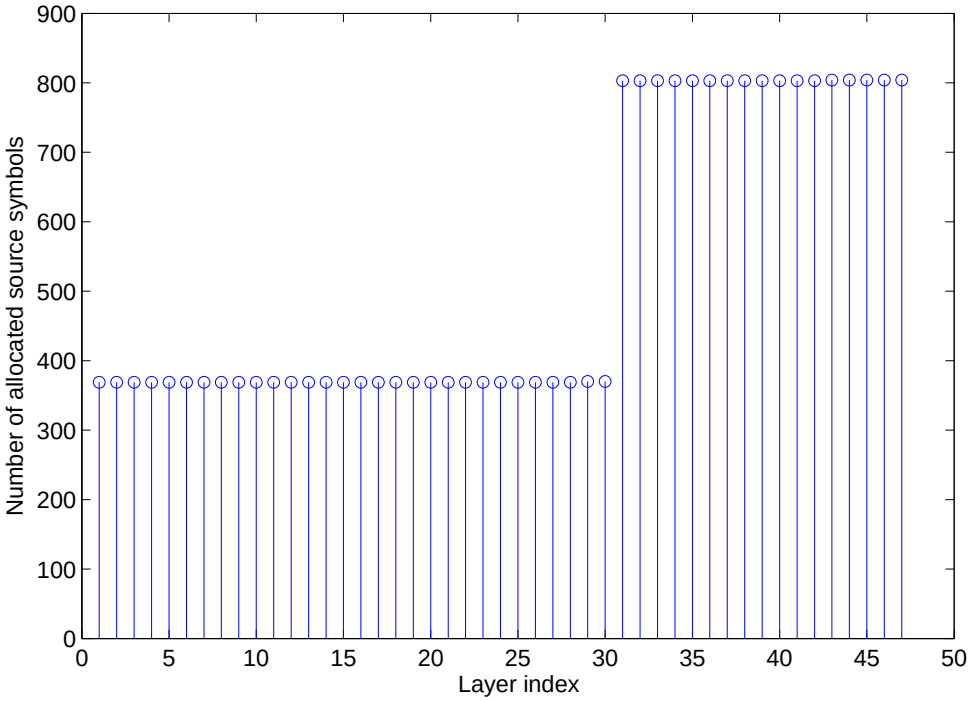

Figure 4.4. Allocation of source symbols for an asynchronous multimedia multicast system with two user classes and no transmission deadline with parameter $w_1 = 0.8$.

the steps described in Section 4.5.2, we first compare the values of $\frac{\sqrt{\alpha_j}}{U_j}$ among different classes. Since $\frac{\sqrt{\alpha_3}}{U_3} > \frac{\sqrt{\alpha_2}}{U_2}$, from Lemma 4.2, the optimal solution must satisfy $\frac{l_2}{U_2} = \frac{l_3}{U_3}$. Therefore, this Problem can be transformed into the form of Problem 4.22 by merging Chunk 2 and Chunk 3 to a new chunk with parameters $\hat{U}_2 = U_2 + U_3 = 755$ and $\hat{\alpha}_2 = \left(\alpha_j/U_j + \alpha_{j+1}/U_{j+1}\right)\left(U_j + U_{j+1}\right) = 554.9$. Since $\frac{\sqrt{\hat{\alpha}_j}}{\hat{U}_j}$ is now non-increasing, the optimal solution of Problem 4.22 can be computed as $\hat{l}_j^0 = \frac{L\sqrt{\hat{\alpha}_j}}{\sum_{i=1}^{\hat{J}}\sqrt{\hat{\alpha}_i}}, j = 1,2,3$. The parameters and solutions for the transformed problem, Problem 4.22, of this example are summarized in Table 4.2.

The optimal solution for Problem 4.21 can thus be obtained by $l_1^0 = \hat{l}_1^0$, $l_2^0 = \hat{l}_2^0 \frac{U_2}{U_2+U_3}$,

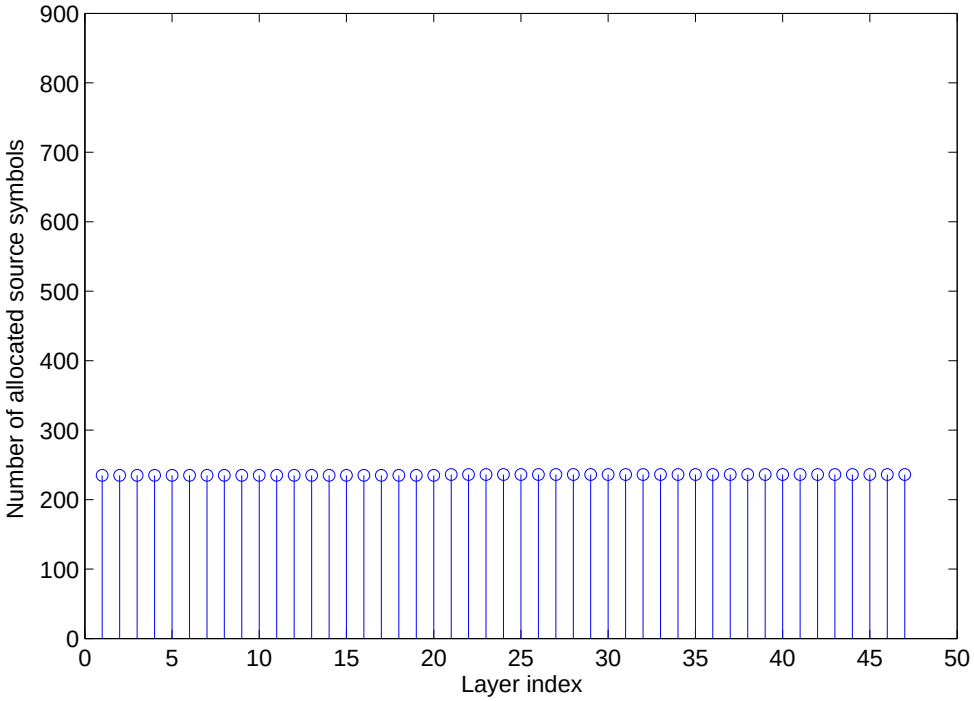

Figure 4.5. Allocation of source symbols for an asynchronous multimedia multicast system with two user classes and no transmission deadline with parameter $w_1 = 1$.

$l_3^0 = \hat{l}_2^0 \frac{U_3}{U_2+U_3}$ and $l_4^0 = \hat{l}_4^0$. With integer constraints, we obtain $l_1 = 14$, $l_2 = 7$, $l_3 = 10$, $l_4 = 16$. Alternatively, if we solve Problem 4.21 with parameters from Table 4.1 using `CVX` [28], a convex optimization software package for Matlab, we obtain the same optimal solution. The number of source symbols allocated to each layer, $K_i, i = 1,2,...,L$, are presented in Fig. 4.8.

Finally, if outage constraints are considered, the optimal allocation scheme can be obtained by solving Problem 4.10, or the transformed problem, Problem 4.11. Assume the maximum number of transmitted packets allowed for all user classes is given by the deadline M_0 and outage probability thresholds are given by $\tau_j = 0.01$, for $j = 1,2,3,4$. To

Table 4.1. Parameters of the asynchronous multimedia multicast example with four user classes.

User class index (j)	1	2	3	4
User erasure rate σ_j	0.6	0.5	0.4	0
User QoS requirement (PSNR threshold γ_j(dB))	25.79	27.25	29	40.28
number of decoded symbols required ($f^{-1}(\gamma_j)$)	400	700	1155	3800
weighting coefficient for Class j users (w_j)	0.4	0.1	0.3	0.2
number of symbols for each chunk (U_j)	400	300	455	2645
$\alpha_j = \frac{w_j U_j (1+\omega)}{1-\sigma_j}$	420.0	63.00	238.9	555.5
$\frac{\sqrt{\alpha_j}}{U_j}$	0.0512	0.0265	0.0340	0.0089

Table 4.2. Parameters and optimal solution of Problem 4.22 after transformation of the asynchronous multimedia multicast example.

User class index (j)	1	2	3
number of symbols for each chunk ($\hat{U}_j$)	400	755	2645
$\hat{\alpha}_j$	420.0	554.9	555.5
optimal solution $\hat{l}_0$	14.2	16.4	16.4

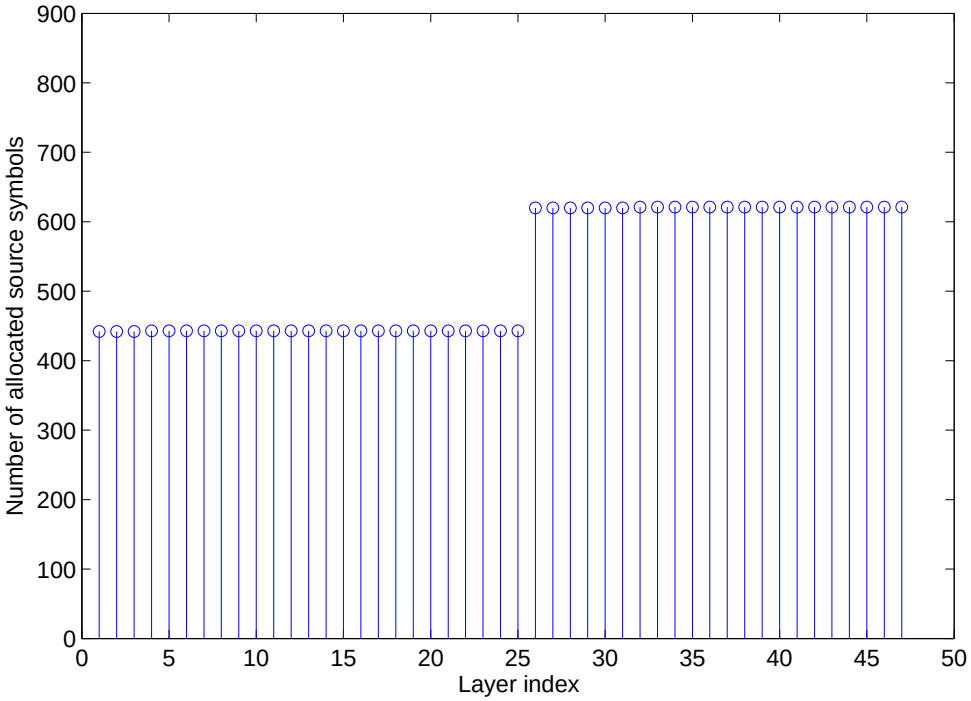

Figure 4.6. Allocation of source symbols for a two-user-class asynchronous multimedia multicast system with outage constraints $l_1 \geq 25, l_2 \geq 17$ and parameter $0 \leq w_1 \leq 0.6$.

show how the outage constraints affect the optimal allocation scheme in different situations, four different values of M_0, $M_0 = 60, 100, 140, 180$, are considered. For each value of M_0, we gradually increase the value of K_{h_j} until the outage probability calculated by Eq. (4.5) reaches the outage probability threshold τ_j. Denote the maximum value of K_{h_j} that keeps the outage probability below the threshold as $K^0_{h_j}$. The outage probability constraint is equivalent to $l_j \geq \theta_j$, where $\theta_j = \lceil U_j / K^0_{h_j} \rceil$. The resulting θ_j for different choices of M_0 are shown in Table 4.3.

When $M_0 = 60$ or $M_0 = 100$, we can see that $\sum_{i=1}^{J} \theta_i > L$, and no feasible solution exists. When $M_0 = 180$, as the optimal solution for Problem 4.5 without the outage constraint is

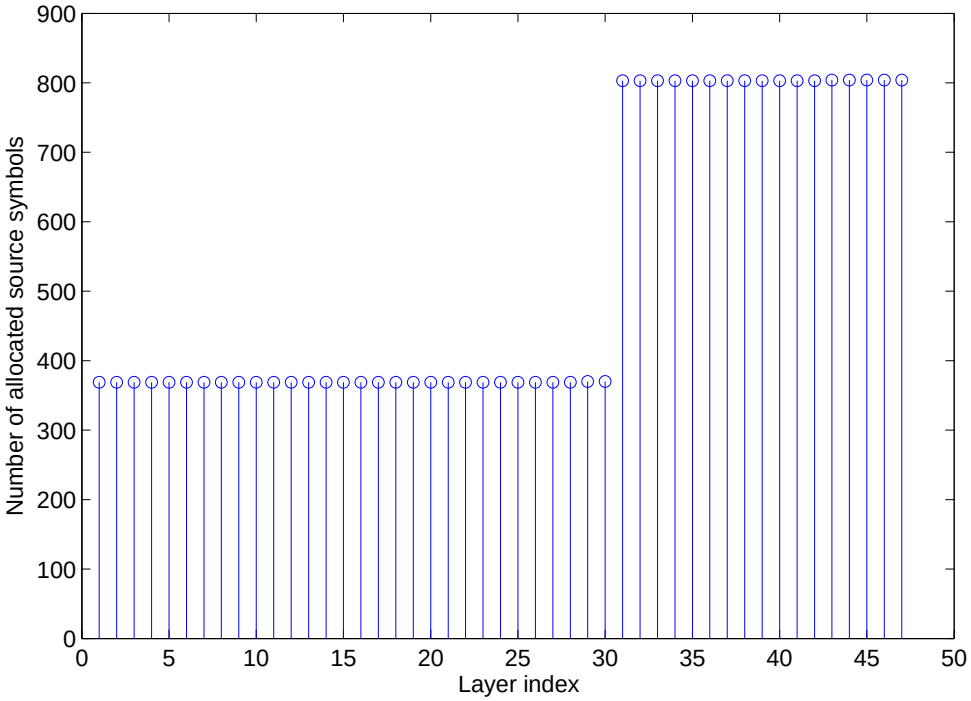

Figure 4.7. Allocation of source symbols for a two-user-class asynchronous multimedia multicast system with outage constraints $l_1 \geq 25, l_2 \geq 17$ and parameter $0.8 \leq w_1 \leq 1$.

given by $l_1^0 = 14$, $l_2^0 = 7$, $l_3^0 = 10$, $l_4^0 = 16$, they all satisfy $l_j \geq \theta_j$ for $j = 1, 2, .., J$. As a consequence, the optimal solution for Problem 4.11 is the same as the optimal solution for Problem 4.21.

When $M_0 = 140$, we can see $l_1^0 > \theta_1$, $l_2^0 > \theta_2$, $l_3^0 > \theta_3$, $l_4^0 < \theta_4$ and $\sum_{i=1}^{J} \theta_i < L$. Therefore, the problem has feasible solutions and the optimal solution is no longer the same as the case without outage constraints. Using the Matlab package, `CVX`, for specifying and solving convex programs [28], we obtain the optimal solution for $M_0 = 140$ as $l_1 = 13.03$, $l_2 = 5.95$, $l_3 = 9.02$ and $l_4 = 19.00$. With integer constraints, the solution is given by $l_1 = 13$, $l_2 = 6$, $l_3 = 9$ and $l_4 = 19$. The number of symbols allocated to each layer is

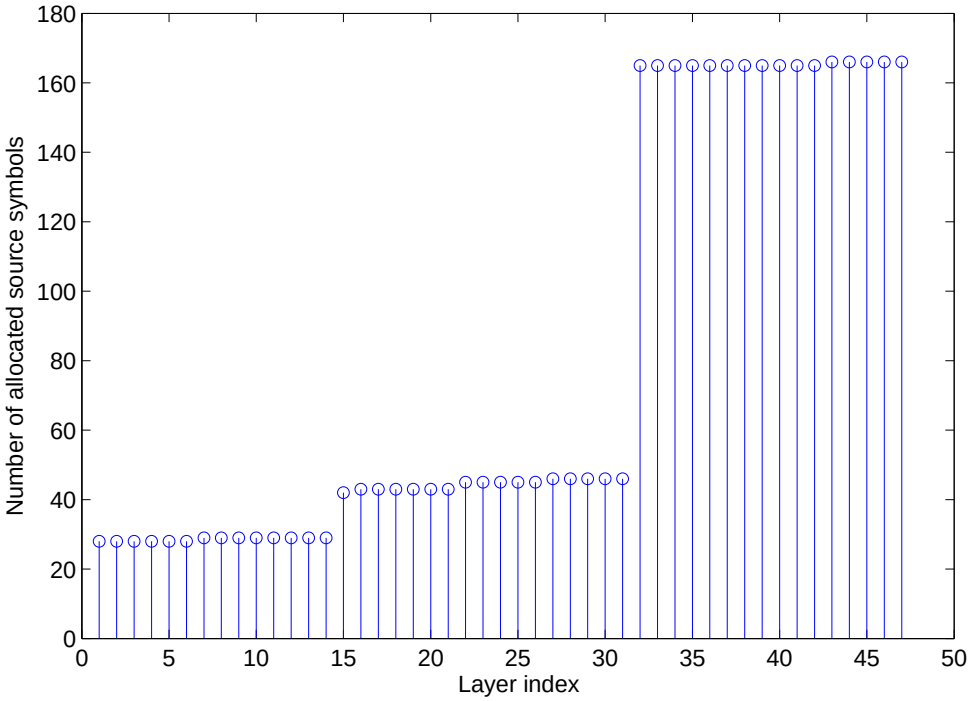

Figure 4.8. Allocation of source symbols for the asynchronous multimedia multicast system with four user classes without transmission deadline.

presented in Fig. 4.9. It can be seen that with a transmission deadline $M_0 = 140$, some of the source symbols that are originally allocated to the layers within Chunk 4 are instead allocated to layers within Chunks 1 to 3.

4.7 Conclusions

We show that the problem of optimal allocation of UEP rateless codes for asynchronous multimedia multicast [25] [26] can be transformed into a convex optimization problem when integer constraints are relaxed. Because the problem is convex and the total number of variables for the transformed problem is equal to the number of user classes, the solution

Table 4.3. The values of thresholds θ_j given different maximum number of transmitted packets M_0

M_0	θ_1	θ_2	θ_3	θ_4
60	25	14	16	44
100	13	7	9	27
140	9	5	6	19
180	7	4	5	15

is much simpler. In addition, an analytical solution is found for the case when there are no outage constraints [25], which has several orders of magnitude lower computational complexity compared to the greedy search algorithm proposed in [25]. For the more general formulation with outage constraints [26], an analytical solution is found for the special case of two user classes, and numerical methods of convex optimization software [27] [28] can be used to solve the more general cases.

