VIDEO STREAMING OVER COGNITIVE RADIO NETWORKS

by

Azin Dastpak

B.Sc., Sharif University of Technology, 2009

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

Master of Science

in the

School of Computing Science
Faculty of Applied Sciences

© Azin Dastpak 2012

SIMON FRASER UNIVERSITY
Fall 2012

All rights reserved. However, in accordance with the Copyright Act of Canada, this work may be reproduced without authorization under the conditions for “Fair Dealing.” Therefore, limited reproduction of this work for the purposes of private study, research, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.
APPROVAL

Name: Azin Dastpak

Degree: Master of Science

Title of Thesis: Video Streaming over Cognitive Radio Networks

Examining Committee: Dr. Steven Pearce
Chair

__________
Dr. Jiangchuan (JC) LIU, Associate Professor
Senior Supervisor

__________
Dr. Mohamed Hefeeda, Associate Professor
Senior Supervisor

__________
Dr. Qianping Gu, Professor
SFU Examiner

Date Approved: September 25, 2012
Partial Copyright Licence

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the “Institutional Repository” link of the SFU Library website (www.lib.sfu.ca) at http://summit/sfu.ca and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author's written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, British Columbia, Canada

revised Fall 2011
Abstract

In recent years, there has been a tremendous growth in multimedia applications over the wireless Internet. The significant bandwidth requirement for multimedia services has increased the demand for radio spectrum. The scarcity of radio spectrum has challenged the conventional fixed spectrum assignment policy. As a result, cognitive radio emerged as a new paradigm to address the spectrum underutilization problem by enabling users to opportunistically access unused spectrum bands. In this thesis, we propose a framework for video transmission over cognitive radio networks. Our objective is to determine the optimal streaming policy in order to maximize the overall perceived video quality while keeping quality fluctuation at minimum. In our framework, we introduce a channel usage model based on a two-state Markov model and estimate the future busy and idle durations of the spectrum based on past observations. On the basis of this scheme, we formulate the streaming optimization problem under the constraint of the available bandwidth budget so that the optimal number of enhancement layer bits are assigned to each frame. We extend this algorithm for three different optimization levels: frame, GOP and scene. We evaluate our algorithm through extensive trace-driven simulation, and show that it improves the perceived video quality and increases bandwidth utilization.

**Keywords:** Cognitive Radio Networks, Scalable Video Coding, Wireless Multimedia
To my beloved mother and grandmother...
“Your time is limited, so don’t waste it living someone else’s life. Don’t be trapped by dogma which is living with the results of other people’s thinking. Don’t let the noise of other people’s opinions drown out your own inner voice. And most important, have the courage to follow your heart and intuition.

They somehow already know what you truly want to become. Everything else is secondary.”

— Steve Jobs
Acknowledgments

First and foremost, I would like to offer my deepest gratitude to my supervisors, Dr. Mohamed Hefeeda and Dr. Jiangchuan Liu, who without their support completion of the this thesis would not have been possible for me. I would like to thank Dr. Hefeeda who supported me through this thesis with his knowledge. It was his patience and encouragement that taught me how to do research. I would like to express my thanks to Dr. Liu for mentoring me in the past two years and giving me valuable insights into research.

I like to thank Dr. Gu, my thesis examiner, for being on my committee and reviewing this thesis. I also like to thank Dr. Pearce for taking the time to chair my thesis defense.

I wish to express my love and appreciation to my mother and grandmother who are always by my side. Needless to say, I would not have been here without their unconditional love and support.

Last, but not least, I would like to express my heartfelt thanks to Ali Khalili who stood by me through it all and gave me the utmost encouragement and support.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approval</td>
<td>ii</td>
</tr>
<tr>
<td>Abstract</td>
<td>iii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Quotation</td>
<td>v</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>vi</td>
</tr>
<tr>
<td>Contents</td>
<td>vii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>viii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>ix</td>
</tr>
</tbody>
</table>

## 1 Introduction

1.1 Introduction .......................... 1

1.2 Problem Statement and Thesis Contributions ......................... 2

1.2.1 Problem Statement ......................... 2

1.2.2 Thesis Contributions ........................ 3

1.3 Thesis Organization ........................... 4

## 2 Background and Related Work

2.1 Background ................................ 5

2.1.1 Cognitive Radio Networks  .................. 5

2.1.2 Architecture ............................... 6
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2</td>
<td>Scalable Video Coding</td>
<td>7</td>
</tr>
<tr>
<td>2.3</td>
<td>Channel and Traffic Model</td>
<td>8</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Primary User</td>
<td>8</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Secondary User</td>
<td>9</td>
</tr>
<tr>
<td>2.3.3</td>
<td>Traffic Models</td>
<td>10</td>
</tr>
<tr>
<td>2.4</td>
<td>Spectrum Occupancy and Traffic Prediction</td>
<td>11</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Hidden Markov Model</td>
<td>11</td>
</tr>
<tr>
<td>2.5</td>
<td>Related Work</td>
<td>12</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Cognitive Radio Networks</td>
<td>12</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Multimedia Services Over Cognitive Radio Networks</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>Problem Formulation and Solution</td>
<td>16</td>
</tr>
<tr>
<td>3.1</td>
<td>Problem Formulation</td>
<td>16</td>
</tr>
<tr>
<td>3.2</td>
<td>Proposed Solution</td>
<td>19</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Overview</td>
<td>19</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Precedence-Constrained Knapsack Problem</td>
<td>21</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Optimal Algorithm</td>
<td>22</td>
</tr>
<tr>
<td>3.2.4</td>
<td>Optimality and Performance Analysis</td>
<td>25</td>
</tr>
<tr>
<td>3.3</td>
<td>Illustrative Example</td>
<td>27</td>
</tr>
<tr>
<td>3.4</td>
<td>Adaptive Prediction Module</td>
<td>31</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Estimating spectrum usage parameters</td>
<td>31</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Prediction Intervals</td>
<td>32</td>
</tr>
<tr>
<td>3.4.3</td>
<td>Prediction Strategy</td>
<td>34</td>
</tr>
<tr>
<td>4</td>
<td>Performance Evaluation</td>
<td>35</td>
</tr>
<tr>
<td>4.1</td>
<td>Simulation Setup</td>
<td>35</td>
</tr>
<tr>
<td>4.2</td>
<td>Simulation Results</td>
<td>36</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Video Frame Quality</td>
<td>36</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Comparison between Aggregation Schemes</td>
<td>38</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Quality Variation</td>
<td>38</td>
</tr>
<tr>
<td>5</td>
<td>Conclusions and Future Work</td>
<td>42</td>
</tr>
<tr>
<td>5.1</td>
<td>Conclusions</td>
<td>42</td>
</tr>
<tr>
<td>5.2</td>
<td>Future Work</td>
<td>43</td>
</tr>
</tbody>
</table>
Bibliography
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Symbols used in the thesis.</td>
<td>19</td>
</tr>
<tr>
<td>4.1</td>
<td>Characteristics of MGS-coded video traces used in the experiments.</td>
<td>36</td>
</tr>
</tbody>
</table>
# List of Figures

3.1 The stream structure. ................................................................. 18
3.2 The structure of a SVC video stream. ......................................... 20
3.3 The Cognitive Radio Streaming (CRS) algorithm. ....................... 23
3.4 The CRS-Optimal-Policy procedure. ....................................... 24
3.5 Illustrative example of the proposed algorithm. Each item is a node in the tree which corresponds to a group layer. Nodes are labeled in DFS order. The number inside the rectangle denotes the weight of that node and the number under the rectangle represents the node’s profit. .................................................. 28
3.6 The result of CRS-Optimal-Quality algorithm is represented by the shaded nodes that constitute the optimal subtree. In our example, the optimal streaming policy involved streaming all three layers of the first and second group and none of the third group. The arrows show the order that the algorithm traversed the nodes. .... 30
4.1 Quality comparison. ............................................................... 37
4.2 Segment average quality. ....................................................... 39
4.3 Coefficient of variation of each segment. .................................. 40
4.4 Min. quality/ Avg. quality ratio. ............................................. 41
Chapter 1

Introduction

In this chapter, we briefly introduce cognitive radio networks along with the motivation for this work. Then, we describe the problem addressed in this thesis and present our contributions. The organization of this thesis is given at the end of the chapter.

1.1 Introduction

The growth of bandwidth-intensive wireless applications has increased the demand for wireless spectrum and raised the challenging issue of spectrum scarcity in recent years. It has been shown in [17] that the spectrum scarcity is largely due to the spectrum underutilization rather than physical shortage of it. According to the US Federal Communication Commission (FCC), the variations in the utilization of the primary spectrum ranges from 15% to 85% [3, 15]. In this regard, another research has been performed in major US cities and showed that many portions of the spectrum below 1 GHz remain unused for significant periods of time [17].

The underutilization of spectrum is mainly due to the conventional fixed spectrum assignment policy where the wireless spectrum is leased to certain operators (primary users) for a fixed, relatively long period of time (several years). Such static spectrum assignments lead to spectrum white spaces, hence inefficient spectrum utilization. This motivated the US FCC to set up a Spectrum Policy Task Force (SPTF) as a possible change in spectrum allocation policies. With regard to such policy, many different organizations such as the US Defense Advanced Research Projects Agency (DARPA), the IEEE 802.22 Working Group and MITRE Corporation worked on standards and technologies to access and share spectrum dynamically [17, 45].

Dynamic spectrum access allows unlicensed (secondary) users to access unused portions of the
spectrum owned by licensed (primary) users. This approach is facilitated by the use of cognitive radio networks. Cognitive Radio (CR) can be defined as a wireless communication system which is aware of the environment and its changes and can adapt its parameters accordingly. Two main characteristics of a cognitive radio network are: (i) Cognitive Capability which is the ability to sense the unused spectrum at any time and location, and (ii) Reconfigurability, which is the ability to send and receive at different frequency bands. In addition, the components of a cognitive radio network can be classified as primary and secondary network. The primary network is licensed to operate in a certain band and consists of primary users and the primary base station. The secondary network can access the licensed band of the primary users in an opportunistic or negotiated manner [3].

1.2 Problem Statement and Thesis Contributions

In this thesis, we investigate the problem of improving the quality of the streamed scalable video in a single-channel transmission scheme in cognitive radio networks. We consider two important quality metrics in this regard: individual frame in terms of Peak Signal-to-Noise Ratio (PSNR) and quality variation between successive frames. Our objective is to maximize the quality of frames while keeping the quality variations at minimum.

1.2.1 Problem Statement

Spectrum scarcity is one of the main obstacles for delivering high quality multimedia services. As mentioned before, the scarcity of spectrum is mainly due to the static spectrum assignment policy. Therefore, multimedia transmission, as a bandwidth hungry application with stringent delay constraints, is one of the candidates that can fully benefit from cognitive radio technology. Our objective is to maximize the overall perceived video quality while minimizing quality fluctuations.

Our scenario involves a base-station transmitting video to a group of users. Beside licensed access to a spectrum band, the base-station can also access the secondary channel opportunistically whenever it is idle. In order to accommodate heterogeneous secondary channel states, we adopt scalable video coding to encode each video into a base layer and several enhancement layers. The base layer contains the basic quality information and the enhancement layer contains more detailed data to improve the service quality. We transmit the base layer on the most stable channel, primary channel, to which base station has licensed access; while enhancement layers are delivered on the secondary channel. We model the availability of the secondary channel as a Markov chain.
and estimate the duration of future idle and busy periods on the secondary channel based on past observations.

As the channel is sensed idle, the server schedules a segment of the enhancement layer video frames to stream over the secondary channel. Our problem is to determine the optimal enhancement layer rate for each frame in the video stream under the estimated bandwidth budget constraint. We propose an optimization algorithm, called Cognitive Radio Streaming (CRS), to solve this problem. We extend our formulation to perform optimization at individual frame level, group of pictures (GoP) level, and scene level. Thus, the optimal enhancement layer assignment problem, which we need to solve, can be stated as follows.

Problem 1. Estimate the next busy and idle intervals on the secondary channel and schedule a segment of enhancement layer frames. For the scheduled segment and a given maximum bit budget, find the optimal layer assignment policy such that: (i) the average quality among all video frames is maximized, (ii) the quality variation among frames in the scheduled segment is kept minimum, and (iii) the transmission rate of scheduled frames does not exceed the bandwidth capacity.

1.2.2 Thesis Contributions

We study and solve the problem of improving the quality of streamed scalable video frames over the secondary channel in cognitive radio networks. Our contributions can be summarized as follows [9].

- We introduce a channel usage model based on the two-state Markov model and estimate the future busy and idle durations of spectrum based on past observations.

- We formulate the streaming rate-distortion optimization problem under the constraint of the available bandwidth budget as a variant of the precedence-constrained Knapsack problem. We extend this formulation for three different optimization levels: frame, GoP and scene.

- We propose a dynamic programming algorithm to find the optimal number of enhancement layers assigned to each frame. The algorithm is pseudo polynomial in time and space. The complexity of the algorithm depends on the size and grouping case of the allocation segment as well as the available rate.

- We evaluate our algorithm through extensive trace-driven simulation and show that it improves the received video quality and increases bandwidth utilization. We also present simulation results under the three mentioned frame grouping in terms of average video quality and quality variation.
1.3 Thesis Organization

The rest of this thesis is organized as follows. In Chapter 2 we provide a brief background on cognitive radio networks and multi-layer videos, and we review related works in the literature. In Chapter 3 we present an overview of our system and formulate an optimization problem to determine the optimal streaming policy. Then, we present an efficient algorithm to solve this problem. We evaluate the proposed algorithm in Chapter 4 using actual video traces, and we conclude the thesis in Chapter 5.
Chapter 2

Background and Related Work

In this chapter, we present an overview of cognitive radio networks and in particular dynamic spectrum sharing in secondary networks. We introduce several traffic models and their usage in evaluating secondary channel’s availability. We describe multilayer scalable video coding and its characteristics, which are employed in the proposed cognitive radio streaming framework. We also summarize the related works done in the field of cognitive radio networks.

2.1 Background

2.1.1 Cognitive Radio Networks

In recent years, the enormous increase in the wireless networks and the devices operating in these networks has led to spectrum scarcity. Further researches in spectrum utilization has shown that the problem main cause of spectrum scarcity is the underutilization of frequency bands in the fixed spectrum policy. In this policy the wireless spectrum (i.e. frequency carriers) are leased to some operators for a long period of time (several years).

The measurements done in major US cities indicates that for significant periods of time, many portions of the spectrum below 1 GHz remain unused [17]. Another similar study showed that at any time and any location in US, approximately 5% of the spectrum is in use in the band below 3 GHz [44].

It became evident that the static spectrum assignment leads to spectrum white spaces resulting in inefficient spectrum utilization. As a result, cognitive radio network emerged as a new paradigm to solve the problem of inefficient spectrum usage by enabling dynamic spectrum access. According to
CHAPTER 2. BACKGROUND AND RELATED WORK

FCC [15], cognitive radio is a wireless communication system aware of its environment and can adjust the transmission parameters dynamically based on the changes in the environment. In addition, as introduced in section 1.1, two main characteristics of a CR network are: Cognitive Capability and Reconfigurability. Such a system allows licensed users (primary users) and unlicensed users (secondary users) to communicate efficiently without interference [45].