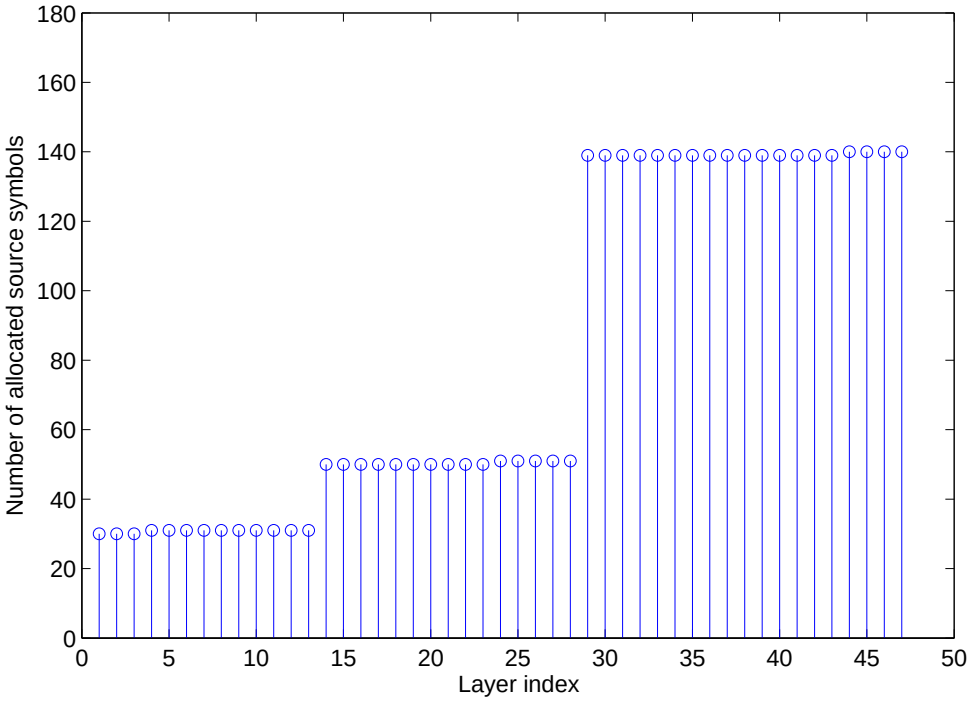

Figure 4.9. Allocation of source symbols for the asynchronous multimedia multicast system with four user classes with maximum number of transmitted packets $M_0 = 140$.

Chapter 5

Cross-layer rate allocation for wireless transmission

5.1 Introduction

To date, most studies of raptor codes focus on memoryless erasure channels where the erasure rate is fixed and known. However, due to the time varying nature of wireless channels, a physical-layer code is required to provide protection against fading and noise. In most communication systems, the physical-layer code and application-layer erasure code are studied separately. The quality of service (QoS) is usually guaranteed by the physical layer code to ensure a packet error rate (PER) below a certain level (1 percent for example). The packets that fail to be decoded by the physical-layer code are then corrected at the packet level by an erasure FEC or retransmitted using an automatic repeat request (ARQ) protocol. This scheme usually results in very little protection at the packet level. From an information theory perspective, since one can always use an ideal physical layer code to

drive the packet error rate to zero, erasure protection at the packet-level is not needed. If such an ideal physical layer code could be designed at a rate nearly equal to the information capacity that the channel supports, then a scheme without packet level erasure protection would be indeed "optimum". However, in practice, such an ideal physical-layer code does not exist, especially for time varying fading channels.

As the system simulation results in [20] suggest, in a raptor coded broadcast system, an optimal combination of physical and application-layer code rates exists. In many cases, a higher physical layer PER that is corrected by the application-layer raptor code can be more efficient than traditional designs. However, the results provided in [20] are from system simulations, and do not provide sufficient insight and quantification of the optimal rate combination. In [51], we also select the best overall code rates by evaluating the capacity of raptor codes in hybrid error-erasure channels when different physical layer code rates are used (see Fig. 4 of [51]), but do not attempt to find the optimal rate combination. Recently, [52] discusses certain aspects of the balance between the physical-layer code rate and packet-level erasure code rate in Rayleigh block fading channels. However, the formulation in [52] considers transmitting a prescribed number of information bits within a given transmission time. The rate of the packet-level rateless code is fixed during the whole transmission period. The objective of the optimization in [52] is to minimize the operational channel average SNR under certain outage constraints. The analysis in [52] provides theoretical insights of the best physical layer and application layer rate combination. However, in systems where the objective is to maximize throughput, the optimal solution in [52] can be very inefficient if the actual channel average SNR is known and much higher than the

minimum operational SNR. In addition, [52] considers an ideal physical-layer code which gives a zero packet error rate as long as the code rate is below the information capacity that the channel supports. Finally, [52] also does not consider rate adaption in the physical-layer when the channel fading is slow.

In this chapter, we investigate the optimal combination of physical-layer code rate and application-layer rateless code rate for systems in fading channels. The application-layer rateless code is assumed to keep generating coded packets on the fly until all erroneous packets are corrected. Therefore, the "rateless" property is exploited in our scheme to drive outage to nearly zero; while in [52], the erasure code rate is fixed during transmission and there is a certain probability of outage. In addition, compared to [52], more practical error curves for physical layer codes are used. We consider Rayleigh fading channels with both slow and fast fading. For fast fading channels, we find the optimal physical-layer modulation and code rate that maximizes the overall system throughput. We compare the proposed choice of the physical-layer code rate to the traditional choice of the physical-layer code rate to demonstrate the advantage of the proposed optimization method.

For slow fading channels where physical-layer rate adaptation is feasible, we propose a cross-layer adaptive modulation and coding scheme which maximizes overall system throughput. Adaptive modulation and coding (AMC) (e.g [53]) have been proposed for the physical layer in the literature for many different communication scenarios. In [54], AMC have been used in combination with truncated automatic repeat request (ARQ). However, all existing adaptive schemes essentially choose the modulation and coding mode to guarantee minimum PER requirement rather than provide the best overall system performance.

We derive the overall system throughput for different scenarios and compare the proposed cross-layer AMC design to traditional non cross-layer AMC designs.

The rest of this chapter is organized as follows: Section 5.2 describes the system setup, channel models, as well as introduces the performance measures in application-layer raptor coding and physical-layer modulation and coding. Section 5.3 discusses and derives the throughput performance of the proposed cross-layer transmission in fast fading channels as well as the cross-layer AMC design in slow fading channels. Section 5.5 presents the numerical results, which includes a comparison between the proposed cross-layer design and traditional non cross-layer design.

5.2 System setup and channel models

5.2.1 System model

The paper considers a two-layer model where the transmitter attempts to deliver messages to single or multiple users. The information bits are divided into data frames. Each data frame contains a total of K source symbols which are encoded by a raptor code to generate a potentially infinite number of raptor-encoded symbols. Each source symbol or raptor-encoded symbol consists of S_R bits. A packet is formed by L_R raptor encoded symbols ($P_d = L_R S_R$ bits) together with packet header information (P_h bits) and cyclic redundancy checks (CRC) (P_{CRC} bits). Thus the total number of bits within each packet, P_P, is given by $P_P = P_d + P_h + P_{CRC}$. Each packet is further protected by a physical-layer code with a code rate R_c and modulated using M-QAM and transmitted to the wireless fading channel.

Finally, pilot symbols are added to every one or more code frames depending on how fast the channel fades. The structure of each packet is shown in Fig. 5.1. For each data frame, the transmitter keeps sending packets until either it receives acknowledgements from all the intended receivers or until the maximum number of packets allowed has been reached. After that, the transmitter starts to transmit the next data frame. On the receiver side, each user first demodulates and decodes the physical-layer code and checks its correctness with a CRC. If there are uncorrected errors, the entire packet is dropped. Otherwise, the correctly decoded packets are further used to decode the original information data using a raptor decoder. Once the receiver successfully decodes the data frame, an acknowledgement is sent to the transmitter. Let N_p be the actual number of packets transmitted for each data frame. Then the number of raptor encoded symbols sent is $K_t = N_p L_R$. For analytical simplicity, we assume $L_R = 1$, i.e, each raptor coded symbol is a physical-layer packet.

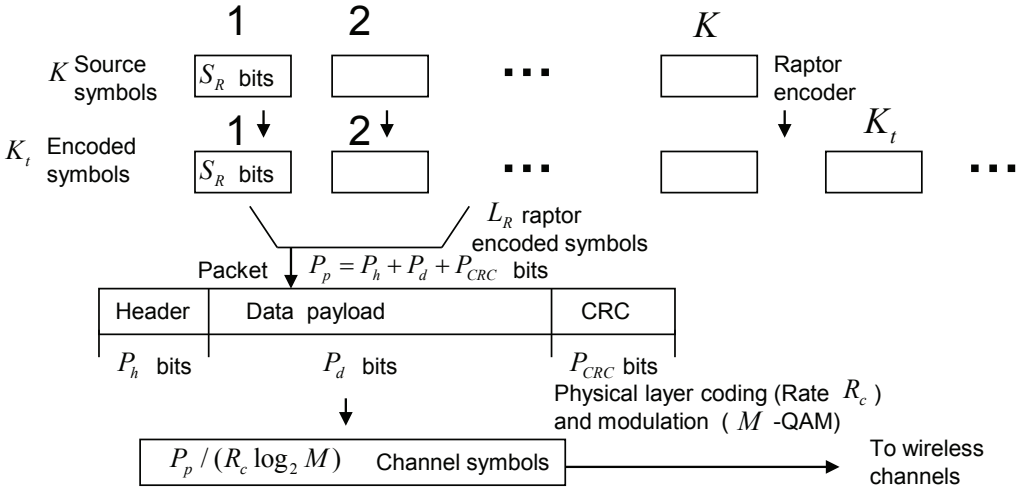

Figure 5.1. System setup and packet structure.

To quantify the overall system performance, we calculate the overall system throughput in terms of the average number of information bits transmitted per channel symbol. To simplify the notation, define $R_R = K/K_t$ as the realized raptor code rate, $\varepsilon_p = (P_{CRC} + P_h)/P_d$ as the packet overhead where P_h, P_{CRC} and $P_d = L_R S_R$ are the numbers of bits representing packet header information, CRCs, and data payload in each physical-layer packet, respectively. $R_m = \log_2 M$ is the modulation rate. Then the system throughput, which is defined by the total number of information bits divided by total number of channel symbols used, can be calculated as

$$\begin{aligned} TP &= \frac{\text{total number of information bits}}{\text{total number of channel symbols}} \\ &= \frac{KS_R}{P_P N_P/(R_m R_c)} \\ &= \frac{KS_R}{P_d N_P} \frac{1}{1+\varepsilon_p} R_m R_c \\ &= \frac{KS_R}{L_R S_R K_t/L_R} \frac{1}{1+\varepsilon_p} R_m R_c \\ &= \frac{1}{1+\varepsilon_p} R_m R_c R_R \text{ bits/channel symbol.} \end{aligned} \tag{5.1}$$

5.2.2 Physical-layer channel model

We assume the link between the server and all the users to be modeled as independently and identically distributed (i.i.d) Rayleigh block fading channels. The channel quality of each user is characterized by the instantaneous signal-to-noise-ratio (SNR) γ. In the Rayleigh fading model, the probability density function (pdf) of the instantaneous channel SNR γ

can be characterized as:

$$f_\gamma(\gamma) = \frac{1}{\bar{\gamma}} \exp -\frac{1}{\bar{\gamma}}, \tag{5.2}$$

where $\bar{\gamma}$ is the average channel SNR.

We also categorize the channel fading into two different scenarios depending on the coherence time of channel fades. In the fast fading scenario, we assume that the coherence time of the channel fades are of a similar length to the packet transmission time. Therefore, the instantaneous channel *SNR* remains the same for each packet, but varies from packet to packet. In the slow fading scenario, we assume the coherence time to be much longer than the transmission time of a raptor coded data frame. Therefore, for slow fading, the instantaneous SNR is assumed to remain the same for the whole raptor coded data frame but varying from frame to frame. It is also assumed that the instantaneous channel SNR is available to the transmitter in the slow fading scenario.

5.2.3 Packet error rate and physical layer design

At the physical layer, we consider possible choices of multiple transmission modes with different combinations of modulation and convolutional coding pairs as in [54], which are borrowed from HIPERLAN/2, IEEE 802.11a and 3GPP standards. For analytical simplicity, we use the PER expressions developed in [54], where for a given modulation and code pair in mode *n*, the packet error rate is approximated by:

$$PER_n(\gamma) = \begin{cases} 1 & \text{if } \gamma < \gamma_{pn} \\ a_n \exp(-g_n \gamma) & \text{if } \gamma \geq \gamma_{pn} \end{cases} \tag{5.3}$$

where γ is the instantaneous channel SNR, which is assumed to remain unchanged for the whole packet. In [54], a_n, g_n and γ_{pn} are parameters that are mode-dependent. The parameter values for different uncoded M-QAM (rectangular) modulations (TM1) and convolutionally coded modulations (TM2) are listed in Table I and Table II of [54], respectively.

5.2.4 Application-layer raptor codes

Raptor codes see a virtual erasure channel in the application layer. The erasure probability P_e is equal to the packet error rate. The decoding failure probability of the raptor code used in the 3GPP standard can be well modeled by the following equation [20],

$$P_{fail}(K_s, K) = \begin{cases} 1 & \text{if } K_s \leq K \\ ab^{K_s - K} & \text{if } K_s > K \end{cases} \tag{5.4}$$

where $P_{fail}(K_s, K)$ denotes the probability that the receiver fails to decode K source packets after K_s packets are successfully received and $a = 0.85$ and $b = 0.567$ are constants. Denote K_s and K_t as random variables representing the number of successfully received packets, and the total number of transmitted packets during the transmission of one raptor code frame, respectively. Consider a continuous transmission of q frames. The effective raptor code rate $R_R = \frac{qK}{\sum_{i=1}^{q} K_t(i)}$, where $K_t(i)$ is the number of packets transmitted during the i-th frame. As q gets large over a long time, according the weak law of large numbers (WLLN), $\frac{1}{q}\sum_{i=1}^{q} K_t(i) \to E(K_t)$. Therefore, the average long-term effective raptor code rate can be evaluated as $R_R = K/E(K_t)$.

To evaluate $E(K_t)$, since the packets transmitted through the erasure channel represent a Bernoulli process with failure probability equal to the erasure probability P_e, it can be

shown (proof-omitted, see e.g. [55]) that the random variable K_t, conditioned on K_s follows a negative binomial distribution, with average value equal to $E(K_t|K_s) = \frac{K_s}{1-P_e}$. Hence $E(K_t) = E(E(K_t|K_s)) = E(\frac{K_s}{1-P_e}) = \frac{E(K_s)}{1-P_e}$. To calculate $E(K_s)$ using Eq. (5.4), we have, for a positive integer x,

$$\begin{aligned} P(K_s = K+x) =& Prob(\text{decoding failure when } K_s < K+x) \\ & \times Prob(\text{decoding success when } K_s = K+x) \\ =& \prod_{i=0}^{x-1}(ab^i)(1-ab^x); \end{aligned} \tag{5.5}$$

Therefore,

$$\begin{aligned} E(K_s) =& K + \sum_{x=0}^{\infty} xP(K_s = K+x) \\ =& K + \sum_{x=0}^{\infty} x[\prod_{i=0}^{x-1}(ab^i)(1-ab^x)] \\ =& K + \sum_{x=0}^{\infty} xa^x b^{x(x-1)/2}(1-ab^x) \\ =& K + C, \end{aligned} \tag{5.6}$$

where C is a constant that can be easily evaluated numerically (here $C \simeq 1.39$). Therefore, the realized raptor code rate [1] can be evaluated as

$$R_R = \frac{K}{E(K_t)} = \frac{K(1-PER)}{K+C} = \frac{(1-PER)}{1+\varepsilon_r}, \tag{5.7}$$

where PER is the packet error rate and $\varepsilon_r = C/K$ is defined as the raptor code overhead, which decreases as the raptor code dimension K increases. It should be noted that Eq. (5.4)

[1] We have assumed that the maximum number of transmitted symbols for each code frame is large enough except when $PER = 1$, in which case an outage event occurs.

used in this section is not critical to the results of this chapter. This is because the optimal choice of physical layer modulation and code pairs does not depend on the average code overhead ε_r.

5.3 Performance analysis in fast fading channels

In the fast fading scenario, each raptor encoded packet experiences a different instantaneous SNR. Therefore the erasure rate for application layer raptor codes is equal to the PER averaged over different channel SNRs. Therefore, for a given transmission mode n (with modulation rate $m_n = \log_2 M$ and code rate R_{cn}), using Eqs. (5.3) and (5.2), the erasure rate for the raptor code is:

$$\begin{aligned} P_e(n) &= \int_0^{\infty} PER_n(\gamma) f_\gamma(\gamma) d\gamma \\ &= \int_0^{\gamma_{pn}} f_\gamma(\gamma) d\gamma + \int_{\gamma_{pn}}^{\infty} a_n \exp(-g_n\gamma) f_\gamma(\gamma) d\gamma \\ &= 1 - \exp(-\frac{\gamma_{pn}}{\bar{\gamma}}) + \frac{a_n}{\bar{\gamma} g_n + 1} \exp(-(g_n + \frac{1}{\bar{\gamma}})\gamma_{pn}). \end{aligned} \tag{5.8}$$

By substituting Eqs. (5.7), (5.8) and $PER = P_e(n)$ into Eq. (5.1), the system throughput in fast fading for given transmission mode n is given by

$$\begin{aligned} TP_{fast}(n) &= \frac{1}{1+\varepsilon_p} m_n R_{cn} R_R \\ &= \frac{1}{(1+\varepsilon_p)(1+\varepsilon_r)} m_n R_{cn} (1 - PER) \\ &= \frac{1}{(1+\varepsilon_p)(1+\varepsilon_r)} m_n R_{cn} (1 - P_e(n)) \\ &= \frac{1}{(1+\varepsilon_p)(1+\varepsilon_r)} m_n R_{cn} (\exp(-\frac{\gamma_{pn}}{\bar{\gamma}}) \end{aligned}$$

$$- \frac{a_n}{\bar{\gamma} g_n + 1} \exp(-(g_n + \frac{1}{\bar{\gamma}})\gamma_{pn})). \tag{5.9}$$

5.3.1 Traditional non cross-layer scheme

Commonly, the physical layer modulation and coding design is independent of the upper layer design. The QoS is guaranteed entirely by the physical layer which ensures that the packet dropping rate is below a certain level. Therefore, when a traditional non cross-layer scheme is used, the transmitter chooses the transmission mode with the highest modulation and coding rate which guarantees that the average PER (or erasure rate) is below a certain value P_{loss}. Therefore, for a known average SNR, transmission mode n is chosen according to the following criterion:

$$n = \arg\max_n (m_n R_{cn}) \text{ subject to } P_e(n) \leq P_{loss}. \tag{5.10}$$

We remark that $P_{loss} = 0.01$ is chosen for all performance comparisons in this thesis.

5.3.2 Proposed cross-layer scheme

In the proposed cross-layer scheme, the physical layer design is aware of the application layer FEC. Here, the QoS is guaranteed by both the physical layer and the application layer FECs. The application layer rateless code is able to correct dropped packets by generating sufficient numbers of raptor encoded packets to drive the outage probability to nearly zero. Therefore, the objective of the proposed cross-layer design is to maximize the overall system throughput. Since the overhead parameters ε_p and ε_r are independent of mode n, the criteria for choosing the transmission mode n based on our proposed cross-layer scheme

can be found by using the mode that maximizes Eq. (5.9):

$$\begin{aligned} n &= \arg\max_n \{TP_{fast}(n)\} \\ &= \arg\max_n \{m_n R_{cn}(\exp(-\frac{\gamma_{pn}}{\overline{\gamma}}) \\ &- \frac{a_n}{\overline{\gamma} g_n + 1}\exp(-(g_n + \frac{1}{\overline{\gamma}})\gamma_{pn}))\}. \end{aligned} \tag{5.11}$$

5.4 Performance analysis in slow fading channels

When the fading is slow, the instantaneous SNR remains the same for each raptor coded data frame. Therefore, the throughput performance in slow fading channel can be obtained by averaging the throughput performance for a given SNR over the distribution of SNR. For any given SNR, it is assumed that the rateless code in the application layer experiences a memoryless erasure channel. Therefore, Eq. (5.7) can still be used to compute the realized rateless code rate. By substituting Eq. (5.7) to Eq. (5.1), we obtain that the system throughput performance for a given SNR γ, which is the same as the throughput performance in an AWGN channel, when mode n is chosen and is given by:

$$TP_{AWGN}(n,\gamma) = \frac{m_n R_{cn}(1 - PER_n(\gamma))}{(1+\varepsilon_p)(1+\varepsilon_r)}, \tag{5.12}$$

where $PER_n(\gamma)$ is given by Eq. (5.3).