2.1.2 Architecture

The components of a cognitive radio network can be classified as primary and secondary network. Primary network is a network infrastructure whose users have exclusive access to licensed spectrum bands. Primary network consists of: (i) primary users and (ii) primary base station. Secondary network is not licensed to operate in a certain frequency band but can access the licensed band of primary users in an opportunistic or negotiated manner. As illustrated in the figure, the secondary network can either be an infrastructure or an ad-hoc network. The components of a secondary network are: (i) secondary users, (ii) secondary base station and (iii) spectrum broker, a scheduling server that shares the spectrum resources between different cognitive radio networks.

It is important to notice that secondary users can communicate with their base-station and each other in an ad-hoc manner in both licensed and unlicensed bands but in order to communicate with primary base-station they have to use the licensed bands. The most important issue in licensed band operations among secondary users is avoiding interference with primary users. This is done through spectrum hand-off, which means that in the presence of any other primary user, secondary users shall vacate the spectrum band immediately and move to the next available band. In licensed band operations, all the secondary users have the same right to access the unlicensed spectrum which necessitates the use of intelligent spectrum sharing algorithms.

In conclusion, there are four main functionalities of a cognitive radio network which can be listed as follows [3]:

1. **Spectrum sensing**: The ability to sense the spectrum at any time and location.

2. **Spectrum management**: The ability to allocate the best available spectrum band base on the availability of the spectrum and other policies.

3. **Spectrum mobility**: The ability to vacate the spectrum in the presence of any primary user and move to the next best available spectrum band.
4. **Spectrum sharing**: The ability to provide a fair and optimal spectrum allocation method among multiple secondary users.

### 2.2 Scalable Video Coding

In recent years, video streaming has become the dominant traffic in wireless networks [27]. The bandwidth-hungry and distortion-sensitive nature of multimedia applications allows it to fully benefit from the potentials of cognitive radio networks. The challenging issue in streaming video over cognitive radio networks is to provide stable and optimal video quality under the time-varying channel capacity. During the past decades, several scalable video coding techniques have been developed to support the heterogeneous network bandwidths and diverse receivers. By using scalable video coding, the video stream is encoded once and can be extracted in several ways according to network conditions and user capabilities.

Scalable video coders encode video stream into a base layer and one or more enhancement layer. The video can be decoded if only the base layer is received. The base layer provides the basic video quality which can be further improved by receiving more enhancement layers. The scalable video coders encode the enhancement layers to be either coarse-grained scalable (CGS), medium grained scalable (MGS), or fine-grained scalable (FGS).

A scalable video coding extension, called H.264/SVC standard [47], is a standard that adds scalability to the widely used H.264/AVC video coding technique [58]. The H.264/SVC standard provides video coding with a number of scalability approaches such as: temporal (frame rate), spatial (frame pixel resolution) and quality (Signal-to-Noise Ratio (SNR)) [48]. In temporal scalability, the frames belonging to the same group of pictures (GoP) form a hierarchical structures where the frames belonging to the higher layers can be predicted from the ones in lower layers. In spatial scalability, each spatial layer employs motion compensated and intra-prediction [48]. For example, the spatial layer $l$ of a frame can be predicted from the $l^{th}$ spatial layer of other frames as well as its own lower spatial layers. H.264/SVC coarse grain scalability (CGS) provides up to eight quality layers which successively increase the quality of frames in SNR [48]. In CGS all quality layers have the same spatial resolution but the increase in quality is achieved as a result of smaller quantization parameters. The H.264/SVC standard also supports combined spatiotemporal-SNR scalability.

The disadvantage of H.264/SVC CGS is its relatively poor rate-distortion performance in a situation where the quality and rate difference between layers are small. To achieve rate-difference in the granularity of one byte between layers, Fine Grain Scalability (FGS) has been initially introduced.
as a part of MPEG-4 standard. The unique characteristic of FGS encoding is that the enhancement layer bit stream can be truncated anywhere and the remaining bits can still be decoded. However, further research showed that FGS had a high computational cost. As a result, H.264/SVC designed a middle approach between CGS and FGS—named medium grain scalability (MGS). The scalability approach splits the residual discrete cosine transform (DCT) coefficients of a CGS layer into multiple MGS layers. In MGS coding, a quality enhancement layer can be partitioned up to 16 layers.

In the simulation of our framework, we use H.264/SVC CGS and MGS coded trace files as well as MPEG-4 FGS SNR scalable coded traces. Using scalable video coding allows the server to adapt to the dynamic changes in the secondary network and stream the optimal number of enhancement layer frames to maximize the perceived quality of the video.

### 2.3 Channel and Traffic Model

Cognitive radio architecture consists of three main components: primary user, secondary user, and shared channel. The licensed spectrum is a shared resource between primary and secondary users. Primary users have higher priority in using licensed channel. Therefore, when a primary user is detected, the secondary users have to vacate the channel in a timely manner to avoid interference. In order to reduce secondary users’ connection loss and primary users’ interference, the secondary users need to avoid the channels with high occupancy probability within a given time period. Moreover, in our single-channel video streaming scenario, the server needs to have an estimation of primary channel’s OFF/ON period in order to assure timely delivery of enhancement layer frames. Therefore, the secondary users should be able to evaluate the availability of the channel through predicting the traffic pattern of primary user. In this section we aim to introduce the modeling of each component and in particular the secondary channel.

#### 2.3.1 Primary User

There are different types of primary users and the primary traffic model is different with each one of them. For example, if the primary users are TV channels, the primary traffic tends to be bursty. This is due to the fact that TV programs are predetermined and when a program is finished, the off-air period will last for a long time and gives the channel to set into another state. These traffic conditions can be modeled as a Markov model chain with two states of 0 (busy) and 1 (Idle and available) [63]. This means that each channel will change states independently of the others and
provide the secondary users with the opportunity to send their data. State transitions happen at the
beginning of each slot with the appropriate probability. The unavailability of a channel can be due
to the fading which should be added into the designed Markov model.

When the primary user’s traffic is more dynamic and time variant, e.g. cellular network, Markov
model may be of less use for modeling the channel. In most cases, there exists more than one
secondary user and the traffic from different users/networks may be more variant and less correlated.
Poisson processes can be of great use here. There has been various studies showing that a wireless
traffic can be modeled by a Poisson process with arrival rate $\lambda_i$. This solution can be adopted to
the Markov chain model by computing the success probability ($P_{\text{success}}$) by estimating the state of
the Markov chain. The maximum likelihood estimator for the Poisson arrival rate $\lambda_i$ is
$\hat{\lambda}_i = \frac{1}{\bar{x}}$, where $\bar{x} = \sum_{j=1}^{n} x_j$, $x = (x_1, x_2, ..., x_n)$ are the $n$ observations and $n$ is the observation window
size [31].

2.3.2 Secondary User

In the secondary usage scenario, the availability of the channel depends on the primary user’s traffic.
Radio spectrum has been divided into and allocated to multiple primary users and there is a fraction
of the radio spectrum that is free at each moment. Isolating these unused portions will create a
spectrum pool that can be later used for spectrum allocation. Spectrum pooling concept is the basis
for the secondary usage system architecture and is called cognitive radio for virtual unlicensed
spectrum (CORVUS) [4].

As mentioned before, the secondary users have to vacate the channel immediately after arrival
of primary user. In this situation, using spectrum pooling is beneficial to find the replacement for the
lost channel in a shorter period of time. Furthermore, since the allocated spectrum to the secondary
user comes from portions of different primary user’s spectrum, arrival of a primary user will not
cause the complete breakdown of secondary user’s link.

Although spectrum pooling has many benefits, it introduces some complexities to the system.
After allocation of each portion of the radio spectrum to a primary user, a subset of the sub-channels
of the available spectrum is chosen. This choice is based on the minimum interference with the
primary user. Usage of any sub-channel by its primary user will result in spectrum shift for the
secondary user to another available sub-channel.
2.3.3 Traffic Models

As mentioned before, due to the opportunistic behavior of cognitive radio networks, the performance of secondary network depends on the spectrum occupancy pattern of primary users. Therefore, providing an accurate and realistic model of such patterns is critical. The occupancy model of primary users depends on the type of traffic data it transmits.