5.4.1 Non adaptive scheme

When a non-adaptive scheme is used, the transmission mode n is fixed during the whole transmission period. The long term average throughput for a non-adaptive scheme using

mode n is then given by:

$$TP_{slow}(n) = \int_0^{\infty} TP_{AWGN}(n,\gamma)\, f_{\gamma}(\gamma) d\gamma. \tag{5.13}$$

By substituting Eqs. (5.12) and (5.3) into Eq. (5.13), we obtain,

$$\begin{aligned} TP_{slow}(n) =& \frac{1}{(1+\varepsilon_p)(1+\varepsilon_r)} m_n R_{cn} (\exp(-\frac{\gamma_{pn}}{\overline{\gamma}}) \\ & - \frac{a_n}{\overline{\gamma} g_n + 1} \exp(-(g_n + \frac{1}{\overline{\gamma}})\gamma_{pn})). \end{aligned} \tag{5.14}$$

5.4.2 Proposed cross-layer AMC scheme

For the slow fading scenario, we propose a cross-layer AMC scheme where the transmission mode is chosen according to the instantaneous SNR to maximize the system throughput. In this case, for a given instantaneous channel SNR γ, the modulation and coding mode is chosen according to

$$n = \arg\max_n \{m_n R_{cn}(1 - PER_n(\gamma))\}, \tag{5.15}$$

where $PER_n(\gamma)$ is given by Eq. (5.3).

Without loss of generality, it is assumed that the transmission modes are ordered such that $m_n R_{cn}$ is monotonically increasing with n, and N is the total number of available transmission modes. As shown later in Figs. 5.2 and 5.3, when the SNR γ is low but greater than γ_{p1}, transmission mode 1 provides the maximum throughput. While SNR γ gradually increases, the index of the best transmission mode n also increases. The design criteria can be further simplified to choose mode n when the SNR γ lies between γ_n and γ_{n+1}, where $\gamma_1 = \gamma_{p1}$, $\gamma_{N+1} = \infty$ and $\gamma_n (n = 1, 2, ..., N)$ is the solution to the following equation,

$$(1 - a_{n-1} \exp(-g_{n-1}\gamma_n)) \times m_{n-1} R_{c(n-1)}$$

$$= (1 - a_n \exp(-g_n \gamma_n)) \times m_n R_{cn} \tag{5.16}$$

for $n = 1, 2, ..., N$, When the channel SNR is below γ_{p1}, the PER approaches 1, and an outage event occurs. The outage probability can be calculated as

$$P_{out} = \int_0^{\gamma_{p1}} f_\gamma(\gamma) d\gamma = 1 - \exp(-\frac{\gamma_{p1}}{\overline{\gamma}}). \tag{5.17}$$

The probability that mode n is chosen is given by:

$$P_n = \int_{\gamma_n}^{\gamma_{n+1}} f_\gamma(\gamma) d\gamma = \exp(-\frac{\gamma_n}{\overline{\gamma}}) - \exp(-\frac{\gamma_{n+1}}{\overline{\gamma}}). \tag{5.18}$$

Therefore, the long-term average system throughput using the proposed cross-layer AMC scheme can be evaluated as:

$$\begin{aligned} TP_{AMC} &= \sum_{n=1}^{N} P_n E(TP_{AMC} | \gamma_n < \gamma \leq \gamma_{n+1}) \\ &= \sum_{n=1}^{N} P_n \int_{\gamma_n}^{\gamma_{n+1}} TP_{AWGN}(n, \gamma) \frac{f_\gamma(\gamma)}{P_n} d\gamma \\ &= \sum_{n=1}^{N} \int_{\gamma_n}^{\gamma_{n+1}} TP_{AWGN}(n, \gamma) f_\gamma(\gamma) d\gamma. \end{aligned} \tag{5.19}$$

Applying Eqs. (5.2) and (5.12) to Eq. (5.19), we obtain the average throughput

$$\begin{aligned} TP_{AMC} = &\frac{1}{(1+\varepsilon_h)(1+\varepsilon_r)} \{ \sum_{n=1}^{N} m_n R_{cn} \times [\exp(-\frac{\gamma_n}{\overline{\gamma}}) \\ &- \exp(-\frac{\gamma_{n+1}}{\overline{\gamma}}) + \frac{a_n}{b_n \overline{\gamma}} \exp(-b_n \gamma_{n+1}) \\ &- \frac{a_n}{b_n \overline{\gamma}} \exp(-b_n \gamma_n)] \}, \end{aligned} \tag{5.20}$$

where $b_n = g_n + 1/\overline{\gamma}$.

5.4.3 Non cross-layer AMC scheme

Traditionally, AMC design in physical layer is independent of data link layer and application layer designs. If AMC is used in a traditional non cross-layer design, the transmitter chooses the transmission mode with the highest modulation and coding rate while still ensuring the packet error rate given the current channel instantaneous SNR is below threshold P_{loss}. Therefore, the choice of the modulation and coding pair is according to the following criterion:

$$n = \arg\max_{n} (m_n R_{cn}) \text{ subject to } PER_n(\gamma) \leq P_{loss}, \tag{5.21}$$

where $PER_n(\gamma)$ is given by Eq. (5.3). The outage probability and overall throughput using a traditional non cross-layer AMC scheme can be obtained similarly except the threshold γ_n is

$$\gamma_n = \frac{1}{g_n} \ln(\frac{a_n}{P_{loss}}). \tag{5.22}$$

Note that the choice of the transmission mode of the non cross-layer AMC scheme described above is the same as that of the AMC design described in [54]. The only difference is that the threshold P_{target} is used in [54] instead of P_{loss} used in this chapter, where P_{target} is a constant determined by the maximum number of re-transmission attempts allowed in the ARQ protocol as well as by the maximum packet loss rate after the re-transmission attempts.

5.5 Numerical Results

With a packet length $P_p = 1080$ bits, the parameters a_n, g_n, γ_{pn} for calculating packet error rates under different modulation and coding modes have been listed in Table I and Table II of [54]. The throughput performance of the combinations of different modulation and coding schemes (uncoded and convolutionally coded) with erasure raptor codes over AWGN channels can be calculated by Eqs. (5.12) and (5.3). The throughput performances of uncoded M-QAM modulations and convolutionally coded M-QAM modulations in AWGN channels are shown in Figs. 5.2 and 5.3, respectively. Without loss of generality, $\varepsilon_p = 0$ and $\varepsilon_r = 0$ are used for Figs. 5.2 and 5.3. Although the focus of the chapter is on fading channels, the results for AWGN channels are also presented to illustrate how different modulation and coding modes are chosen at different SNRs in adaptive schemes. From

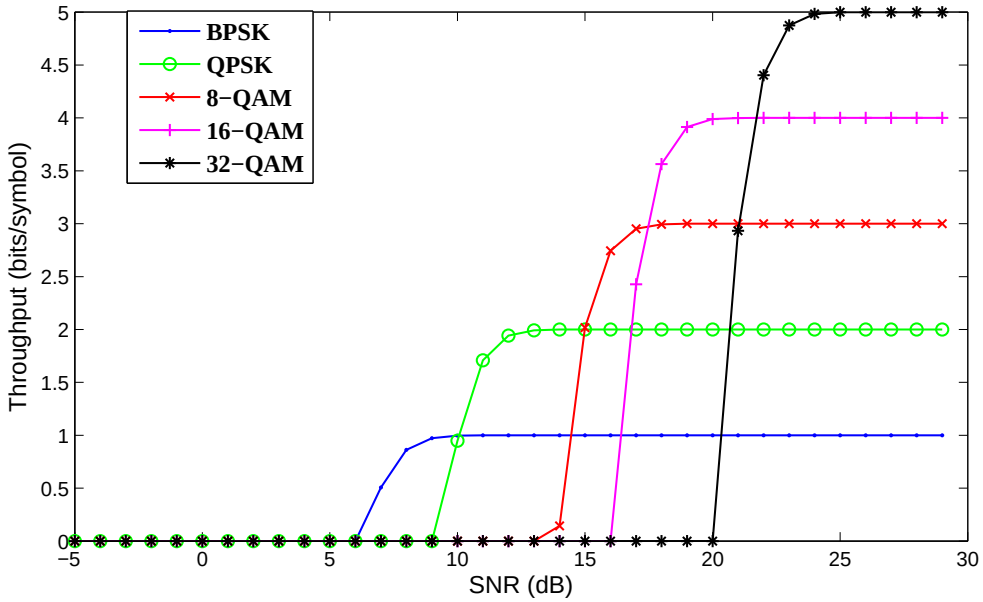

Figure 5.2. Throughput performance of uncoded packets in AWGN channels.

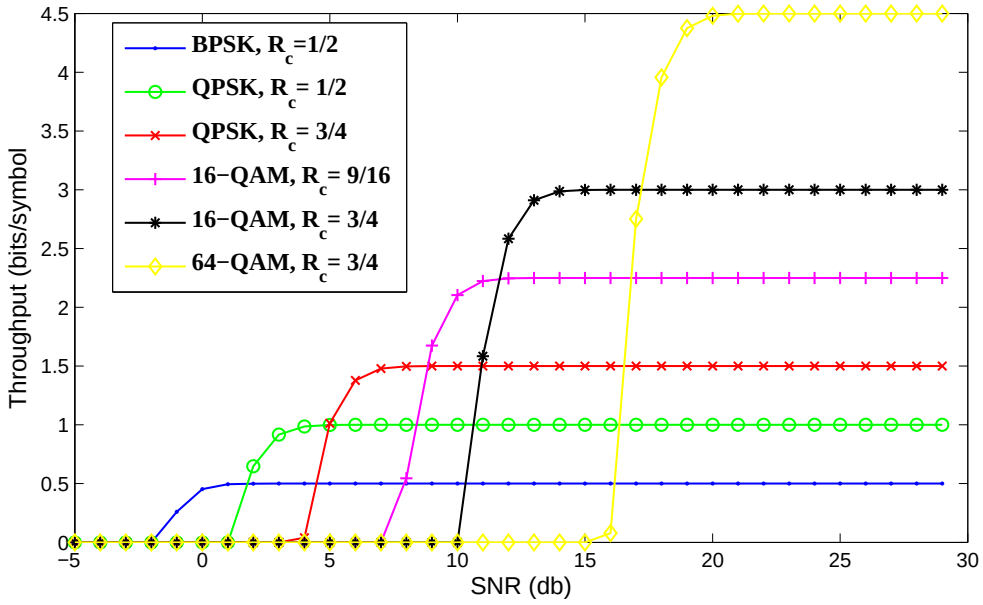

Figure 5.3. Throughput performance of convolutinally coded packets in AWGN channels.

Figs. 5.2 and 5.3, it can be seen that the best modulation and coding mode which maximizes the throughput differs for different SNRs. In very low SNR, coded BPSK offers the best throughput performance; while at high SNR, uncoded high level modulation is best. This is to be expected, as in low SNR, low-rate codes and modulation schemes should be chosen to achieve an acceptable packet error rate; On the other hand, in high SNR, a small packet error rate can be easily corrected by the raptor codes, and a high-rate modulation and coding scheme can achieve a higher rate. Therefore, it is beneficial to use AMC to select the best modulation and coding mode according to the SNR when the channel varies over time. Comparing all the curves in both Figs. 5.2 and 5.3, it can be seen that the uncoded schemes using BPSK, QPSK, 8-QAM and 16-QAM does not offer maximum throughput at any SNR. Therefore, in the proposed AMC scheme for slow fading channels, only the

coded modulation mode and uncoded 32-QAM, 64-QAM, 128-QAM transmission modes are candidates.

We next compare the performance of different choices of modulation and coding schemes in fast fading channels. Without loss of generality, for all the numerical results presented afterwards, it is assumed that packet overhead $\varepsilon_p = 2\%$, raptor code dimension $K = 256$ and raptor code overhead $\varepsilon_r = C/K = 0.0054$. Fig. 5.4 shows the throughput performance of the proposed cross-layer optimized transmission mode for Rayleigh fading channels for different average SNRs. For comparison, throughput performance of three specific transmission modes as well as the transmission mode chosen by the non cross-layer scheme are also shown in Fig. 5.4. For all the curves in Fig. 5.4, the throughput is calculated using Eq. (5.9), with the transmission mode n being selected according to the corresponding criteria described as follows: for the "proposed cross-layer selection scheme", the transmission mode n is selected according to the criteria described by Eq. (5.11); for the traditional "non cross-layer selection scheme", the transmission mode n is chosen according to the criteria described by Eq. (5.10); for the curves representing specific transmission mode n, the modulation and coding pair used for the transmission is pre-designed and fixed for all the average SNR values. It can be seen that the proposed cross-layer transmission scheme offers significantly better overall system performance over most of the SNR range compared to the traditional non cross-layer transmission scheme.

Fig. 5.5 shows the resulting average packet error rate $P_e(n)$ for the proposed cross-layer scheme and the traditional non cross-layer scheme in fast fading channels. The average PER is calculated using Eq. (5.8) with the transmission mode n selected according to the

corresponding criteria. It can be observed that while traditional non cross-layer schemes use strong and conservative physical layer coding to guarantee a low PER, the optimal choice of transmission mode from a throughput maximization perspective involves transmission with a high PER to be corrected by the application-layer raptor code. Note that in the case of low SNR where a PER below 0.01 cannot be guaranteed for the traditional non cross-layer scheme no matter which transmission mode is used, the PER is computed by choosing the transmission mode $n = 1$ which corresponds to *BPSK* modulation with code rate $R_c = 1/2$.

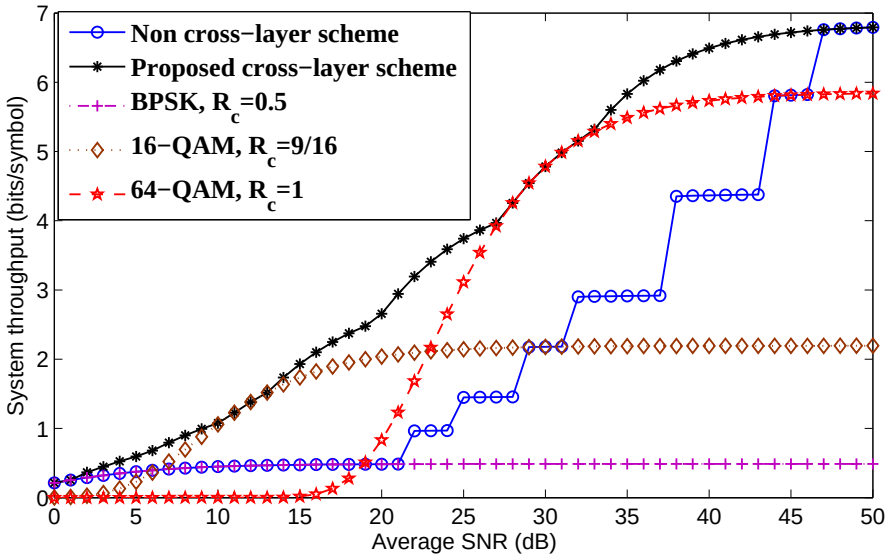

Figure 5.4. Throughput performance comparison for fast fading ($\varepsilon_p = 0.02, K = 256, \varepsilon_r = 0.0054$).

In Fig. 5.6, for the slow fading scenario, the throughput performance of the proposed cross-layer AMC scheme is shown in comparison with the best performing non-adaptive

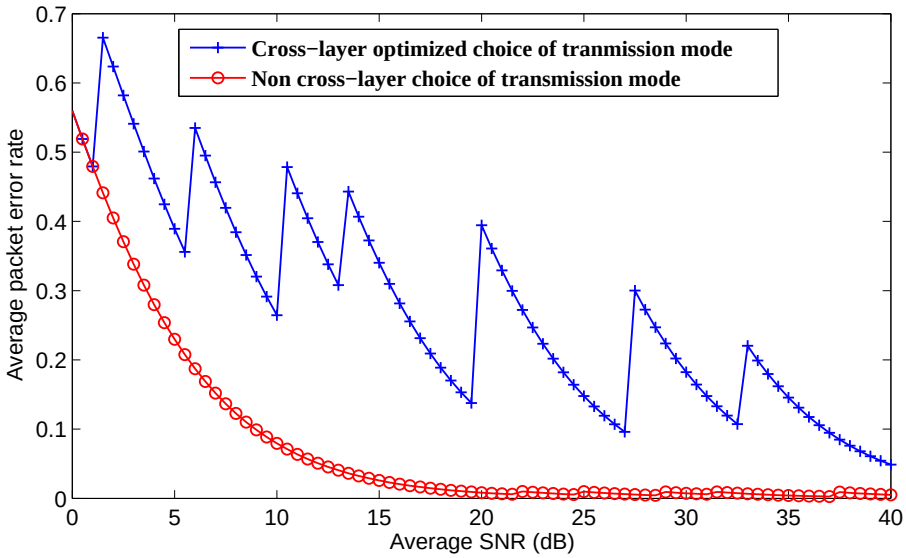

Figure 5.5. Average packet error rate in fast Rayleigh fading channels.

scheme, as well as AMC using traditional non cross-layer design criteria. When the proposed cross-layer AMC scheme is used, the transmission mode n for any given instantaneous SNR γ is chosen according to the criteria described in Eq. (5.15). The throughput performance of the cross-layer AMC scheme is given by Eq. (5.20), where the threshold parameters γ_n is given by the solution of Eq. (5.16). When the non cross-layer AMC scheme is used, the transmission mode n for any given instantaneous SNR γ is selected according to criteria described by Eq. (5.21) and the throughput of non cross-layer AMC scheme is calculated using Eq (5.20) with parameters γ_n given by Eq. (5.22). For the non-adaptive schemes, the throughput performance is calculated using Eq. (5.14), where for each given average SNR $\bar{\gamma}$, the best performing transmission mode $n = \arg\max_n\{TP_{slow}(n)\}$ is chosen, where $TP_{slow}(n)$ is the throughput given by Eq. (5.14). Note that we have shown

two curves of non-adaptive schemes where the candidate transmission modes are limited to either uncoded modulation modes or convolutionally coded modulation modes. It can be seen that both non cross-layer and cross-layer AMC schemes outperform the best performing non-adaptive schemes. The cross-layer AMC design performs better than traditional non cross-layer AMC design across the entire average SNR range considered. For example, at an average SNR of 10dB, the cross-layer AMC, traditional AMC, and non AMC designs achieve throughputs of 1.68, 1.42, and 1.10 bits/symbol, respectively.