Traffic data patterns can be categorized into deterministic patterns and stochastic patterns [23]. TV transmitters are one example of deterministic data pattern in which each primary user is assigned a fixed time slot to transmit data. After the transmission is done, the channel will be idle. Stochastic models refer to dynamic models where the state of the channel or traffic changes in time. Data packets traffic, where the arrival of packets are modeled as a Poisson process with exponentially distributed service time, is an example of stochastic traffic. Stochastic parameters vary slowly and can be estimated from historical data [38]. Primary types of traffic can be categorized as follows.

Voice Data

The nature of a speech process switches between talk and silent periods. Thus, it can be modeled as a two-state Markov model [36]. Hong et al. propose a traffic model for radio telephone systems which assumes that new call origination rate is distributed uniformly over the service area while the call holding time follows a negative exponential distribution [24]. In [38], Li et al. attempt to address the prediction of traffic pattern of primary users in voice communications where two crucial factors are the call arrival rate and call holding time. They assume that the call arrival rate process is periodic with a period of 24 hours. They set of hourly observations of call arrival rate for this period can be considered as discrete-time series. Therefore, they use the Seasonal Auto-Regressive-Moving-Average model to predict the call arrival rate of \( n + 1 \) interval based on the known \( n \) past observations of call arrival rates. They also assume that the call holding time follows Gamma distribution and is mean-ergodic. Therefore, they use the mean in different time intervals to estimate the call holding time during that interval. By incorporating the prediction results, they predict the probability that the frequency bands are available within a given time period.
CHAPTER 2. BACKGROUND AND RELATED WORK

Video Data

To address different characteristics of variable bit rate traffic, different models such as Markov renewal model [40], Markov Modulated Fluid Flow (MMFF) [41], Markov Modulated Poisson Process [51] and Auto-Regression stochastic model [41] are proposed to represent the basic characteristics of variable bit rate traffic.

In [60], Zhao et al. propose a traffic model for multi-layer scalable video based on Markovian Arrival process with marked transitions. In this model the pairs of base and enhancement layer video frame size are analyzed and grouped into clusters which correspond to one state of the Markov chain. Such model is suitable in queuing analysis to study the performance of video transmission. In [16], the authors try to represent the arrival of calls, packets and bursts by use of a multilayer Markov model for the traffic model of general packet data.

2.4 Spectrum Occupancy and Traffic Prediction

Traffic pattern prediction enables the secondary users to estimate the utilization of the shared spectrum in near future. The traffic prediction technique needs to take into account various characteristics of network traffic such as Short Range Dependencies (SRD), Long Range Dependencies (LRD), self-similarity, and etc. The prediction models can be categorized as stationary and non-stationary.

Examples of stationary models are Poisson, Markov, and autoregressive models. Nonstationary models include neural network which can be used to predict different types of traffic data. However they are complicated to implement and their applicability is limited. We will introduce the hidden Markov model in the following subsections.

2.4.1 Hidden Markov Model

According to the definition by [13], a hidden Markov process (HMP) is a discrete-time finite state homogeneous Markov chain observed through a discrete-time memoryless invariant channel. Mathematically, a HMP is defined as the pair of \( \{X_t, Y_t; t \in N\} \) on the probability space \((\Omega, F, P)\) where \(X_t\) and \(Y_t\) represent the hidden state and observation sequence respectively. The mathematical model that generates an HMP is called the Hidden Markov Model (HMM) that is a finite state machine in which the observation sequence is a probabilistic function of states [16].

A discrete HMM with \(N\) states and \(M\) symbols has an \(N \times N\) state transition matrix which includes the probability of switching from one state to the other. The output probability matrix of
HMM is an $N \times M$ matrix that has the probability different outputs while being in a certain state. An $N$ dimensional vector, named initial probability matrix, defines the probability of being in a certain state in the beginning of the process.

The HMM parameters can be estimated using Baum-Welch Algorithm (BWA) \textsuperscript{[54]}, derived from the Expectation-Maximization (EM) algorithm. In some cases such as primary user spectrum usage prediction where these parameters have to updated constantly, Forward-only BWA can be used.

There has been several studies about Hidden Markov model (HMM) and traffic prediction. The idea of predictive dynamic spectrum access has been introduced in \textsuperscript{[8]}. In \textsuperscript{[22]} Ghosh et al. validate the existence of a Markov chain for sub-band utilization by primary users using real-time measurements and formulate a spectrum sensing paradigm as a hidden Markov model that predicts the true states of a sub-band. In \textsuperscript{[39]}, a discrete-time Markov chain with deterministic and stochastic duty cycles is designed to accurately describe the spectrum occupancy in the time domain. In \textsuperscript{[2]}, the spectrum occupancy of bands are represented by binary sequence of zero and one. These binary vectors of bands are then used in training process of HMMs assigned to each band. The parameters of each HMM is then estimated by using Foward BWA algorithm. As the secondary user selects a channel, it computes its probability of being occupied and idle and compares them to predict the channel’s behavior.

2.5 Related Work

2.5.1 Cognitive Radio Networks

Cognitive radio (CR) research covers a broad range of topics. As seen in \textsuperscript{[3]} and \textsuperscript{[62]}, most of them have mainly focused on physical and MAC layers such as spectrum sensing and spectrum access. In \textsuperscript{[6]}, Chen et al. design an approach to maximize the throughput of secondary users while minimizing the collision with primary users. The design consists of three components: spectrum sensor, spectrum strategy, and access strategy. It is formulated as a constrained partially observable Markov decision process (POMDP) with the aim to decouple the access strategy from spectrum sensors. Zhao et al., in \textsuperscript{[61]}, formulate the opportunistic spectrum access problem as a Markovian decision process (CMDP) with the objective to maximize the channel utilization while limiting the interference. The benefits of cooperation in spectrum sensing is introduced in \textsuperscript{[19][20]} for two-user and multi-user networks. The results showed that the cooperation between cognitive nodes increases
the overall agility of the network. In [29], Zhu et al. present a game theoretic overview of dynamic spectrum access. They study the behavior of cognitive users to develop efficient dynamic distributed spectrum sharing schemes and ensure fairness and optimality.

The important QoS issue has been considered in [14, 28, 52, 56]. The focus of these papers are on metrics such as throughput and delay. Fattahi et al. [14] propose a game theoretic approach for the problem of resource allocation for multimedia transmission in spectrum-agile wireless networks, where each station plays a resource management game coordinated by a network moderator. In [56], a cognitive MAC protocol is proposed to establish a CR adhoc network. The proposed protocol supplements the legacy CSMA/CA MAC protocol in terms of throughput and delay to fulfill the goals of cognitive wireless networks.

2.5.2 Multimedia Services Over Cognitive Radio Networks

Multimedia services over cognitive radio networks have also attracted considerable efforts from the research community. In multimedia transmission, QoS at the application layer is the most important from users’ point of view. The previously mentioned works considered maximizing throughput of secondary users as the main criterion of their design approaches. However, recent studies show that maximizing throughput does not necessarily provide QoS at the application layer for multimedia applications [33]. Therefore, application layer parameters such as video distortion play the most important part in multimedia transmission over CR. Cross-layer multimedia transmission, where different optimization parameters are considered jointly across different OSI layers has been widely studied in the literature. We classify these works based on the transmission type into unicast and multicast categories and review them in the following subsections.

Unicast Transmission

Shiang et al. [50] focus on delay-sensitivity characteristic of multimedia applications and consider a scenario where multimedia sharing users share a single-hop wireless adhoc network. A virtual priority queue is proposed as an interface for heterogeneous multimedia users. In this scheme, the users exchange their information about the channel and time share the various frequency bands based on the proposed virtual priority queue. A dynamic learning algorithm is then developed based on the priority queue analysis which helps users to adapt their channel selection strategy accordingly.

In [5], the scenario involves a multimedia cognitive network with one primary user and N secondary users. The secondary users compete with each other to buy the primary spectrum once it
becomes idle. In this scenario, each secondary user transmits multimedia stream to a corresponding receiver. Chen et al. [5] formulate the spectrum allocation problem as auction game and propose three distributed auction-based allocation algorithms. These algorithms consider social welfare and being cheat-proof as performance metrics in their evaluations.

A distributed joint routing and spectrum sharing algorithm is proposed in [12] where the cross-layer control scheme aims to maximize the channel throughput under the constraints of delay sensitive video application using prioritized queuing model. In this model, each node maintains a separate queue for each waiting unicast session for which it is a source or intermediate node. The expected delay of competing flows is then estimated considering the dynamic spectrum condition and traffic characteristics.

**Multicast Transmission**

Video multicast over cognitive radio network has also been studied in a few research papers. For example in [26], Hu et al. took the scalable video coding, video rate control, spectrum sensing, dynamic spectrum access, modulation, scheduling, and primary user protection as design factors to solve the video multicast problem and formulate it as mixed integer nonlinear programming (MINLP) problem. In their approach to solve the problem, they determine the optimal partition and modulation-coding (MC) of FGS data and adjust the solution based on the channel sensing feedback. Then use a tile scheduling algorithm to assign video packets to available channels. In [59], Yu et al. design an integrated approach to optimize the multimedia intra refreshing rate with spectrum sensing and access. The primary network usage is modeled as a finite state Markov process and the optimization problem is a partially observable Markov decision process (POMDP) and dynamic programming framework obtains the optimal policy. They consider the total end-to-end distortion including both source and channel distortion. In [30], Jin et al. propose a multihop multicast protocol by employing techniques of power control, cooperative communication and network coding. In this protocol, secondary users perform cooperative transmissions in order to mitigate loss and network coding is adapted to provide error control and recovery. They evaluate the performance of framework in terms of interference to primary users and throughput utility of secondary users. In [43], Mao et al. address the problem of video delivery over multi-hop CR network and formulate it as MINLP optimization problem with video quality and proportional fairness as objectives. They solve the problem using sequential fixing scheme.

In addition to the previously mentioned works which focus on cross-layer designs, a few works
investigate channel adaptive multimedia transmission over a cognitive radio network and aim to optimize the application layer QoS, video distortion, for multimedia transmission. In [42], Mansour et al. consider multi-user bit-rate and latency control of scalable video content in a cognitive radio network. The video transmission system is formulated as a switching control dynamic Markovian game which results in an improvement in video quality. In [34, 57], single-layer approaches are designed to provide QoS for multimedia transmission over CR networks. In [57], an analytical model is proposed to estimate the capacity for secondary users and a cognitive MAC scheme is used to keep secondary users from interfering with primary users. In [34], Kushwaha et al. model the arrival of primary users as a Poisson process and design a metric to measure the quality of sub-channels to select the required ones from the spectrum pool to establish a secondary link. They optimize the spectral resources with respect to primary user occupancy and availability of sub-channels and use fountain codes to compensate for the interference loss affecting secondary user link.

What differentiates our work from the previous works is that it decouples the spectrum sensing and control from spectrum usages. Our intention is to investigate the application-level characteristics of the video to be transmitted independent of routing, dynamic spectrum allocation and power control. The work is helpful in a situation where the video is streamed from a separate server other than the base-station which controls the power and deals with spectrum sharing and management. The streaming server does not sense the spectrum itself but rather relies on the feedback information from the base-station to get notified about the channel state and uses the history of secondary channel usage to estimate bandwidth, schedule the frames with the adaptive optimal bit-rate. Hence, our focus is on the server side implications on the video quality.

In our optimization framework, the QoS target is on the application layer quality metric such as video distortion and quality fluctuation. Furthermore, we address quality variation among successive frames as one of the important parameters that can degrade the clients’ experience. Although there are design approaches for multimedia transmission over cognitive radio in the literature, most of them focus on latency and distortion as the main optimization parameter and quality variation, to the best of our knowledge, has not been addressed in previous works.
Chapter 3

Problem Formulation and Solution

In this chapter, we first estimate the available bandwidth budget on secondary channel and then formulate the streaming rate-distortion optimization problem under the available bandwidth constraint. We solve the optimization problem by proposing an algorithm to determine the optimal SVC enhancement layer rate for the video sequence. As mentioned before, the transmission of the base layer is made reliable by transmitting it on the primary channel and we will focus on streaming the enhancement layer frames.

3.1 Problem Formulation

The problem we investigate in this thesis is improving the quality of streamed video by taking advantage of dynamic spectrum access to a secondary channel. Our scenario involves a base-station streaming scalable coded video to a group of users. In this scenario, the base layer of video frames are transmitted over the primary channel to ensure timely delivery. The enhancement layer frames are scheduled according to the availability pattern of the secondary channel. The goal is to find the optimal streaming policy in regards of the number of enhancement layers for each frame to maximize the overall perceived quality by the users.

In our formulation, we model the occupancy pattern of secondary channel as a two-state Markov model. As explained in Chapter 2, the availability of secondary spectrum depends on the type of the data it transmits. Thus, we assume that the primary users who occupy secondary channel are TV stations who broadcast programs according to preset schedules. In this case the secondary channel occupancy pattern tends to be bursty and can be modeled by a two-state Markov model.

The two state Markov model provides an estimation of the next busy and idle periods, $T_{busy}$ and
Therefore, prior to the upcoming idle interval, the server is capable of scheduling a stream of frames to be transmitted when the channel becomes available. The number of frames in stream $s$ can be computed as $N_s = (T_{busy} + T_{idle})/T_f$, where $T_f$ is a constant that denotes the frame duration in seconds. The first frame in stream $s$ is the undelivered frame with the smallest index number which has not missed its deadline yet. A maximum bit-budget of $B_{max} = T_{idle} \cdot B_{sec}$ can be allocated across stream $s$, where $B_{sec}$ is the secondary channel bandwidth.

To determine the optimal streaming policy, the server can further group consecutive frames in stream $s$ into $G$ groups and perform rate-distortion optimization on these groups of frames. Depending on the level of optimization we want to achieve, the grouping policy can either be one of the three following cases:

- scene-based, where all images belonging to the same scene are grouped together
- GoP-based, where each group contains all frames belonging to the same group of pictures
- frame-based, where each group consists of just an individual frame.

Figure 3.1 shows a stream of frames consisting of 4 enhancement layers, which are segmented into two groups according to the GOP-based case.

Given a stream $s$, let $G$ be the number of groups in stream $s$ and $N_g$ be the number of frames in each group $g$, where $g = 1, ..., G$. With the maximum bit-budget $B_{max}$, let $\pi_g$ denote the number of bits allocated to each frame in group $g$, with $\pi = (\pi_1, ..., \pi_G)$ being the streaming policy for the stream. We define $Q(\pi)$ as the overall quality of the stream of frames (in PSNR) under the streaming policy $\pi$. Therefore, $Q_{n,g}(\pi_g)$ is defined as the quality of frame $n$, $n = 1, ..., N_g$, in group $g$ when the enhancement layers of frames in that group are encoded with $\pi_g$ bits.