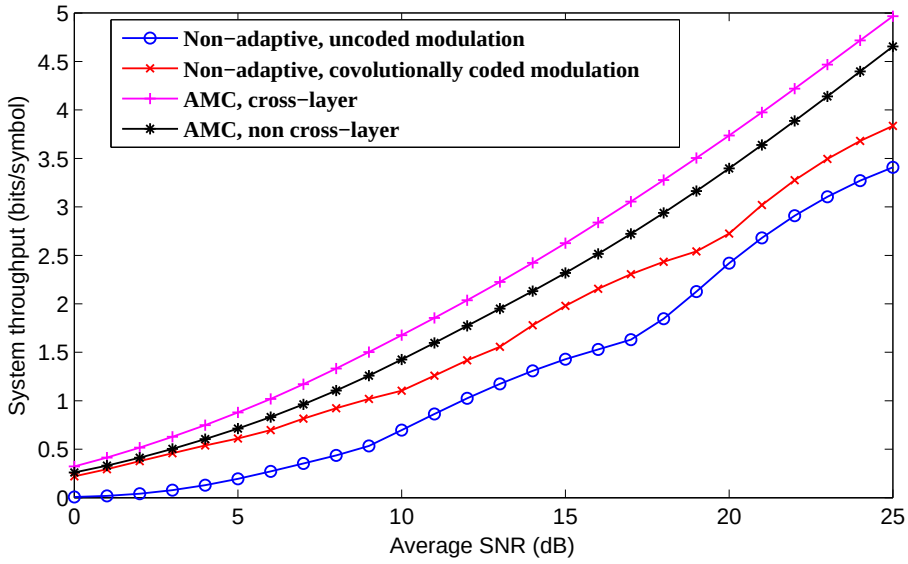

Figure 5.6. Throughput performance comparison for slow fading ($\varepsilon_p = 0.02, K = 256, \varepsilon_r = 0.0054$).

Fig. 5.7 shows the PER of the transmission mode chosen by the proposed cross-layer and traditional non cross-layer schemes as a function of instantaneous SNRs for the slow fading scenario. It can be seen that while traditional non cross-layer schemes keep the PER very low (below $P_{loss} = 0.01$), the proposed optimized cross-layer scheme allows much

larger PER to be corrected by the application-layer raptor codes.

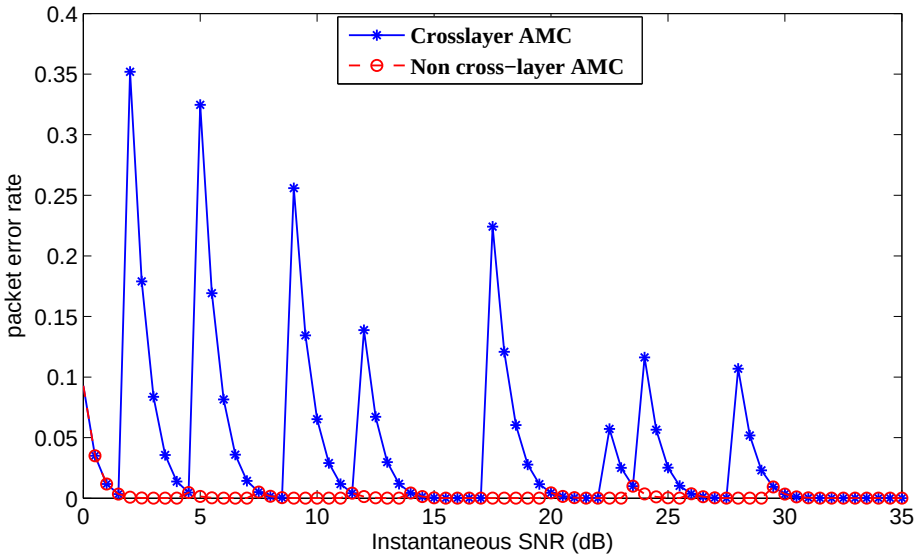

Figure 5.7. Packet error rate of traditional and optimal choice of transmission mode as a function of instantaneous SNR.

5.6 Conclusions

This chapter studies the optimal combination of the physical layer and application-layer raptor code rates for rateless coded communication systems. We consider both slow and fast Rayleigh fading channels. We propose a cross-layer scheme to optimize the set of available physical-layer modulation and coding pairs. In addition, a cross-layer adaptive modulation and coding design is proposed for the slow fading scenario. The system throughput performance that considers both physical-layer code rate and application-layer erasure code rate is analyzed. Numerical results show that the proposed cross-layer design outperforms

traditional non cross-layer design significantly in both slow and fast fading channels.

Chapter 6

Hybrid error-erasure decoding of raptor codes over wireless channels

6.1 Introduction

In wireline internet communications, as packets can be lost due to network congestion, the channel can be easily modeled by a packet erasure channel. In wireless communications, packets can be both lost and corrupted. Traditionally, corrupted packets are discarded and not forwarded to the application layer. These dropped packets are either re-transmitted using an automatic repeat request (ARQ) protocol or are recovered using application layer forward error correction (FEC) codes. These schemes can result in large packet drops and hence very low throughput when the channel condition is poor. To mitigate the inefficiency of such schemes, newer cross-layer protocols allow corrupted packets to be relayed into the application layers. With these protocols, the application layer FEC sees both erasures and errors. Such channels can be modeled as hybrid error-erasure channels. The simplest form

of a hybrid error-erasure channel is the binary symmetric channel with erasures (BSCE).

To date, most of the applications envisioned for raptor codes assume a perfect erasure channel [20]. For example, authors in [20] investigated the application of raptor codes in download delivery of multimedia broadcast and multicast services (MBMS). In their system, packets which contain errors that are not fully corrected by the physical layer turbo code are discarded. Such schemes may result in large numbers of dropped packets in poor channel conditions. The performance of raptor codes over noisy channels have been investigated using soft decoding [22] [23] [56]. It is found in [22] that unlike binary erasure channels (BECs), the optimal degree distribution for raptor codes is no longer universal for binary input additive white gaussian noise (BIAWGN) and binary symmetric channels (BSCs), but depends on the noise level. Nevertheless, a raptor code designed for BEC performs quite well in BSC and BIAWGN channels [22] [23]. Therefore, raptor codes are good FEC candidates in correcting both erasures and errors. However, soft decoding schemes are usually more complex than traditional erasure decoding schemes. In addition, raptor codes are not originally designed as physical layer codes, and incorporating raptor codes directly in the physical layer would require modification to the whole system design which is not practical. To date, the performance of raptor codes for hybrid error-erasure channels has not been explored.

In reality, wireless fading channels are correlated, which results in a burstiness of bit errors and erasures. A FEC usually performs best when each code symbol experiences independent fading. The traditional way to deal with channel memory is to interleave the encoded symbols prior to transmission. However, interleaving introduces large delays

and complexity, motivating the study of the performance of raptor codes in channels with memory. To the author's best knowledge, only [56] [57] have addressed this issue. In [56], the authors study the performance of fixed-rate raptor codes over Rayleigh fading channels with memory via simulation. However, [56] considers raptor codes as a pure physical layer code, while the conventional usage of raptor codes are application-layer FEC codes. Also, hybrid error-erasure channels have not been considered in [56]. In [57], application layer raptor codes are evaluated over DVB-H where shadowing and effects of memory are taken into consideration. However, [57] considers traditional non-cross-layer protocols which drop all corrupted packets and render pure erasure channels for the raptor codes. Finally, [57] assumes that raptor codes have a fixed two-percent overhead, which is not realistic.

Recently, two general cross-layer communication protocols, known as hybrid error-erasure protocols (HEEPs), are applied to Reed Solomon (RS) codes and low density parity check (LDPC) codes in wireless multimedia/video transmission [58]. These HEEPs allow corrupted packets to be relayed into the application layers. However, the protocols in [58] have not been applied to rateless raptor codes and do not model practical physical layer channels and the behavior of physical layer FECs. In addition, channel memory has not been considered in [58].

In this chapter, we first analyze and simulate the performance of actual raptor codes over BSCE and Gilbert-Elliott (GE) channels, which have not been studied in the literature. These are fundamental channel models that are formed by different cross-layer and non-cross-layer protocols. By using the rateless property of raptor codes, we reveal the

relationship between the overheads of raptor codes over BSC and BSCE channels; by simulating the performance of actual raptor codes over GE channels using iterative decoding, we demonstrate the desirable performance of raptor codes in channels with memory and investigate the effect of channel memory as well as the availability of channel state information (CSI).

We then investigate performance of raptor codes in MBMS file downloading services when either different cross-layer protocols or conventional protocols are applied. The raptor coded packets experience both packet erasures due to network congestion and packet corruptions due to wireless fading and noise. Channel memory in both the wireline channel and the wireless channel have been considered. By taking channel memory and the behavior of physical layer turbo codes into account, we model the channel that the raptor code experiences as a hierarchical Markov model. We derive the transition probabilities based on the turbo code rate and parameters of correlated Rayleigh fading. The main difference between this channel model and a regular Markov-type model (such as the well known GE channel) is the choice of channel states. Rather than the usual choice of good and bad states, the three states used in this model (erasure states, corrupt states and correct states) directly represent the results of the physical layer decoder. With this model, the two cross-layer protocols considered here only differ in the availability of side information about the instantaneous channel state. Therefore, based on this model, we are able to easily evaluate and compare the performance of different cross-layer and conventional protocols in channels with memory.

We also propose a hybrid erasure-soft raptor decoding scheme to implement cross-layer

protocols with respect to application layer raptor codes. The decoding scheme improves system performance substantially compared to that of conventional protocols [20] with modification required only on the receiver side. The main new idea of the decoding scheme is to perform traditional erasure decoding based on the correct packet transmission first and soft iterative decoding based on corrupted packets afterwards. In broadcasting applications, each user/receiver also has the flexibility to choose whether to use traditional erasure decoding or the proposed hybrid decoding depending on channel conditions and individual quality of service (QoS) requirements. Therefore, the hybrid decoding provides a flexible balance between performance and complexity. For example, when the channel quality is good and the physical layer code is able to correct all the errors in most packets, the receiver can recover all information using simpler traditional erasure decoding methods; when those non-corrupt packets are insufficient to decode all the source information, the receiver can collect soft information from corrupted packets to help in the decoding process. Third, we evaluate the system throughput using different turbo code rates and simulate performance of an actual raptor code using different protocols in various channel conditions.

The rest of the chapter is organized as follows. In Section 6.2, performance of raptor codes over BSCE channels is derived. Section 6.3 presents performance of raptor codes over GE channel using iterative decoding. In section 6.4, we first describe the overall MBMS system model and the application of cross-layer protocols to our system. We then illustrate the channel modeling and derive the transition probabilities based on the physical layer parameters. In Section 6.5 , we evaluate the application-layer capacity and maximum system throughput when different turbo code rates are used. In Section 6.6, we first provide

relevant background information of raptor codes. We then describe the proposed hybrid erasure-soft decoder for different cross-layer protocols. Section 6.7 shows the simulation results of raptor codes using conventional and cross-layer protocols in various channel conditions. Section 6.8 provides the main conclusions for this chapter.

6.2 Performance of raptor codes over BSCE channels

6.2.1 Hybrid error-erasure channels

The simplest form of hybrid error-erasure channel is the BSCE channel (Fig. 6.1). The

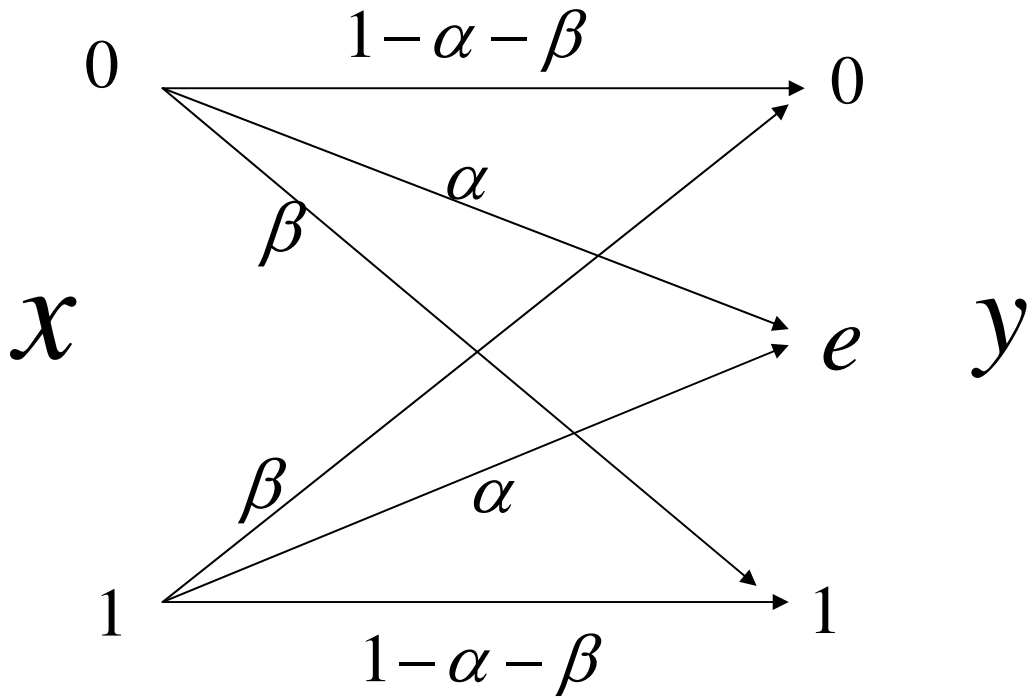

Figure 6.1. BSCE channel model with erasure probability α and bit error probability p

BSCE channel is a discrete-input, discrete-output channel with input alphabet $x \in \{0,1\}$ and output alphabet $y \in \{0,E,1\}$. The erasure probability $\alpha = Pr(y = E|x = 0) = Pr(y = E|x = 1)$, and the error probability $\beta = Pr(y = 1|x = 0) = Pr(y = 0|x = 1)$. Denote the probability that a bit error occurs conditioned on the case that the information bit is not erased as p. Then $p = \frac{\beta}{1-\alpha}$ and the information capacity of such a hybrid channel can be easily calculated as [5],

$$C = (1-\alpha)(1-h_b(p)), \tag{6.1}$$

where $h_b(p) = -p\log p - (1-p)\log(1-p)$ is the binary entropy function. In the following, we use $BSCE(\alpha, p)$ to represent a BSCE channel with erasure probability α and conditional error probability p.

6.2.2 Raptor codes over BSCE channels

As a rateless code, the performance of raptor codes over BSCE channels can be measured by the average overhead for which all the information bits are successfully decoded. To decode raptor codes, the receiver collects output bits and records the reliability of the bit as a measure of the amount of information received. Once the total amount of information received exceeds that of the source, the receiver starts to decode. If decoding fails, the receiver waits for a certain number of bits and again attempts to decode. For a raptor code with K information bits, let K_t represent the number of generated coded bits required for successive decoding. The overhead of raptor codes is defined as

$$\varepsilon = \frac{K_t - (K/C)}{(K/C)} = \frac{C}{R} - 1, \tag{6.2}$$

where C is the channel capacity and $R = K/K_t$ is the realized rate. Then the following property is satisfied:

Property 6.1: The average overhead of a raptor code (or LT code) required for successful decoding over a $BSCE(\alpha, p)$ channel is the same as that over a $BSC(p)$ channel.

Intuitively, the property can be interpreted as a compensation for erasure loss: the raptor codes need to generate more code bits by the factor corresponding to the erasure rate for $BSCE(\alpha, p)$ channels compared to that of $BSC(p)$ channels. However, the number of bits erased, successfully received as well as the ratio between them vary for each realization. Although the numbers and the patterns of bits discarded are irrelevant to the code performance, they are important in practice as they contribute to the time that the receiver needs to wait for successful decoding. A simple formal proof for this property is as follows:

Proof: Consider a rateless raptor code transmitted over a BSCE(α,p) channel. Denote S and X as random variables representing the numbers of coded bits received and the coded bits erased on successful raptor decoding, respectively. The average number of coded bits generated is then $\bar{K_t} = E\{S+X\}$. By taking the expectation conditioned on S, we have $\bar{K_t} = E\{S+X\} = E_S\{E\{S+X|S\}\} = E_S\{S+E\{X|S\}\}$. If an un-erased bit is considered as a successful trial and an erased bit is considered as a failed trial with failure probability equal to α, then conditional on the number of successful trials $S = s$, it is easily shown that X follows a negative binomial distribution which gives the probability of $s-1$ successes and x failures in $x+s-1$ trials, and success on the $(x+s)$-th trial, i.e.,

$$P_{X|S}(x|s) = \binom{x+s-1}{s-1} (1-\alpha)^s \alpha^x. \tag{6.3}$$

Since the average of the negative binomial distribution is easily computed as $E\{X|S\} = \frac{\alpha}{1-\alpha}S$, $E\{S+X\} = \frac{1}{1-\alpha}E\{S\}$. By substituting (6.1) into (6.2), the average overhead is obtained as $\bar{\varepsilon} = \frac{\bar{K_t}-(K/C)}{(K/C)} = E\{S\}(1-h_b(p))K^{-1} - 1$. Since erased bits do not contribute to code performance, and each coded bit is generated independently, $E\{S\}$ is the same for the $BSCE(\alpha, p)$ channel and $BSC(p)$ channel. Hence $\bar{\varepsilon}$ is also independent of erasure rate α. □

As a corollary to of Property 6.1, a raptor code has the same overhead for BEC channels of all erasure rates. This also confirms the property of "universality" of raptor codes over BEC channels [11].