Given the above definitions, the video streaming optimization problem can be formulated as follows. For the stream $s$ and a given maximum bit-budget $B_{max}$, find the optimal policy $\pi = (\pi_1, ..., \pi_G)$ that:

$$\text{maximize } Q(\pi) = \sum_{g=1}^{G} \frac{1}{N_g} \sum_{i=1}^{N_g} Q_{i,g}(\pi_g),$$

subject to: $B(\pi) = N_1 \cdot \pi_1 + \cdots + N_G \cdot \pi_G \leq B_{max}$,

where $\pi_g \leq \pi_{max}$ and $g = 1, \cdots, G$. \hfill (3.1)

The number of possible solutions to the optimization problem is exponential in $N_g$ and trying all of them to find the optimal one is extremely expensive. In the following sections, we propose an
efficient and optimal algorithm to solve our problem with the use of the dynamic programming approach.

For quick reference, we list all symbols used in this chapter in Table 3.1.
### 3.2 Proposed Solution

#### 3.2.1 Overview

The optimization problem is a resource allocation problem. To solve this problem, we employ a dynamic programming algorithm. At each step of the algorithm, we increment the enhancement layer of frames in a group by a fixed size unit $U$. The size of $U$ can be in bytes, layers or other appropriate units depending on the type of scalable coded video that the server is streaming, (i.e.
Coarse Grain Scalable (CGS), Medium Grain Scalable (MGS) or Fine Grain Scalable (FGS) coded [48]).

It is important to mention that in our evaluation with FGS video traces, we use the rate-distortion curves provided in [10] to measure the quality of the streamed video. As it is not accurate to model the enhancement layer rate-distortion curves by a simple function, a piecewise linear approximation, using the 200 Kbps spaced sample points, is employed in [10] which gives an accurate characterization of the rate-distortion curve. Therefore, in the implementation of our FGS streaming algorithm, we choose our step unit $U$ to be 833 bytes and increment the possible values of enhancement layer bits per frame in steps of 833 bytes ($200 \text{ Kbit/sec} \times 0.033 \text{ sec/8}$), where 0.033 is the time duration of each frame in our scenario. When the video traces are MGS or CGS coded, enhancement layer frames are incremented one layer at each step and $U$ is the size of the layer. In the rest of this chapter we will use the terms unit and layer interchangeably.

Figure 3.2 shows the structure of an SVC video stream, where each group is divided into a number of layers. Each layer contains the enhancement layer of all frames in the group. Since the
layers have discrete sizes, our problem resembles a packing problem in which each layer in a group corresponds to an item in a knapsack. The item’s weight is equal to the layer’s size. The item’s profit is equivalent to the increased video quality gained by including the corresponding layer. We try to find the optimal subset of items to achieve the maximum quality given the available resource limit (bit-budget). If all items were independent of each other, our problem could be easily reduced to 0-1 knapsack problem. However, this is not the case across the items belonging to the same group, because an additional layer is only useful if the precedent layers are also transmitted. Therefore, our problem is a variant of precedence-constrained knapsack problem (PCPK) \[32\]. As PCPK is a special case of knapsack problem it is NP-complete \[21\]. However the problem is solvable via a dynamic programming solution in $\theta(KR)$, with $K$ being the number of items and $R$ being the available rate \[7\].

3.2.2 Precedence-Constrained Knapsack Problem

In the precedence-constrained knapsack problem, each item can be placed in the knapsack if all its precedent items are already included in the knapsack. Such dependencies among items in the knapsack can be illustrated as a tree structure. In the undirected tree $T = (V,E)$ rooted at node 0, where $V$ is the set of nodes representing items in the knapsack. Node 0 is an abstract item and the other nodes each represent item $I_{l,g}$ which is equivalent to layer $l$ of group $g$ in the stream. Figure 3.5 shows the dependency tree of the items in Figure 3.2. For each node $i \in V$, $w_i$ and $q_i$ denote the weight and profit of item $i$, respectively. Let $p_i$ be the precedent of item $i$ and $P[i,j]$ be the unique path from item $i$ to item $j$. We assume that items are labeled in a depth-first search (DFS) order. Let $B$ be the capacity of the knapsack. Then, our problem is reduced to finding the subtree $T' = (V',E')$ rooted at item 0 as follows:

$$\max \sum_{i=0}^{n} q_i x_i$$

s.t. $x_{p_i} \geq x_i$, $i = 1, 2, ..., n$ \hspace{1cm} (3.2)

$$\sum_{i=0}^{n} w_i x_i \leq B$$

$$x_i = \begin{cases} 1 & \text{if } i \in V' \\ 0 & \text{otherwise} \end{cases}$$
3.2.3 Optimal Algorithm

We assume that nodes in the knapsack tree are labeled in DFS order. Let $L_k = 0, 1, 2, ..., k$ denote the set of visited nodes and $T_k = (L_k, E_k)$ be the induced subtree. For each node $v \in L_k$ we define an optimal profit $P_{L_k}(v, b)$ under capacity constraint $b, b = 0, 1, ..., B$. By employing a dynamic programming algorithm we can find the optimal profit $P_{L_0}(0, B)$ in $\theta(nB)$ time. Therefore, the problem can be broken down to the following subproblems and solved in a recursive manner:

1. Initialization:
   
   $P_{L_0}(0, b) = \begin{cases} 
   q_0 & \text{if } w_0 \leq b \leq B \\
   -\infty & \text{otherwise}
   \end{cases} $

2. Move forward as long as there is an unvisited node. For $k \neq 0$ and for every $b = 0, 1, ..., B$,
   
   $P_{L_k}(k, b) = \begin{cases} 
   P_{L_{k-1}}(p_k, b - w_k) + q_k & \text{if } \sum_{i \in P[0,k]} w_i \leq b \\
   -\infty & \text{otherwise}
   \end{cases} $

3. Move backward and revisit labeled nodes when reaching a leaf node or a node that all of its successors have been visited. Let $v \neq 0$ and $L = L_{v-1} \cup T(v)$. For each $b = 0, 1, 2, ...B$,
   
   $P_L(p_v, b) = \max \{P_{L_{v-1}}(p_v, b), P_L(v, b)\}$.

After finding the optimal profit value, we can find the optimal solution by defining the index of node $v$ in the following manner:

1. Initialization:
   
   $I_{L_0}(0, b) = \begin{cases} 
   1 & \text{if } w_0 \leq b \leq B \\
   0 & \text{otherwise}
   \end{cases} $

2. Let $v \neq 0$ and $L = L_{v-1} \cup T(v)$
   
   $I_L(v, b) = \begin{cases} 
   1 & \text{if } P_{L_{v-1}}(p_v, b) < P_L(v, b) \\
   & \text{and } \sum_{i \in P[0,v]} w_i \leq b \\
   0 & \text{otherwise}
   \end{cases} $

We assign values of 0 and 1 to keep track of adding or removing a node to the solution.

The high-level pseudo-code of the Cognitive Radio Streaming (CRS) algorithm is given in Figure 3.3. The algorithm works as follows. In the initialization part, the server obtains an estimation of future busy and idle periods and hence computes the available bit-budget and schedules a stream of frames to be transmitted. The server further segments the stream into groups of frames based
CRS-Optimal-Quality Algorithm

1. // Inputs: $B_{\text{max}}$ maximum bit-budget, $S = (g_1, g_2, \ldots, g_G)$ the set of groups in current stream
2. // Output: $P(0, B)$ optimal quality value
3. // Initialization:
4. For every layer $l$ of group $G$, add Item $I_{l,g}$ to precedence Items’ tree $T = (V, E)$ rooted at abstract item 0
5. For every node in Items’ tree, label the node in DFS order and compute the corresponding weight, $w_i$, and quality $q_i$
6. $w' \leftarrow$ minimum weight of items
7. $P(0, b) = -\infty$, $I(0, b) = 0$ for all $b$, $w' \leq b \leq w_0 - 1$
8. $P(0, b) = q_0$, $I(0, b) = 1$ for all $b$, $w_0 \leq b \leq B$
9. $w_{\text{path}} = w_0$
10. for $k \leftarrow 1$ to $n$ do
11. Forward($k$)
12. if $k$ is a leaf node then
13. $v = k$
14. Reverse($v$)
15. $v = p_v$
16. while $v$ has no successor $i$ such that $i > k$ and $v \neq 0$ do
17. Reverse($v$)
18. $v = p_v$
19. CRS-Optimal-Policy(I, P)

Procedure Forward ($u$)

1. // Computes $P_{L_u} u, b$
2. $w_{\text{path}} = w_{\text{path}} + w_u$
3. $P(u, b) = -\infty$ for all $w' \leq b \leq w_{\text{path}}$
4. $P(u, b) = P(p_u, b - w_u) + c_u$ for all $w_{\text{path}} \leq b \leq B$

Procedure Reverse ($u$)

1. // Computes $P_{L_u} p_u, b$ by comparing the current value of two sub-trees
2. // one containing node u and one not including it.
3. $w_{\text{path}} = w_{\text{path}} - w_u$
4. for $b = w' \text{ to } B$
5. if $P(p_u, b) \geq P(u, b)$
6. $I(u, b) = 0$
7. else
8. $P(p_u, b) = P(u, b)$
9. $I(u, b) = 1$

Figure 3.3: The Cognitive Radio Streaming (CRS) algorithm.
Procedure CRS-Optimal-Policy\((I, P)\)

1. // Input: The items’ tree, \(T = (V, E)\), with \(I(i, b)\) and \(P(i, b)\) computed for all \(i \in V\) and \(b \leq B\)
2. // Output: \(\pi^* = (\pi^*_1, ..., \pi^*_G)\) the optimal streaming policy
3. \(L(i) \leftarrow 0\) for every node \(i\) //\(L(i) = 1\) if item \(i\) is included in knapsack
4. Mark node 0
5. \(i \leftarrow n\)
6. \(b \leftarrow B\)
7. \(last \leftarrow -1\) //keeps the last node of the optimal subtree within bitbudget \(b\)
8. while \((i > 0 \text{ and } b > 0)\) do
9.   while \((i \text{ is unmarked})\) do
10.      if \((I(i, b) = 0)\) do
11.         \(i \leftarrow i - 1\)
12.      else
13.         if \((last = -1)\) then \(last \leftarrow i\)
14.             Mark node \(i\)
15.         \(i \leftarrow p_i\)
16.      end if
17.   end while
18.   \(L(last) \leftarrow 1\)
19.   \(b \leftarrow b - w_{last}\)
20.   \(i \leftarrow last - 1\)
21.   \(last \leftarrow -1\)
22. end while
23. for (every \(i \in V\)) do
24.     if \((L(i) = 1)\) then
25.         Find the corresponding layer and group, \(l\) and \(g\), of item \(i\)
26.         Include layer \(l\) in streaming policy of group \(g\)
27. end if

Figure 3.4: The CRS-Optimal-Policy procedure.
on the intended aggregation case. The maximum bit-budget and the groups of items will be fed as input to the algorithm to find the optimal streaming policy. The algorithm initializes items corresponding to the group layers in a tree structure based on dependencies between layers. To find the optimal value for video stream quality, the algorithm starts by initializing the value $P_{L_0}(0, b)$ for all weights $b = 0, 1, 2, ..., B$ and continues by visiting the tree nodes in DFS order. Whenever a node is visited, it will be added to the set of labeled nodes and its $P_{L,v}, b$ value is computed for all weights $b = 0, 1, 2, ..., B$ in Forward procedure. If node $v$ is a leaf node or a node that all of its successors have been visited, the Reverse procedure is called, in which we revisit the predecessor of node $v$, $p_v$, and evaluate $P_{L}(p_v, b)$ for all weights $b$. In Reverse procedure, the basic idea is that we can find $P_{L}(p_v, b)$ by comparing the current value of two sub-trees, one containing node $v$ and one not including it. If there exists an unvisited successor node of node $p_v$, we continue by calling the forward procedure on the first unlabeled successor of node $p_v$. The procedure continues until we reach the root node from its last successor. The value $P(0, B)$ at the end of the algorithm would be the optimal quality given the bit-budget constraint $B$.

After the CRS algorithm in Figure 3.3 finds the optimal quality value, we find the optimal streaming policy which specifies the number of layers assigned to each group. To find the optimal solution we perform DFS in reverse order in CRS-Optimal-Policy algorithm presented in Figure 3.4. The algorithm starts with the last node $n$ and maximum capacity $B$. It follows a reverse order DFS and keeps track of the last node of the optimal sub-tree with capacity $b$.

### 3.2.4 Optimality and Performance Analysis

In this section, we first prove the correctness of the CRS streaming algorithm. We also show that the algorithm is pseudo polynomial in time and space. It finds the optimal policy and quality in $\Theta(nB)$ time. The space complexity of the algorithm is also $\Theta(nB)$. As $n$ is the number of items and each item corresponds to a group layer, the computational effort of our solution depends on the grouping case of the stream, i.e., scene-based, frame-based, or GoP-based. It also depends on the number of frames and scenes within a stream.

The following theorem proves the optimality of CRS-algorithm in Figure 3.3 using induction.

**Theorem 1.** The CRS-algorithm is optimal, i.e., it returns feasible and valid solution if and only if it exists.

**Proof.** To prove that the optimal value $P(0, B)$ at the end of the algorithm is indeed the optimal value, we need to show that the solution to each subproblem is optimal. Therefore, we want to show
that the value of \( P_L(v, b) \) at each step of the algorithm, for all \( v \in P[0, n] \) and all \( b = 0, 1, 2, \ldots, B \) is optimal and equal to

\[
P^*_L(v, b) = \max \sum_{i=0}^{n} q_i x_i | x_{p_i} \geq x_j, 0 < j \leq n, \sum_{i=0}^{n} w_i x_i \leq b, \text{ and } x_0 = 1. \quad (3.3)
\]

For the basic step, \( n = 0 \), the claim is trivial. Using induction, we assume that the claim is true for all trees with \( m \) nodes, \( m \leq n \). Assume that we have a tree rooted at node 0 with \( n + 1 \) nodes. For the last node \( v = n \), we obtain the value \( P_L^n \) from Forward Procedure as follows:

\[
P_L^n = P_L^{n-1}(p_n, b - w_n) + q_n. \quad (3.4)
\]

As \( p_n \in P[0, n] \) from the induction hypothesis we have:

\[
P_L^{n-1}(p_n, b - w_n) = P^*_L(n-1, b). \quad (3.5)
\]

Therefore, rewriting Eq. (3.4) gives us

\[
P_L^n = P^*_L(n-1, b) + q_n = P^*_L(n, b). \quad (3.6)
\]

We want to prove the claim for all nodes \( v, v = 0, 1, \ldots, n \). For node \( v = n - 1 \), according to the Reverse Procedure, we have:

\[
P_L^{n-1}(n-1, b) = \max \{ P_L^{n-1}(n-1, b), P_L^n(n, b) \}. \quad (3.7)
\]

As \( n - 1 < n \) and \( n - 1 \in P[0, n] \) by induction hypothesis we have:

\[
P_L^{n-1}(n-1, b) = P^*_L(n-1, b). \quad (3.8)
\]

Therefore we have

\[
P_L^n(n-1, b) = \max \{ P^*_L(n-1, b), P^*_L(n, b) \} = P^*_L(n-1, b). \quad (3.9)
\]

By going up the tree along the path \( P[0, n] \), we can prove \( P_L^n(v, b) = P^*_L(v, b) \) in the same manner. Consequently, we prove that for every tree with \( n + 1 \) nodes the claim is true for all the nodes along the path \( P[0, n] \) and for every weights \( b = 0, 1, \ldots, B \).

In the following theorem, we prove that the time and space complexities of the CRS algorithm is \( \Theta(nB) \).
Theorem 2. The time and space complexities of CRS-Optimal-Quality algorithm are $\theta(nB)$, where $n$ is the number of groups layers and $B$ is the maximum bit-budget capacity.