6.2.3 Simulation results and analysis

Due to the decoding complexity, the performance in terms of realized rate is not simulated. As an alternative, the performance of raptor codes over $BSCE(\alpha, p)$ channels is measured by bit error rate (BER) versus the inverse of code rate R^{-1}, which is proportional to the number of coded bits generated. The decoding of raptor codes over BSCE channels is performed by using initial LLRs according to (2.5) and (2.6) and performing belief propagation (BP) decoding in the same way as for BSC channels. Unless mentioned otherwise, the raptor code used in this chapter is the same as the raptor code introduced in Section 2.1, with a code dimension $K = 9500$, LT degree distribution $\Omega_r(x)$ given by Eq. (2.4) and a rate 0.95 LDPC code as the precode. Fig. 6.2 shows the performance of raptor codes over $BSCE(\alpha, p)$ channels for different parameters α and p. The figure includes both *BEC* ($p = 0$) and *BSC* ($\alpha = 0$) as a special case. The capacity of BSCE channels is evaluated

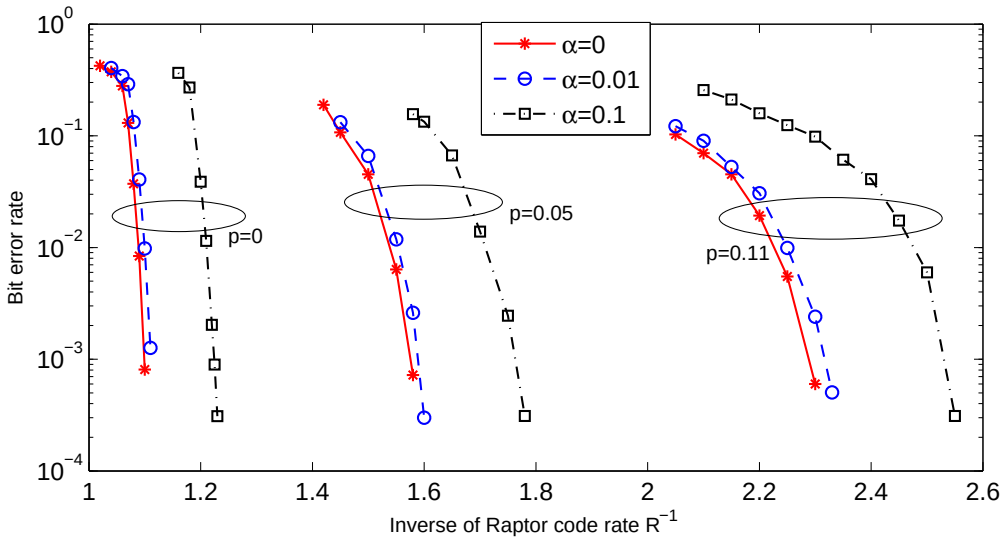

Figure 6.2. Raptor code performance over BSCE (α,p) channels with erasure rate α and error probability p.

by Eq. (6.1). From Fig. 6.2, it can be seen that to achieve an average BER of 10^{-2}, the raptor codes require approximately 9%,10% and 11% overhead for BEC, BSC($p = 0.05$), and BSC($p = 0.11$), respectively, compared to their own information capacity bounds. It can also be found that a BSCE(α,p) channel requires approximately the same overhead as a BSC(p) channel with the same error rate p. For Fig. 6.2, the BER curve of the BSCE(α,p) is not an exact horizontal shift of the BER curve of the BSC(p) channel. This is because for a fixed rate raptor code over BSCE(α,p) where $\alpha \neq 0$, the number of non-erased bits varies with each implementation, while this value does not change for BSC(p). The simulation demonstrates that in the case of channel with no memory, the variation of the number of received bits in different realizations has little effect on the performance of raptor codes.

6.3 Performance of raptor codes over GE channels

6.3.1 Gilbert-Elliott (GE) channels

GE channels are models for time varying binary symmetric channels containing a good state G and a bad state B (Fig. 6.3). For a given state, the channel can be modeled as

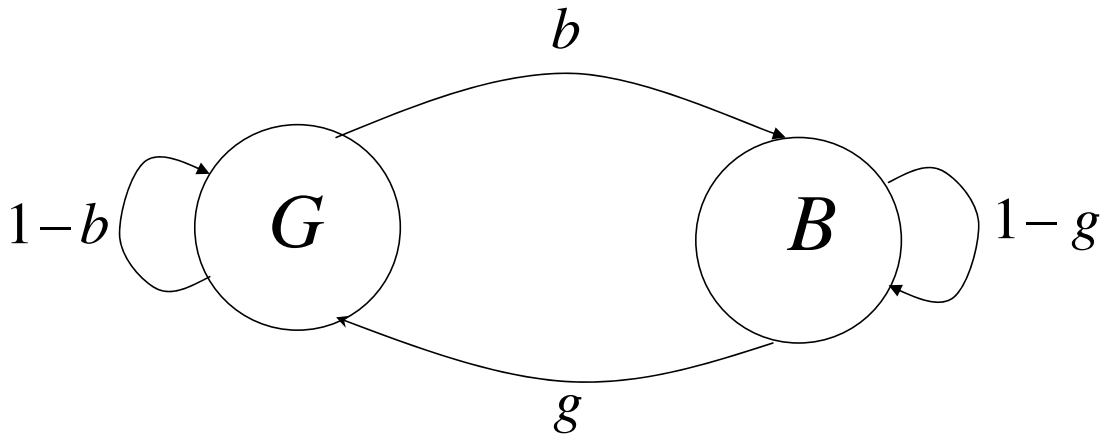

Figure 6.3. Structure of the Gilbert Elliott channel

a BSC channel. Let P_G and P_B represent the channel error probabilities in good and bad states, respectively, we have $0 \leq P_G \leq P_B \leq 1$. The transition between the two states form a binary Markov process. Let g and b represent the transition probabilities between state G and state B, $\{s_l\}_{l=0}^{\infty}$ represent the states at time l, $s_l \in \{G, B\}$. Then the stationary distribution can be obtained as $[\pi_g, \pi_b] = \left[\frac{g}{b+g}, \frac{b}{b+g}\right]$. The good-to-bad ratio ρ is defined as $\rho = P(s_l = G)/P(s_l = B) = g/b$. By induction on l, it can be easily verified that [59], for

$\xi \in \{G,B\}$, $P(s_l = \xi | s_0 = \xi) - P(s_l = \xi | s_0 \neq \xi) = (1-g-b)^l$. Denote parameter

$$\mu = 1 - g - b \tag{6.4}$$

as the channel memory ($\mu \in [0,1]$). When $\mu = 0$, the channel is memoryless, i.e., the current state is independent of previous states.

The capacity of GECs depends on the availability of channel state information (CSI) [1] at the receiver [59]. Let C_{CSI} represent the channel capacity when CSI is available at the receiver, i.e., when the receiver knows the current state. Let C_{NM} and C_μ represent the channel capacity when the channel has no memory, and when the channel memory is μ, respectively. For a fixed P_G, P_B and ρ, it is shown in [59] that

$$C_{NM} \leq C_\mu \leq C_{CSI}, \tag{6.5}$$

where C_{CSI} remains the same for different μ.

6.3.2 Simulation results and analysis

Fig. 6.4 shows the performance of raptor codes for the special case of GE channels where the probability of erasure is 0 in the good state and 1 in the bad state. These GE channels correspond to non-cross-layer protocols and can well model the bursty behavior of packet losses due to network congestion or packet drops due to corruptions. The average erasure rate P_{era}, which is equal to the steady-state probability of the bad state π_b, of the four curves are chosen to be the same ($P_{era} = 0.1$). It is not surprising that memory generally has a

[1] Throughout this chapter, CSI only refers to the information of the current channel state. The values of P_G and P_B are assumed to be known by the receiver in all cases.

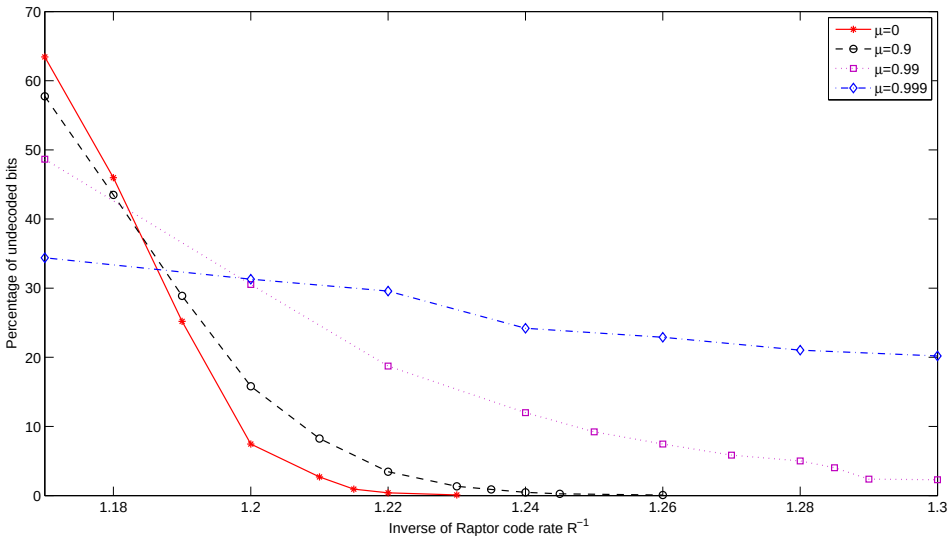

Figure 6.4. Raptor code performance over erasure channels with different amount of memory plotted as a function of the number of decoded bits

negative effect on the performance of raptor codes. The performance loss is a result of the distribution of bad states (erasures) over one code block. Although the average numbers of bad states are the same for the four different cases, a channel with higher correlation (memory) has higher variation of the number of bad states within one block length. The probability density function (PDF) of the number of bad states n_b that occur among n consecutive bits is provided in [60]. It is interesting to observe that at the left part of each curve, i.e., when most of the information bits are not able to be decoded, the channel with higher memory actually shows better performance than channels with no memory. This can be explained by the fact that raptor codes have a very steep performance curve in erasure channels. When not enough coded bits are received, most of the information bits cannot be

decoded. In this case, the higher variation of the number of bad states n_b is actually helpful.

Fig. 6.5 shows the performance of raptor codes over more general GE channels. When

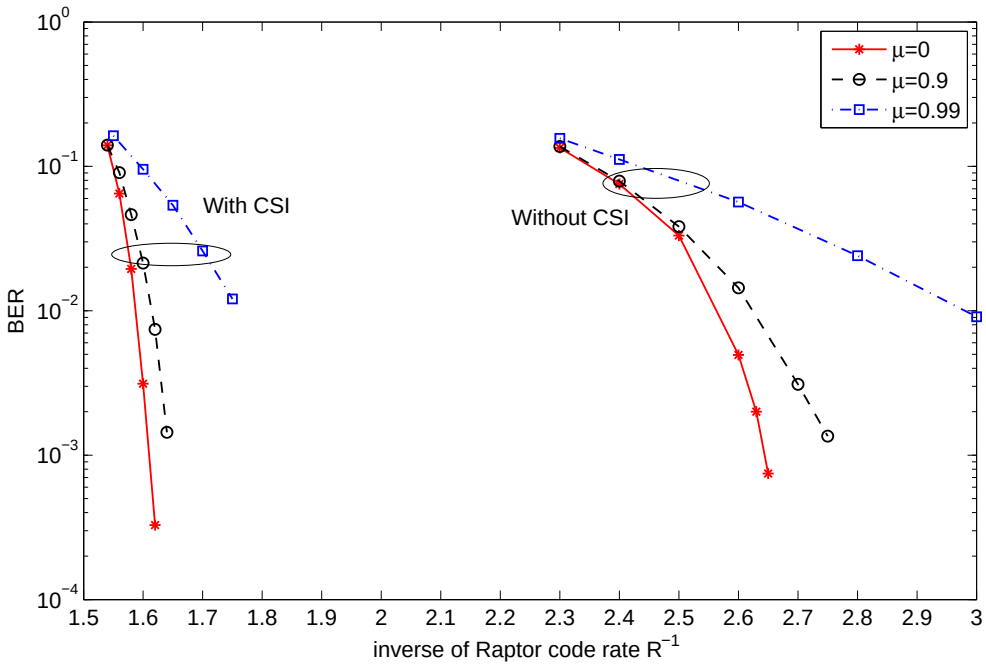

Figure 6.5. Raptor code over GE channels with and without channel state information. Here $P_G = 0.01$, $P_B = 0.5$ and $\rho = g/b = 3$.

CSI is available, the decoding is performed using iterative BP decoding with initial LLR equal to 0 for the code bit experiencing a bad state and $(-1)^y \ln((1-P_b)/P_b)$ for the code bit experiencing a good state. For GE channels when no CSI is available, the initial LLRs for all code bits are set according to the average error probability over all the states. It can be seen from Fig. 6.5 that there is a significant difference in code performance between the case when CSI is available and the case when it is not. It can also be observed that memory has a negative effect on performance regardless of the availability of CSI. The reason is the

same as that for the case of erasure channels; memory increases the variation of the number of bad states in a given block length.

The results in Fig. 6.5 seem to be contrary to the capacity study [59] which shows that memory increases capacity when CSI is not available. This suggest that decoding performance can be improved. To exploit the improved capacity, the decoder needs to utilize channel correlation for better estimation of CSI [59]. This improvement may possibly be obtained by employing similar estimation and decoding techniques used for LDPC codes, though applying such decoding methods to raptor codes is beyond the scope of this thesis. Since no attempt is made to estimate CSI, the performance of raptor codes over GE channels with no CSI is still bounded by C_{NM}.

6.4 System and channel models

6.4.1 System model and cross-layer protocols

Two layers of FEC are used in multimedia broadcast and multicast service (MBMS): turbo codes in the physical layer and raptor codes in the application layer [20]. Since the protocol stack in MBMS systems is overly detailed to present here, a much simpler two layer model that captures the essential performance is considered. In our model, the information data are first segmented into data-bearing packets. Multiple data packets are coded by a raptor code where each packet is considered as a bit (a vector of binary bits) of a raptor code. Cyclic redundancy checks (CRC) and packet header information are then appended to each output packet to form the transmitted packets. Each packet is further protected by a physical

layer turbo code, modulated by BPSK and transmitted over the physical channel.

The packets experience a hybrid type of channel where transmitted packets can be lost due to network congestion. Packets that are not lost are still subjected to channel fading and noise. When the packet is not lost, the receiver first demodulates and decodes data using the turbo decoder. The correctness of the turbo decoded output is checked by the CRC embedded in each packet. In the current MBMS standard, the entire packet is dropped if

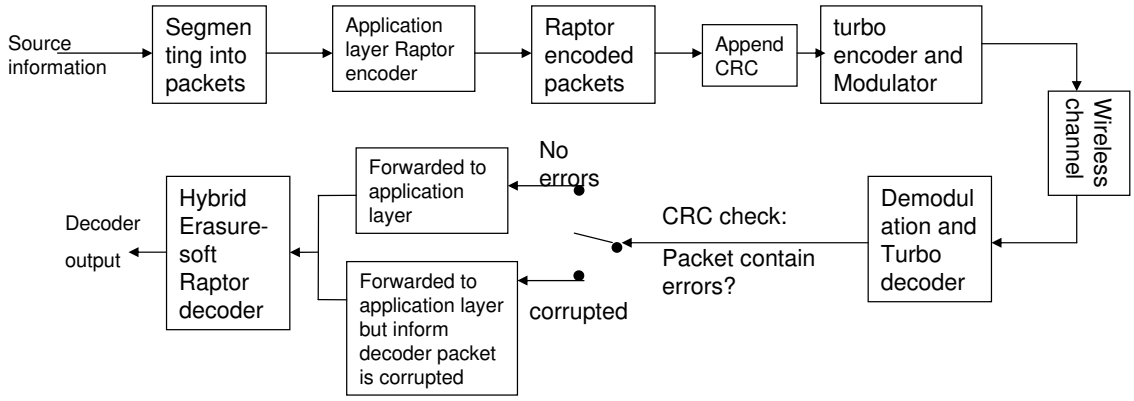

Figure 6.6. Cross-layer decoder with side information (CLDS) system diagram.

the CRC fails, which we denote as the conventional (CON) scheme. Therefore only packets that do not contain errors are forwarded to the raptor decoder. Two general cross-layer protocols, known as cross-layer design (CLD) and cross-layer design with side information (CLDS), are proposed in [58]. For the CLD protocols, CRC information is simply ignored and all turbo decoder outputs are forwarded to the raptor decoder. For the CLDS protocols, all outputs of the turbo decoder are forwarded to the raptor decoder along with the side information provided by the CRC indicating whether the packet is corrupted. The system diagram for the CLDS protocol is illustrated in Fig. 6.6.

6.4.2 Channel modeling

Packets are lost in bursts when network congestion is severe. Therefore a more accurate model for packet loss should take channel memory into account. We model the behavior of packet losses as a GE channel, a well known two state Markov model for modeling channels with memory (Fig. 6.3). The transition between the two states form a binary Markov process. We consider a special case of GE channel where in the bad state (erasure state), the packet loss probability is 1 while in the good state (non-erasure state), the packet loss probability is 0, which is equivalent to $P_G = 0, P_B = 0.5$. Let g_1 and b_1 represent transition probabilities from bad state to good state, and from good state to bad state, respectively. The average packet loss rate $\lambda = b_1/(g_1 + b_1)$. The channel memory is defined as $\mu_1 = 1 - g_1 - b_1$ [59]. The two parameters λ and μ_1 determine g_1 and b_1 and the packet loss behavior.

The physical layer wireless channel is assumed to be a correlated Rayleigh fading with Doppler frequency f_d and average received SNR $\bar{\gamma}$. The "water-fall" region of the turbo code is narrow [20] and has the following property: for a given rate R_{turbo}, there exists a *SNR* threshold γ_t such that when the channel SNR $\gamma > \gamma_t$, the turbo decoder almost always decodes the information correctly; and when $\gamma < \gamma_t$, the decoder almost always fails due to errors in the decoder output. The cutoff rate of the turbo coder satisfies $R_{turbo}(\gamma_t) = 1 - \log_2(1 + \exp(-\gamma_t))$ [20]. Hence for a given turbo code rate R_{turbo},

$$\gamma_t = -\ln(2^{1-R_{turbo}} - 1). \tag{6.6}$$

To model the combination of correlated fading channels and the two cross-layer protocols,

a good state is used to represent the case when the instantaneous channel SNR $\gamma > \gamma_t$, while a bad state represents the case when $\gamma < \gamma_t$. In the good state (correct state), the turbo code always decodes the information correctly and the CRC succeeds. In the bad state (corrupt state), there are errors present in the turbo decoder output and the CRC fails.

To match the two-state GE channel to the correlated Rayleigh fading, the steady-state probability $\pi_b = \int_0^{\gamma_t} f_\gamma(\gamma) d\gamma = 1 - \exp(-\gamma_t/\overline{\gamma})$ where $f_\gamma(\gamma) = \frac{1}{\gamma}\exp(-\gamma/\overline{\gamma})$ is the PDF of the instantaneous SNR of Rayleigh fading channels. Next, by matching the average fade duration, e.g., the time the fading amplitude is below the threshold to the average time of the GE channel staying in the bad state, it can be shown that [60],

$$g_2 = \frac{\sqrt{\gamma_t/\overline{\gamma}} f_d T \sqrt{2\pi}}{\exp(\gamma_t/\overline{\gamma}) - 1} \tag{6.7}$$

$$b_2 = \sqrt{\gamma_t/\overline{\gamma}} f_d T \sqrt{2\pi} \tag{6.8}$$

where g_2 and b_2 are the transition probabilities of the "fading" GE channels, f_d is the Doppler frequency, T is the packet duration and $f_d T$ is the normalized Doppler frequency.

The overall channel for the application-layer raptor codes can be represented as a hierarchical Markov channel model (Fig. 6.7). At the higher level, the channel can be in the erasure state (packet loss) or the non-erasure state with transition probabilities g_1 and b_1, respectively where the erasure probability is 1 in the erasure state and 0 in non-erasure state. Conditional on the event that the packet is not erased, the channel is a GE channel with transition probabilities g_2 and b_2, where the error probabilities in the good (correct) state and bad (corrupt) state are 0 and ε, respectively, where ε is also termed the *packet corruption level*. Note that g_1 and b_1 are independent of g_2 and b_2 because the packet

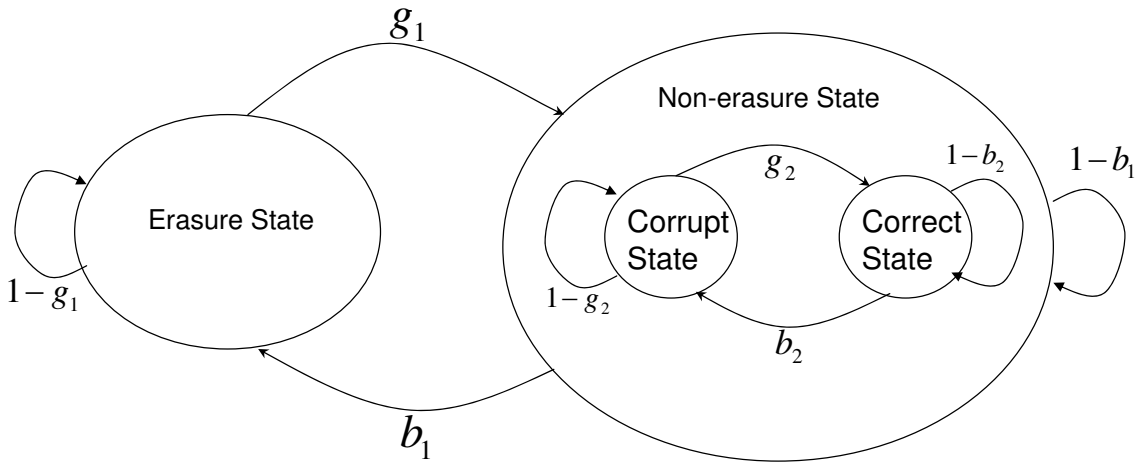

Figure 6.7. The hierarchical markov model for cross-layer protocols

loss in this channel model is caused by network congestion as opposed to packet header corruption assumed in [58].

6.5 Capacity and system throughput evaluation

The capacity of CON, CLD and CLDS protocols for memoryless channels have been summarized in [58]. Let δ and λ represent the packet dropping rate in the CON and CLD schemes, respectively. Let *p* represent the probability that an error occurs in a data bit in an unerased packet. The packet corruption level ε is equal to the conditional probability that an error occurs in a random data bit in an unerased packet given that the CRC fails. The capacity of the three schemes for memoryless channel can be easily obtained as [58],

$$C_{CON}^{NM} = 1 - \delta \tag{6.9}$$

$$C_{CLD}^{NM} = (1-\lambda)(1-h_b(p)) \tag{6.10}$$

$$C_{CLDS}^{NM} = (1-\delta) + (\delta-\lambda)(1-h_b(\varepsilon)) \tag{6.11}$$

where the superscript NM represents no memory and $h_b(.)$ is the binary entropy function, defined as in Eq. 6.1 as in $h_b(p) = -p\log p - (1-p)\log(1-p)$, and where

$$p = (\delta-\lambda)\varepsilon/(1-\lambda). \tag{6.12}$$

In the channel model of Fig. 6.7, CLD and CLDS schemes only differ by the availability of instantaneous channel state at the receiver. In [59], it is shown that channel memory does not change the capacity of GE channel with erasure states for the same average erasure rate. As a consequence, we can conclude

$$C_{CON} = C_{CON}^{NM}. \tag{6.13}$$

As CLD scheme using the hierarchical Markov model is essential a CLDS scheme without CSI, from (6.5), we obtain

$$C_{CLD}^{NM} < C_{CLD} < C_{CLDS}^{NM}. \tag{6.14}$$

Finally, receivers using CLDS scheme is fully aware of CSI, i.e., they know whether the channel is in erasure, correct or corrupt state. Since C_{CSI} is not affected by channel memory, we have

$$C_{CLDS} = C_{CLDS}^{NM}. \tag{6.15}$$

Nevertheless, as the raptor decoder does not attempt to estimate side information of the instantaneous channel state, the performance of raptor codes over these decoding schemes

is still bounded by the capacity for the case of no memory. Eqs. (6.9), (6.10) and (6.11) can be used to evaluate C_{CON}^{NM}, C_{CLD}^{NM} and C_{CLDS}^{NM}. By definition, $1-\delta$ is steady state probability of the correct state in the hierarchical Markov model while $1-\lambda$ is the steady state probability of the non-erased state. Therefore, $\delta = 1-(1-\lambda)\frac{g_2}{b_2+g_2}$.

The application layer capacity provides a bound to the performance of raptor codes. However, the application layer capacity does not take into account the extra protection overhead used in the physical layer to protect the information bits. To compare the system performance using different turbo code rates, we use the maximum system throughput which is equal to $C \times R_{turbo}$, where C is the application layer capacity as a performance metric. Fig. 6.8 shows a comparison of the maximum achievable system throughputs of each of the three schemes as a function of channel SNR for two different turbo code rates. It is quite obvious that the proposed hybrid scheme using the CLDS protocol can achieve much higher throughput over most of the SNR range. When channel SNR is very high, the difference becomes negligible. We also observe that a higher turbo code rate is preferable except for the extremely low SNR regime. We remark that [20] reaches similar conclusions about the combinations of the turbo code and raptor code rates, but the results there are only limited to simulations in the CON scheme.

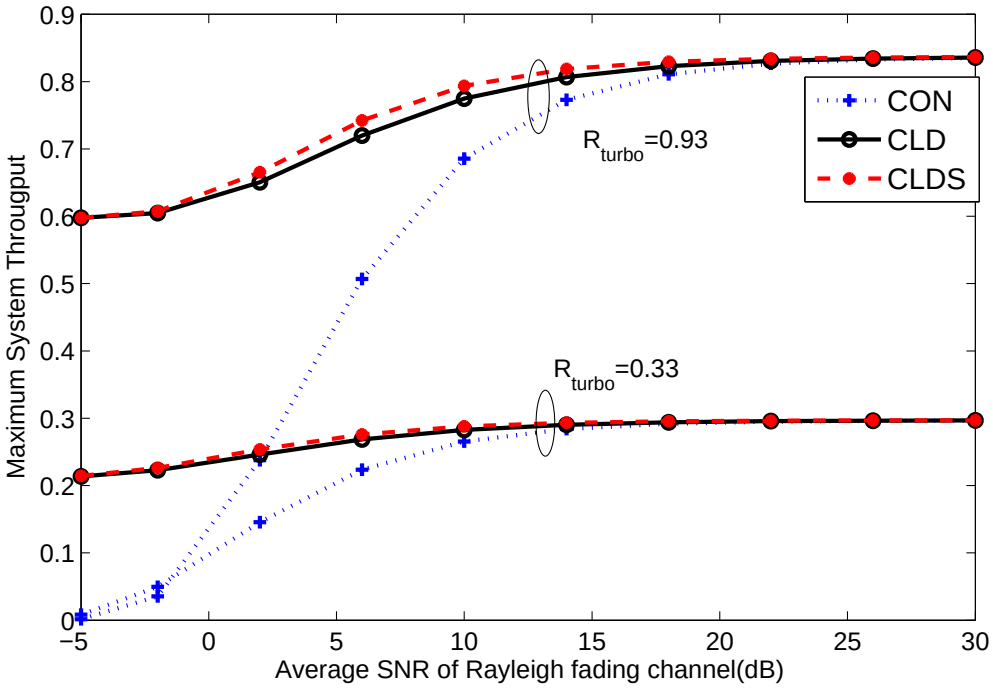

Figure 6.8. Maximum system throughput as a function of channel SNR for the fading channel in Fig. 6.7 with parameters $\lambda = 0.1$, $\mu_1 = 0.9$, $fdT = 0.01$, $\varepsilon = 0.05$.

6.6 Raptor codes and the hybrid erasure-soft decoder

The first practical realization of fountain codes is known as the class of Luby Transform (LT) codes [9] that encode k information bits $(x_1, x_2, \ldots, x_k)$ into a potentially infinite number of output bits $(z_1, z_2, z_3, \ldots)$. The encoding process is performed by first sampling a probability distribution Ω and a degree of d distinct information bits are then chosen uniformly at random from the k input bits. The value of each output bit is the modulo 2 bit-wise summation of the d chosen input bits. The output bit stream is generated independently until the transmitter receives an acknowledgement (ACK) of successful decoding

from the receiver of successful decoding or until a predesigned code rate is achieved. The degree distribution Ω is described by its generating polynomial $\Omega(x) = \Sigma_{i=1}^{k} \Omega_i x^i$, where Ω_i represents the probability that value i is chosen. Shokrollahi [11] extended the idea of LT codes to raptor codes to reduce the decoding complexity to be linear for binary erasure channels (BEC). A raptor code with parameters (k, C, Ω) is constructed by concatenating a block code C with a LT code with degree distribution Ω. To encode a raptor code, the precoder C first encodes k information bits into $\tilde{k}$ intermediate bits. The output bit streams are then generated by applying the inner LT code on the $\tilde{k}$ intermediate bits.

Decoding of the raptor codes for a binary symmetric channel (BSC) can be performed iteratively using belief propagation (BP) algorithms over the Tanner graph of the raptor code [23]. For the BEC, the BP algorithm can be significantly simplified, which allows for linear decoding complexity of raptor codes [11]. In this paper, we term the decoding method for BEC as erasure decoding, and iterative decoding that uses soft information as soft decoding. Note that the complexity of erasure decoding is much lower than soft decoding.

To implement a cross-layer protocol for raptor coding in the MBMS system, we propose a hybrid erasure-soft decoder. The hybrid decoder works as follows:

Step 1) The Tanner graph of the raptor code is constructed as shown in Fig. 6.9. For each LT encoded bit, a corresponding check node is added to form a Tanner graph of the LT code. In the final Tanner graph, there are two types of variable nodes (input variable nodes and output variable nodes) and two types of check nodes (LDPC check nodes and LT check nodes).

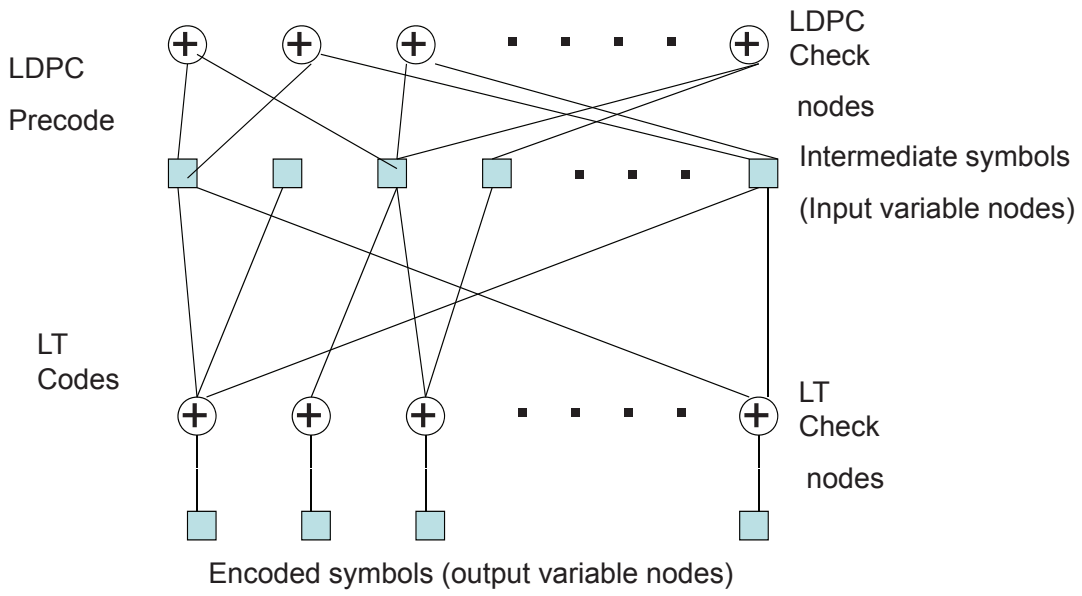

Figure 6.9. Tanner graph of raptor code

Step 2) The LLRs of the variable nodes are initialized. The initial LLRs for input variable nodes are all set to 0 because they have not been transmitted. For all output variable nodes connected to a packet that is lost or discarded, the initial LLRs should also be set to 0.

In the CLD scheme, since the CRC is turned off, the raptor decoder does not know whether the channel is in the correct state or the corrupt state, i.e, the decoder does not know the instantaneous channel state. Therefore, the decoder treats the channel as a BSC with crossover probability p at non-erasure states, where p is given by (6.12). Hence the decoder will set the initial LLRs of output bits to 0 for the erasure state and $(-1)^y \ln((1-p)/p)$ [30] for the non-erased state, where $y \in \{0,1\}$ is the physical layer decoder output.

In the CLDS scheme, the receiver knows which state the current channel is in. Therefore, the decoder sets the initial LLR to 0 for the erasure state, $(-1)^y \cdot \infty$ for the correct state, and $(-1)^y \ln((1-\varepsilon)/\varepsilon)$ for the corrupt state.

Step 3) The decoder eliminates all the nodes and edges that are associated with encoded bits that are in the erasure state since they provide zero reliability.

Step 4) Based on the value of all the encoded bits in the correct state, the decoder performs erasure decoding on the decoding Tanner graph. Any information bit that can be decoded and any edges associated with these decoded nodes are removed from the graph. In CLD scheme, this step is not performed because the receiver does not identify the correct states using CRC.

Step 5) Iterative BP decoding based on LLRs from the corrupt state is performed on the remaining graph. Because the number of remaining edges is smaller than that of the original decoding graph, the decoding complexity of the hybrid scheme is simpler than the traditional iterative decoding scheme. The updating equation for the BP algorithm is the same as that used for LDPC codes [30].

6.7 Simulation results

To simulate the actual performance of raptor codes, the raptor code described in [23] is used. The pre-code of this raptor code is a left regular and right Poisson LDPC code with rate 0.95, and the variable nodes of this LDPC code have constant degree of 4. The code

dimension $k = 9500$ and the inner LT codes use the degree distribution,

$$\begin{aligned}\Omega(x) = {} & 0.007969x + 0.493570x^2 + 0.166622x^3 \\ & + 0.072646x^4 + 0.082558x^5 + 0.056058x^8 + 0.037229x^9 \\ & + 0.055590x^{19} + 0.025023x^{65} + 0.003135x^{66}. \end{aligned} \tag{6.16}$$

Fig. 6.10 depicts the performance of raptor codes of the three communication schemes with channel memory for different raptor code rates. It can be seen that the CLDS and CLD schemes perform significantly better than the CON scheme. To achieve an average BER of 10^{-2}, the difference between the number of raptor coded bits that needs to be generated in the CON and CLDS schemes is approximately 21% when the corruption level $\varepsilon = 0.02$ and 18% when $\varepsilon = 0.05$. The CLDS scheme performs slightly better than the CLD scheme and the difference between their performance increases as the corruption level increases. It can be seen that the gap between the CLD and CLDS schemes is 0.6% for $\varepsilon = 0.02$ and 1.5% for $\varepsilon = 0.05$ to achieve a BER of 10^{-2}. The raptor codes require less than 12% overhead for CLD and CLDS schemes to achieve a BER of 10^{-2} compared to their own capacity bounds obtained by evaluating (6.10) and (6.11), respectively. It should be noted that among the three schemes, CLD has the highest decoding complexity. Compared to CLD, CLDS not only provides additional performance gain, but also reduces decoding complexity. It can also be observed that the performance curve of the raptor code is very steep. Therefore, the rateless property of raptor codes is very important to provide the flexibility of different code rates to accommodate different channel conditions.

Fig. 6.11 shows the effect of channel memory on the three different schemes caused by

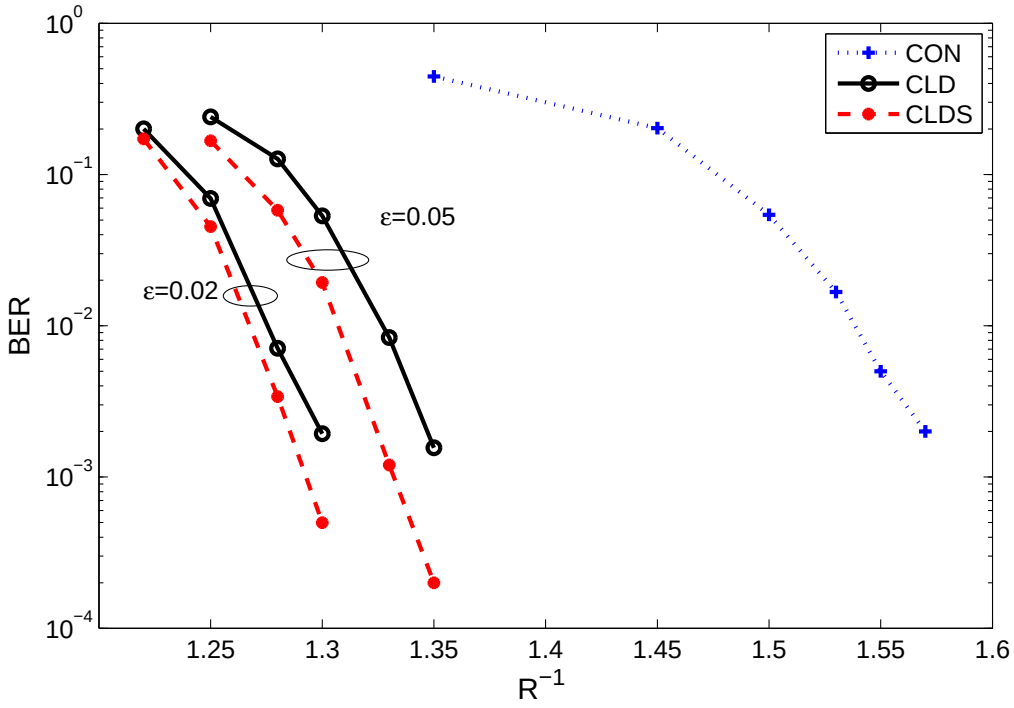

Figure 6.10. Raptor code over hybrid error-erasure fading channels with parameters $\lambda = 0.1$, $\mu_1 = 0.9$, $R_{turbo} = 0.93$, $\overline{\gamma} = 10dB$, $f_d T = 0.01$.

fading correlation. It can be observed that memory decreases the performance of the cross-layer schemes. However, the cross-layer protocols are quite robust to fading correlation as the effect of memory is only significant for CLD and CLDS schemes when the normalized Doppler frequency is below 0.01. This can be explained by the fact that packet corruption only results in a small probability of error for a particular bit inside a packet.

Figs. 6.12 and 6.13 show the influences of channel SNR and corruption level ε. An increase of average SNR decreases the average number of corrupt states, and hence improves the performances of all three schemes. It can be observed that the performances of CLD

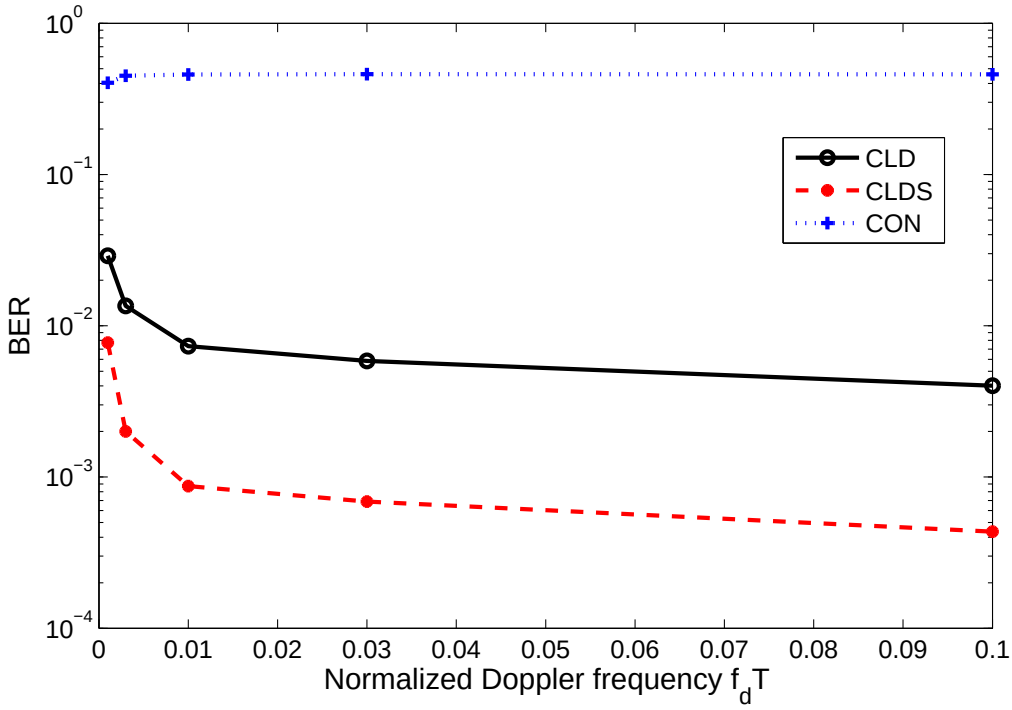

Figure 6.11. Performance of CLD, CLDS and CON schemes as a function of normalized doppler frequency with parameters $\lambda = 0.05$, $\mu_1 = 0$, $R_{turbo} = 0.93$, $\bar{\gamma} = 10dB$, $\varepsilon = 0.05$.

and CLDS are less sensitive to SNR than CON. The differences between the SNR requirement to achieve BERs of 10^{-1} and 10^{-2} is approximately 7dB for CLD and CLDS, and 5dB for CON. This also shows that the combination of an application layer raptor code and a physical layer code is very robust to variations in channel quality, as a significant drop in channel SNR can be compensated by a slightly lowered raptor code rate. The corruption level also has a significant impact on the performance of CLD and CLDS schemes. As shown in Fig. 6.13, for the same raptor code rate and with all the other parameters equal, the performances of CLD and CLDS are reasonable at a corruption level of 0.005 (BER

below 10^{-2}) but very poor at a corruption level above 0.5. The change of corruption level does not affect the performance of CON since it does not change the average number of corrupt states.

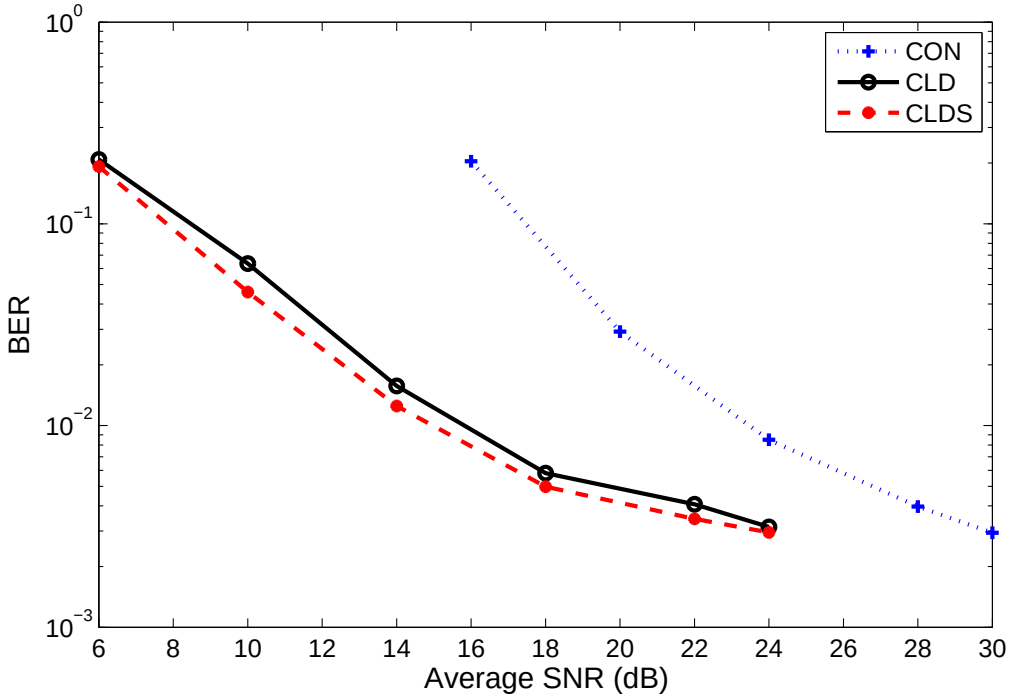

Figure 6.12. Performance of raptor codes in correlated fading channels as a function of different SNR with parameters $\lambda = 0.1$, $\mu_1 = 0.9$, $R_{turbo} = 0.93$,$\varepsilon = 0.02$, $R^{-1} = 1.25$, $fdT = 0.01$.)

6.8 Conclusion

This chapter proposes a hybrid erasure-soft decoding scheme for application-layer raptor codes used in broadcasting services with cross-layer protocols. By taking channel memory

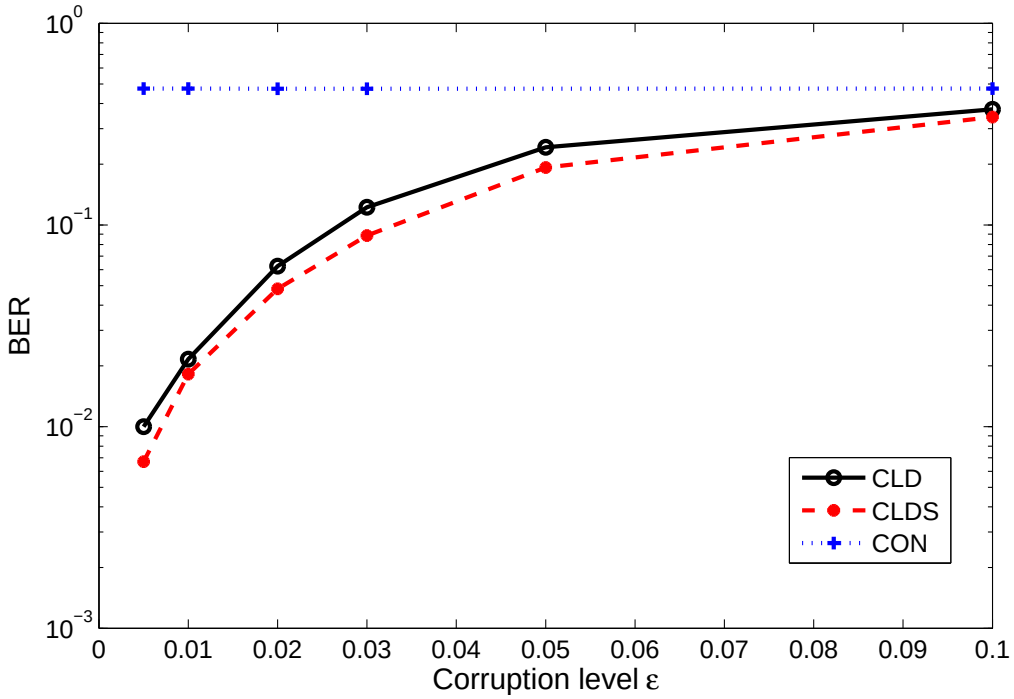

Figure 6.13. Bit error rate performance of raptor codes using CLD, CLDS and CON schemes as a function of corruption level with parameters $\lambda = 0.1$, $\mu_1 = 0.9$, $R_{turbo} = 0.93$, $\overline{\gamma} = 10dB$, $R^{-1} = 1.25$, $fdT = 0.01$.

into account, the composite channel is modeled by a hierarchical Markov model which includes erasure, correct and corrupt states. For this channel model, the CLD and CLDS schemes differ only by the availability of side information about instantaneous channel state. The proposed cross-layer decoding schemes outperform conventional (CON) scheme using erasure decoding significantly. The difference in the number of raptor coded bits required to achieve the same BER of 10^{-2} for the CLDS and CON schemes can be 20% as shown in Fig. 6.10. Channel correlation decreases the performance of raptor codes for

all three schemes and the impact is significant when the normalized Doppler frequency is small. The effect of the choice of turbo code rate on the system throughput is also discussed.

Chapter 7

Conclusions and Future Work

In this chapter, we summarize the major contributions in this thesis and suggest several topics for future research.

7.1 Conclusions

In this thesis, we address two major challenges of the application of rateless codes over wireless multimedia multicast. To address the challenges due to user heterogeneity, we propose UEP rateless code designs and present an optimization framework and different solutions depending on the constraints of the multimedia multicast system.

In Chapter 3, two general problems are formulated for optimizing unequal error protection (UEP) rateless codes for scalable multimedia multicasting systems with heterogeneous users. The design objective is to either minimize transmission overhead for guaranteed quality of service (QoS) or provide best-effort QoS for a given transmission overhead. A random interleaved UEP raptor code design is proposed that can take advantage of the high

performance of existing standardized raptor codes. The formulated problem is converted into a convex optimization problem which can be solved analytically. Numerical results demonstrate that the optimized proposed UEP raptor codes perform better than existing UEP raptor code designs when the same degree distribution and iterative decoding is applied. Large additional gains for the proposed UEP scheme can be obtained by using the superior existing standardized raptor codes which also takes advantage of a corresponding efficient maximum likelihood (ML) decoder.

In Chapter 4, we show that the problem of optimal allocation of priority encoding transmission (PET) based rateless codes for asynchronous multimedia multicast [25] [26] can be transformed into a convex optimization problem when integer constraints are relaxed. Because the problem is convex and the total number of parameters for the optimization is equal to the number of user classes, the solution simplifies. In addition, an analytical solution is found for the case when there are no outage constraints [25]. For the more general formulation with outage constraints [26], an analytical solution is found for the special case of two user classes. Numerical methods of convex optimization software [27] [28] are required to solve the more general cases.

To address the second challenge due to the time varying wireless channel, we first find the optimal combination of physical layer code rate and application layer code rate that maximize overall throughput. We then propose a hybrid erasure-soft decoder of rateless codes for wireless channels with memory to improve performance and reduce complexity.

In Chapter 5, the balance between physical layer rate and application-layer raptor code rate for rateless coded communication systems is investigated. Both slow and fast Rayleigh

fading have been considered. Unlike traditional approaches which choose physical layer modulation and coding pairs to meet a target outage probability, modulation and coding is optimally chosen in our proposed cross-layer scheme in order to maximize system throughput. In addition, a cross-layer adaptive modulation and coding design is proposed for the slow fading scenario. The proposed cross-layer design shows significantly higher throughput compared to traditional non cross-layer design. It can also be seen from this study that allowances for high packet error rates to be corrected by application-layer erasure codes can be efficient in many situations.

In Chapter 6, performance of raptor codes over BSCE and GE channels are evaluated, where it is first shown that the average overhead of raptor codes over BSCE and BSC channels with the same cross-over probabilities are the same. A hierarchical Markov model is proposed which includes erasure, correct and corrupt states to model hybrid erasure-error channels with memory. A cross-layer hybrid erasure-soft decoder for raptor code is proposed. It has been shown that the cross-layer hybrid decoder outperforms the erasure decoder, which also takes advantage of the build-in CRC mechanism that exists in current protocols.

7.2 Future Work

There are several areas that are suggested for future research:

- In Chapter 3, the best-effort QoS formulation is only solved numerically. It is thus of interest for future studies to find a systematic and lower complexity method to

solve the best-effort QoS problem, which is to find the optimal selection probabilities of proposed UEP rateless code that maximize the average fidelity measure of multimedia content for multimedia multicasting over heterogeneous users.

- In Chapter 4, for the general asynchronous multicast optimization problem with outage constraints, an analytical solution is not found despite that it is a convex optimization problem. It would be of interest if a similar analytical solution can be found as in the case without outage constraints.

- In Chapter 5, the results are primarily presented for single user cases. It is thus of interest to investigate the optimal combination of physical layer code rate and application layer code rate in multicast, multiple access and other multi-user scenarios. In addition, the results are presented for slow and fast fading channels. It would be of interest to provide analytical result for correlated fading channels with a given normalized Doppler frequency.

- In Chapter 6, although the relationships among the capacities of three cross-layer and conventional schemes are expressed in relative terms, it would be of interest to derive the exact capacity values as a function of channel parameters.

- Rateless codes also have potential applications in relay channels. It would be interesting to investigate the optimal degree distribution of a rateless code design when utilized in relay channels.

Bibliography

[1] N. Rahnavard, B. N. Vellambi, and F. Fekri, "Rateless codes with unequal error protection property," *IEEE Trans. Inform. Theory*, vol. 53, no. 4, pp. 1521–1532, April 2007.

[2] D. Sejdinovic, D. Vukobratovic, A. Doufexi, V. Senk, and R. Piechocki, "Expanding window fountain codes for unequal error protection," *IEEE Trans. on Communications*, vol. 57, no. 9, pp. 2510–2516, Sept 2009.

[3] D. J. C. MacKay, "Fountain codes," *IEE proceedings*, vol. 152, no. 06, pp. 1062–1068, Dec. 2005.

[4] L. Shu and D. Costello, *Error Control Coding: Fundamentals and Applications*. Prentice-Hall Inc., Upper Saddle River, NJ, 2004.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, New York, 1990.

[6] L. Rizzo, "Effective erasure codes for reliable computer communication protocols," *Comput. Commun. Rev.*, vol. 27, no. 2, pp. 24–36, April 1997.

[7] J. Nonnenmacher, E. W. Biersack, and D. Towsley, "Parity-based loss recovery for reliable multicast," *IEEE/ACM Trans. Networking*, vol. 6, no. 4, pp. 349–361, Aug. 1998.

[8] J. W. Byers, M. Luby, M. Mitzenmacher, and A. Rege, "A digital fountain approach to reliable distribution of bulk data," in *Proc. ACM SIGCOMM*, Vancouver, BC, Canada, 1998, pp. 56–57.

[9] M. Luby, "LT-codes," in *Proc. 43rd Annu. IEEE Symp. Foundations of Computer Science (FOCS)*, Vancouver, BC, Canada, Oct. 2002, pp. 271 – 280.

[10] A. Shokrollahi, "Raptor codes," Digital Fountain, Technical Report, DR2003-06-01, available at http://algo.epfl.ch/pubs/raptor.pdf, 2003.

[11] ——, "Raptor codes," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2551–2567, June 2006.

[12] J. Byers, M. Luby, and M. Mitzenmacher, "Accessing multiple mirror sites in parallel: Using tornado codes to speed up downloads," in *Proc. INFOCOM*, Mar. 1999, pp. 275–283.

[13] J. W. Byers, M. Luby, and M. Mitzenmacher, "A digital fountain approach to asynchronous reliable multicast," *IEEE J. Select. Areas Commun.*, vol. 20, no. 8, pp. 1528–1540, Oct. 2002.

[14] "3rd Generation Partnership Project;Technical Specification Group Services and System Aspects; Multimedia Broadcast/Multicast Service (MBMS);Protocols and codecs," *3GPP TS 26.346 V8.3.0*, June 2009.

[15] A. Mohr, E. Riskin, and R. Ladner, "Unequal loss protection: graceful degradation of image quality over packet erasure channels through forward error correction," *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, pp. 819 – 828, May 2000.

[16] P. A. Chou, H. J. Wang, and V. N. Padmannabhan, "Layered multiple description coding," in *Proc. Packet Video Workshop*, 2003.

[17] P. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Trans. Multimedia*, vol. 8, no. 2, pp. 390–404, Apr. 2006.

[18] A. Talari and N. Rahnavard, "Unequal error protection rateless coding for efficient MPEG video transmission," in *Proc. Military Communication Conference*, Oct. 2009, pp. 1–7.

[19] D. Vukobratovic, V. Stankovic, D. Sejdinovic, L. Stankovic, and Z. Xiong, "Scalable video multicast using expanding window fountain codes," *IEEE Trans. on Multimeida*, vol. 11, no. 6, pp. 1094–1104, Oct 2009.

[20] M. Luby, T. Gasiba, T. Stockhammer, and M. Watson, "Reliable multimedia download delivery in cellular broadcast network," *IEEE Tran. on Broadcasting*, vol. 53, no. 1, pp. 235–246, March 2007.

[21] B. Furht and S. Ahson, *Handbooks of mobile broadcasting: DVB-H, DMB, ISDB-T, and mediaFLO*. 1st ed. Auerbach Publications, 2008.

[22] O. Etesami and A. Shokrollahi, "Raptor codes on binary memoryless symmetric channels," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2033–2051, May 2006.

[23] R. Palanki and J. S. Yedidia, "Rateless codes on noisy channels," in *Proc. IEEE International Symposium on Information Theory*, Chicago, IL, Jun. 2004, p. 37.

[24] R. Y. S. Tee, T. D. Nguyen, L. L. Yang, and L. Hanzo, "Serially concatenated luby transform coding and bit-interleaved coded modulation using iterative decoding for the wireless interenet," in *Proc. Vehical Technology Conference*, May 2006, pp. 22–26.

[25] W. Sheng, W. Y. Chan, S. D. Blostein, and Y. Cao, "Asynchronous and reliable multimedia multicast with heterogeneous QoS constraints," in *Proc. IEEE Int. Conf. on Communication*, May 2009, pp. 1–6.

[26] W. Sheng, W. Y. Chan, and S. D. Blostein, "Rateless code based multimedia multicasting with outage probability constraints," in *Proc. 25th Biennial Symposium on Communications*, May 2010, pp. 134–138.

[27] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, Cambridge, 2004.

[28] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 1.21,” http://cvxr.com/cvx, Feb. 2011.

[29] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Johns Hopkins University Press, Baltimore, Maryland, 1983.

[30] W. Ryan, *An introduction to LDPC codes*. CRC Handbook for Coding and Signal Processing for Recoding Systems (B. Vasic, ed.), CRC Press, 2004.

[31] C. Berrou, A. Glavieux, and P. Thitimajshima, “Near shannon limit error-correcting coding and decoding: Turbo codes,” in *Proc. IEEE International Conference on Communications*, Geneva, Switzerland, May 1993, pp. 1064–1070.

[32] J. Hagenauer, E. Offer, and L. Papke, “Iterative decoding of binary block and convolutional codes,” *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 429–445, March 1996.

[33] J. Castura and Y. Mao, “Rateless coding over fading channels,” *IEEE Commun. Lett.*, vol. 10, no. 1, pp. 46–48, Jan. 2006.

[34] K. Hu, J. Castura, and Y. Mao, “Performance-complexity tradeoffs of raptor codes over gaussian channels,” *IEEE Communications Letters*, vol. 11, no. 4, pp. 343–345, Apr. 2007.

[35] ——, “Reduced-complexity decoding of raptor codes over fading channels,” in *Proc. Global communications Conference*, Nov. 2006, pp. 1–5.

[36] H. Jenkac, J. Hagenauer, and T. Mayer, "The turbo-fountain," *Euro. Tran. Telecomms.*, vol. 17, no. 3, pp. 337–349, Jun. 2006.

[37] A. Tarable and S. Benedetto, "Efficiency of precode-only raptor codes and turbo-fountain codes," in *IEEE Information Theory Workshop*, Oct. 2006, pp. 61–65.

[38] H. Jenkac and T. Mayer, "Soft decoding of LT-codes for wireless broadcast," in *Proc. IST Mobile Summit*, Dresden, Germany, Jun. 2005.

[39] J. W. Byers, M. Luby, and M. Mitzenmacher, "Asynchronous and reliable on-demand media broadcast," *IEEE Network*, vol. 20, no. 2, pp. 14–20, Mar. 2006.

[40] "Specification text for systematic raptor forward correction code," *3GPP TSG-SA WG4 S4-AHP238*, April 2005.

[41] M. Mitzenmacher, "Digital fountains: A survey and look forward," in *Proc. IEEE Info. Theory Wksp.*, San Antonio, TX, USA, Oct. 2004, pp. 271–276.

[42] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission," *IEEE Trans. Inform. Theory*, vol. 42, no. 6, pp. 1737–1744, Nov 1996.

[43] P. A. Chou, A. E. Mohr, A. Wang, and S. Mehrotra, "Error control for receiver-driven layered multicast of audio and video," *IEEE Trans. Multimedia*, vol. 3, no. 1, pp. 108–122, Mar. 2001.

[44] Y. Cao, S. D. Blostein, and W. Y. Chan, "Unequal error protection rateless coding design for multimedia multicasting," in *Proc. International Symposium on Information Theory*, June 2010, pp. 2348–2442.

[45] H. Jenkac, J. Hagenauer, and T. Mayer, “The turbo-fountain and its application to reliable wireless broadcast,” in *Proc. European Wireless 2005*, Nicosia (Cyprus), April 2005, pp. 1–7.

[46] A. Talari, B. Hahrasbi, and N. Rahnavard, “Efficient symbol sorting for high intermediate recovery rate of LT codes,” in *Proc. International Symposium on Information Theory*, Austin, TX, USA, June 1998, pp. 2443–2447.

[47] L. Benacem and S. D. Blostein, “Raptor-network coding enabled strategies for energy saving in DVB-H multimedia communications,” in *Proc. First Int. Conf. on Green Circuits and Systems*, June 2010, pp. 527–532.

[48] M. Luby, M. Mitzenmacher, and A. Shokrollahi, “Analysis of random processes via and-or tree evaluation,” in *Proc. 9th SIAM Symp. Discrete Algorithms (SODA)*, Jan. 1998, pp. 364–373.

[49] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable video coding extension of the H.264/AVC standard,” *IEEE Trans. Circuits Syst. Video Technol*, vol. 17, no. 9, pp. 1103–1120, Sep 2007.

[50] R. Puri and K. Ramchandran, “Multiple description source coding through forward error correction codes.” in *Proc. Asilomar Conference on Signals, Systems and Computers*, CA, Oct. 1999.

[51] Y. Cao and S. D. Blostein, “Cross-layer raptor coding for broadcasting over wireless channels with memory,” in *Proc. 11th Canadian Workshop on Information Theory*, 2009, pp. 130–135.

[52] T. Courtade and R. Wesel, “A cross-layer perspective on rateless coding for wireless channels,” in *Proc. IEEE International Conference on Communications*, Jun. 2009, pp. 1–6.

[53] M. S. Alouini and A. J. Goldsmith, “Adaptive modulation over nakagami fading channels,” *Kluwer J. Wireless Commun.*, vol. 13, no. 1-2, pp. 119–143, May 2000.

[54] Q.Liu, S.Zhou, and G.B.Giannakis, “Cross-layer combining of adaptive modulation and coding with truncated arq over wireless links,” *IEEE Transactions on Wireless Communications*, vol. 3, no. 5, pp. 1746–1755, Sept. 2004.

[55] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, 2002.

[56] B. Sivasubramanian and H. Leib, “Fixed-rate raptor code performance over correlated Rayleigh fading channels,” in *Pro. IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, April 2007, pp. 912–915.

[57] L. R. Wilhelmsson, “Evaluating the performance of raptor codes for DVB-H by using Gilbert-Elliott channel,” in *Proc. IEEE Vehicular Technology Conference (VTC)*, Oct. 2007, pp. 1932–1936.

[58] S.S.Karande and H. Radha, “Hybrid erasure-error protocols for wireless video,” *IEEE Tran. Multimedia*, vol. 9, no. 2, pp. 307–319, Feb. 2007.

[59] M. Mushkin and I. Bar-David, “Capacity and coding for the Gilbert-Elliott channels,” *IEEE Trans. Inf. Theory*, vol. 35, no. 6, pp. 1277–1290, Nov. 1989.

[60] L. Wilhelmsson and L. B. Milstein, “On the effect of imperfect interleaving for the Gilbert-Elliott channel,” *IEEE Trans. on Commun.*, vol. 47, no. 5, pp. 681–688, May 1999.

Appendix A

Decoding failure probability evaluation of rateless codes when the number of received symbols is random

In Chapter 3, a simplified channel model is considered which assumes that the number of received symbols of class j users does not change and equals δ_j times the number of transmitted symbols. In this appendix, we show numerically that the difference created by using this model and a model which takes into account the randomness of the number of received symbols is insignificant.

A.1 Decoding failure probability evaluation

As described in Chapter 3, when standardized raptor codes are used, for $k > 200$, the probability that the receiver fails to fully recover k source symbols after m symbols are successfully received can be well modeled by the empirically determined equation [20],

$$P_e^r(m,k) = \begin{cases} 1 & \text{if } m \leq k \\ ab^{m-k} & \text{if } m > k \end{cases} \tag{A.1}$$

where a and b are constants given by $a = 0.85$, $b = 0.567$.

To calculate the decoding failure probability in the QoS constraints of Eq. (3.3) when the channel experiences independent packet losses, let n, m, k, $1-p$ represent the total number of transmitted symbols, received symbols, information symbols and the erasure rate for each code block, respectively. Then m is a Binomial random variable with probability density function (PDF) $Prob(m = x) = \frac{n!}{(n-x)!x!}p^x(1-p)^{n-x}$. Therefore, the probability of successfully decoding the whole code frame is equal to

$$1 - P_e^f(n,k,p) = \sum_{x=k}^{n}(1-ab^{x-k})P(m=x) \tag{A.2}$$

$$= \sum_{x=k}^{n}(1-ab^{x-k})\frac{n!}{(n-x)!x!}p^x(1-p)^{n-x}. \tag{A.3}$$

Since n is large, the above equation can be computationally complex. Letting $h = \min(k+8, n)$, we approximate the above equation while significantly reducing computational complexity via

$$1 - P_e^f(n,k,p) \geq \sum_{x=k}^{h}(1-ab^{x-k})\frac{n!}{(n-x)!x!}p^x(1-p)^{n-x}$$

$$+ (1-ab^8)\sum_{x=h}^{n}(1-ab^{x-k})\frac{n!}{(n-x)!x!}p^x(1-p)^{n-x}$$

$$\approx \sum_{x=k}^{h} (1-ab^{x-k}) \frac{1}{\sqrt{2\pi np(1-p)}} \exp \frac{-(x-np)^2}{2np(1-p)}$$
$$+ (1-ab^8)(Q(\frac{h-np}{\sqrt{np(1-p)}}) - Q(\frac{n-np}{\sqrt{np(1-p)}})) \quad \text{(A.4)}$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-x^2/2)dx$ is the Q-function. The difference between the two sides of the inequality in the first step is minimal because when $x > k+8$, the outage probability $ab^8 < 0.01$ is very small and can be ignored. The approximations made in the second step use the normal approximation (known also as DeMoivre-Laplace Theorem) [55], which are also very accurate for large n. The probability that the decoder of a class j user fails to fully decode layer l, $P_e(l, j)$ in (3.3), can be evaluated as $P_e^f(n, k, p)$ of (A.4) with the number of transmitted symbols $n = t_l = (1+\varepsilon)K\rho_l$, the code dimension $k = S_l$ and $p = \delta_j$.

A.2 Numerical results

In Fig. 3.4, the minimum overhead required to achieve user' QoS requirements when different value of ρ_1 is used. The results in Fig. 3.4 are shown using the simplified channel model. In Fig. A.1, the same results are shown using the channel model described in this appendix. The proposed random interleaved UEP design employing standardized raptor codes is used. The dashed line represents the results of the original memoryless erasure channel model based on the decoding failure probability evaluated in Section A.1, while the solid line represents the results of simplified channel model that is analyzed in Section 3.4.2. The operating point at $\rho_1/\rho_2 = S_1/S_2$ on each curve is marked with a star, which indicates the performance of the equal error protection (EEP) scheme. It can be seen that

the minimum required transmission overhead is very sensitive to the choice of ρ_1: the minimum transmission overhead with the optimal choice of ρ_1 performs significantly better than that of the EEP scheme and other arbitrary non-optimized allocation schemes. Also, the performance difference between the two curves is very small. As shown in Fig. 3.2 and Fig. 3.3, the use of an inferior raptor encoder/decoder has a much larger impact on performance. This validates the use of the simplified channel model which allows for a much simpler solution based on convex analysis.

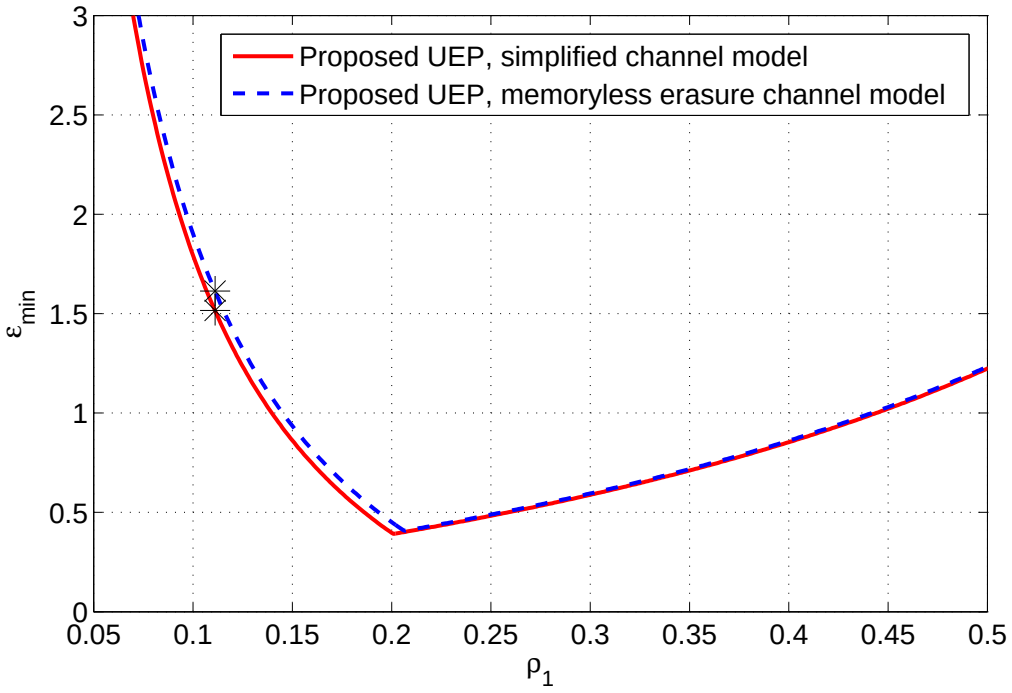

Figure A.1. Minimum transmission overhead required to meet users' QoS constraints for different layer allocation probabilities ρ_1. Two different channel models with proposed UEP rateless codes employing standardized raptor codes are used, where parameters $L = 2; K = 9000; S = [1000, 8000]; \delta = [0.4, 0.8]; P = [0.95, 0.8]$.

Appendix B

Proof of the last part of Lemma 3.1

In this appendix, we prove the last part of Lemma 3.1, i.e., after performing Algorithm 3.1, the transmission overhead cannot be further reduced with additional layer partitioning and selection probability re-assignment. Let Scheme A denote the source-to-channel layer mapping produced by Algorithm 3.1 and denote Scheme B as one which further partitions Layer l into Layers m and n with dimensions S_m and S_n, respectively. Denote the resulting optimal selection probabilities for Scheme B which minimize the transmission overhead as ρ_m and ρ_n for Layers m and n, respectively. We now show that the minimum required transmission overhead is no larger by using Scheme A with selection probability $\rho_l = \rho_m + \rho_n$ assigned to Layer l. For the same number of total transmitted symbols M, the effective average raptor code rates for Layer l in Scheme A, Layer m in Scheme B and Layer n in Scheme B are $R_l = \frac{S_l}{M\rho_l}$, $R_m = \frac{S_m}{M\rho_m}$ and $R_n = \frac{S_n}{M\rho_n}$, respectively. Without loss of generality, we assume $\rho_m/S_m \geq \rho_n/S_n$. Then it can be shown that $R_l = \frac{S_m+S_n}{M(\rho_m+\rho_n)} \leq \frac{\frac{S_n}{\rho_n}\rho_m+\frac{S_n}{\rho_n}\rho_n}{M(\rho_m+\rho_n)} = \frac{S_n}{M\rho_n} = R_n$. As the decoding failure probability of the raptor codes is monotonically increasing with code rate for the same user class, we have $(1-P_e(l,j)) \geq (1-P_e(n,j)) > (1-P_e(m,j))(1-P_e(n,j))$

for any class index j, where $P_e(.)$ is the same decoding failure probability function as defined in (3.3). This means that for the same number of transmitted symbols, the original mapping scheme (Scheme A) has higher probability of successfully decoding all the symbols in Layer l than Scheme B for all user classes. Therefore, for the same QoS constraints described by (3.3), Scheme A requires less minimum transmission overhead compared to Scheme B. Finally, raptor codes with larger dimension have better performance for the same code rate, which also favors no further layer partitioning.

Appendix C

General asynchronous multicast system setup

In Chapter 4, the rateless coded asynchronous multicast system is described where fixed channel erasure rates are used and all the users are assumed to access the multicast system at the same time. In this appendix, a more general and dynamic system model is investigated where both the channel erasure rate and user access time are random. The cost functions and outage probabilities are first derived using the more general system model. Then the transformation of Problem 4.10 to 4.11 under the new model is presented. It is shown that the methods and results presented in Chapter 4 for solving the asynchronous multicast optimization problems can still be applied to the more general system model used in this Appendix.

In the dynamic system model, it is assumed that the packet erasure rate σ_j for a class j user at any given time is a random variable with probability density function (PDF) denoted by $f_{\sigma_j}(.)$, however, the packet erasure rate for each user is fixed for each rateless code frame. This usually corresponds to a block fading channel model used in the physical layer.

In order to reach the target QoS of a Class j user, the user is required to decode Layer h_j. The number of packets need to decode layer h_j is given by $\lceil K_{h_j}(1+\omega)\rceil$. To compute the average number of transmitted packets required, a given value of σ_j is first considered. If an un-erased packet is considered as a successful trial and an erased packet is considered as a failed trial with failure probability equal to σ_j, then given the number of successful trials $\lceil K_{h_j}(1+\omega)\rceil$, the number of transmitted packets required, M_j, follows a negative binomial distribution with probability of $M_j = x$ equal to the probability of $\lceil K_{h_j}(1+\omega)\rceil - 1$ successes and x failures in $\lceil K_{h_j}(1+w)\rceil + x - 1$ trials, and success on the $\lceil K_{h_j}(1+\omega)\rceil + x$-th trial, i.e.,

$$Pr(M_j = x) = \begin{pmatrix} x + \lceil K_{h_j}(1+\omega)\rceil - 1 \\ \lceil K_{h_j}(1+\omega)\rceil - 1 \end{pmatrix} (1-\sigma_j)^{\lceil K_{h_j}(1+\omega)\rceil} \sigma_j^x. \tag{C.1}$$

It can be shown that the average value of M_j of the above negative binomial distribution conditional on a given σ_j is [55]

$$E(M_j)|_{\sigma_j} = \frac{\lceil K_{h_j}(1+\omega)\rceil}{1-\sigma_j}. \tag{C.2}$$

When σ_j is a random variable with a PDF $f_{\sigma_j}(.)$,

$$\begin{aligned} E(M_j) =& E_{\sigma_j}\left(E(M_j)|_{\sigma_j}\right) \\ =& \lceil K_{h_j}(1+\omega)\rceil E\left(\frac{1}{1-\sigma_j}\right) \\ =& \lceil K_{h_j}(1+\omega)\rceil \int_0^1 \frac{f_{\sigma_j}(x)}{1-x} dx. \end{aligned} \tag{C.3}$$

Finally, the cost function for all the user classes is given by

$$M_{av} = \sum_{j=1}^{J} w_j E(M_j)$$

$$=(1+\omega)\sum_{j=1}^{J} w_j K_{h_j} \int_0^1 \frac{f_{\sigma_j}(x)}{1-x} dx. \tag{C.4}$$

To compute the outage probabilities, we first introduce the access time random variable T_j for Class j users, which is defined by the number of packets that have been transmitted before the user accesses the multicast session. Users that join a multicast session too late may not be able to achieve their QoS requirements before the deadline. Users that join a session within a certain time interval, or access window, after transmission starts are allowed to access the multicast content. Therefore, the value of T_j for active users in Class j are truncated with a maximum value $T_{0,j}$. The probability mass function (PMF) of the access time T_j of Class j users is denoted by $Pr(T_j = x) = p_{T_j}(x), 0 \le x \le T_{0,j}$.

For a given erasure rate σ_j and access time T_j, the outage probability is equal to the probability that the receiver collects fewer than $\lceil K_{h_j}(1+\omega) \rceil$ encoded packets after M_0 packets have been transmitted, which can be expressed as

$$O_j|_{\sigma_j, T_j} = \sum_{i=0}^{\lceil K_{h_j}(1+\omega) \rceil - 1} \binom{M_0 - T_j}{i} (1-\sigma_j)^i \sigma_j^{M_0 - T_j - i}. \tag{C.5}$$

Thus the outage probability can be expressed as

$$O_j = \sum_{i=0}^{\lceil K_{h_j}(1+\omega) \rceil - 1} \sum_{y=0}^{T_{0,j}} p_{T_j}(y) \int_0^1 \binom{M_0 - y}{i} (1-x)^i x^{M_0 - y - i} f_{\sigma_j}(x) dx. \tag{C.6}$$

Comparing the cost functions given by Eqs. (C.4) and (4.8), it can be seen that the cost function of this dynamic model is equivalent to the cost function of the model described in Chapter 4 with fixed erasure rates for Class j users, $\sigma_{0,j}$, if

$$\frac{1}{1-\sigma_{0,j}} = \int_0^1 \frac{f_{\sigma_j}(x)}{1-x} dx. \tag{C.7}$$

In addition, observe that the outage probability O_j given by Eq. (C.6), despite being more complex than the outage expression, Eq. (4.5), in Chapter 4, is also a monotonically increasing function of K_{h_j}. Hence the outage constraint $O_j \leq \tau_j$ can also be simplified as l_j being greater than a constant for $j = 1, 2, ..., J$. Therefore, all the results provided by Chapter 4 for solving the asynchronous multicast optimization problem can also be applied to the more general and dynamic model described in this appendix.