Proof. The algorithm consists of two basic procedures, Forward and Reverse. If a node has any unvisited successors then the Forward procedure will be called; otherwise the Reverse procedure will be carried out. Therefore, for every node except for the root node we will perform two procedures, as we just perform the Forward procedure at the root node. As a result we have $n + 1$ forward calls and $n$ backward calls. Since each procedure requires $O(B)$ time, the overall complexity will be $\theta(nB)$. During the algorithm, two values $P(v, b)$ and $I(v, b)$ need to be stored for each node $v = 0, 1, ..., n$ and each rate $b = 0, 1, ..., B$. Therefore, the space complexity is also $\theta(nB)$.

The obtained time and space complexities result in reasonable performance for practical values of $n$ and $B$. It’s important to note that the value of $B$ in our algorithm is not incremented in the granularity of bits but in the granularity of fixed unit size $U$, which can be in layers depending on the type of SVC video. On the other hand, the value $n$ depends on the aggregation method (scene, GoP, or frame based). In the frame-based method, $n$ is equal to the number frames scheduled to be transmitted during the next idle period. Even in that case, the algorithm will be able to find the optimal streaming policy, given the sufficient computational power in a reasonably long busy interval.

In the following section, we illustrate how CRS algorithm works to find the optimal quality by showing a complete example.

### 3.3 Illustrative Example

In this section, we provide a simple example to illustrate how CRS-Optimal-Quality algorithm works. For the sake of simplicity, in our example the stream consists of 48 frames and three enhancement layers. The frames are segmented into 3 groups based on GOP-aggregation case. By mapping each group layer to an item in knapsack, the algorithm constructs a tree structure of items based on the dependencies among items. Figure 3.5 shows the items’ tree, where all nodes are labeled in DFS order. The numbers in the square represent the weights of the items in bytes and the numbers below the boxes stand for the items’ quality in PSNR. The item’s quality is defined as the induced increase in video quality that results from including its corresponding enhancement layer. It is worth mentioning that node 0 is an abstract node whose weight and quality are equal to 0. Having a maximum bit-budget of 20,000 bytes, the CRS-Optimal-Quality algorithm finds the optimal
streaming policy which results in optimal video quality in the following manner:

First we begin with node 0. As it’s an abstract node with $q_0 = 0$ and $w_0 = 0$, we have:

$$P_0(0, b) = \begin{cases} 
0 & \text{if } 0 \leq b \leq 20000 \\
-\infty & \text{otherwise}
\end{cases}$$

and

$$I_0(0, b) = \begin{cases} 
1 & \text{if } 0 \leq b \leq 20000 \\
0 & \text{otherwise}.
\end{cases}$$

We continue by calling forward procedure on node 1, 2 and 3 and evaluate $P_{0,1}(1, h)$, $P_{0,1,2}(2, h)$ and $P_{0,1,2,3}(3, h)$ as follows:

$$P_{0,1}(1, b) = P_{0}(0, b - w_1) + q_1$$

$$= \begin{cases} 
1.94 & \text{if } 5789 \leq b \leq 20000 \\
-\infty & \text{otherwise}.
\end{cases}$$
As node 3 is a leaf node, we perform the Reverse procedure and revisit node 2 and evaluate $P_{[0,1,2,3]}(2, b)$ and $I_{[0,1,2,3]}(3, b)$ as follows:

$$P_{[0,1,2,3]}(2, b) = \begin{cases} 3.81 & \text{if } 8236 \leq b \leq 20000 \\ -\infty & \text{otherwise} \end{cases}$$

$$I_{[0,1,2,3]}(3, b) = \begin{cases} 1 & \text{if } 9841 \leq b \leq 20000 \\ 0 & \text{otherwise} \end{cases}$$

Since node 2 has no unlabeled successor, we continue calling Reverse procedure to evaluate $P_{[0,1,2,3]}(1, b)$ and $I_{[0,1,2,3]}(2, b)$ as follows:

$$P_{[0,1,2,3]}(1, b) = \begin{cases} 4.54 & \text{if } 9841 \leq b \leq 20000 \\ 3.81 & \text{if } 8236 \leq b < 9841 \\ 1.94 & \text{if } 5789 \leq b < 8236 \\ -\infty & \text{otherwise} \end{cases}$$

$$I_{[0,1,2,3]}(2, b) = \begin{cases} 1 & \text{if } 8236 \leq b \leq 20000 \\ 0 & \text{otherwise} \end{cases}$$

As Node 1 has also no unlabeled successor, we revisit node 0 via Reverse procedure to obtain

$$P_{[0,1,2,3]}(0, b) = \max \{ P_{[0]}(0, b), P_{[0,1,2,3]}(1, b) \},$$
Figure 3.6: The result of CRS-Optimal-Quality algorithm is represented by the shaded nodes that constitute the optimal subtree. In our example, the optimal streaming policy involved streaming all three layers of the first and second group and none of the third group. The arrows show the order that the algorithm traversed the nodes.

\[
I_{[0,1,2,3]}(1, b) = \begin{cases} 
1 & \text{if } 5789 \leq b \leq 20000 \\
0 & \text{otherwise.}
\end{cases}
\]

The algorithm continues to go through nodes 4 to 9 in a similar way and terminates after revisiting root node via its last successor, i.e, node 7. At the end of the algorithm, we obtain the optimal value \( P_{[0,1,...,9]}(0, 20000) = 9.04 \) and the optimal subtree of items that should be included in knapsack, which are nodes 1, 2, 3, 4, 5 and 6. Figure 3.6 depicts the order that the nodes are visited in our algorithm and the shaded nodes are the items that are included in the optimal subtree. Based on the optimal subtree, we conclude that the optimal streaming policy includes streaming all three
layers of group 1 and 2 and not streaming any layer from group 3.

3.4 Adaptive Prediction Module

Our optimization framework consists of streaming and prediction module. In the previous section we described the streaming module which takes in as input the results of the prediction module. As the streaming and prediction components are strictly independent and modular, any prediction model (e.g. multistage Markov model) will work in conjunction with our streaming module. In our framework, we adopt a simple two-state Markov chain model which portrays the characteristics of our primary users well. In this section we’ll go over our adaptive prediction method and traffic model.

3.4.1 Estimating spectrum usage parameters

The secondary spectrum usage depends on the activity of the primary users. In our framework, we model the spectrum usage with a two-state Markov model which switches between busy and idle periods. In this model, the arrival of secondary spectrum to each state follows a Poisson distribution. Therefore, the busy/idle periods of secondary spectrum follow a negative exponential distribution.

As spectrum usage is dynamic and varies over time, we make sure that our model learns from the changes of primary users’ activity over time and adapts to it by updating the parameters and adopting a recovery strategy.

Before the streaming module starts streaming, the prediction module gains a priori knowledge of the occupancy pattern by observing the channel for a period of time. The module gathers enough sample data of channel usage history. Let’s assume that base-station records the durations of \( N \) busy and \( M \) idle periods on secondary channel in two lists \( \{ BT_1, BT_2, ..., BT_N \} \) and \( \{ IT_1, IT_2, ..., IT_M \} \). The base-station uses maximum likelihood estimator (MLE) to calculate the average busy and idle interval (\( \overline{IT} \) and \( \overline{BT} \)).

The busy and idle intervals can be modeled by a negative exponential distribution with the following density function:

\[
p(t) = \alpha e^{-\alpha t}
\]  

(3.10)

In this case, the joint pdf of all idle duration samples, the likelihood function, is given by
CHAPTER 3. PROBLEM FORMULATION AND SOLUTION


\[ L(TI_1, TI_2, ..., TI_M, \alpha) = \prod_{i=1}^{N} \alpha e^{-\alpha TI_i} \]

\[ = \alpha^N e^{-\alpha \sum_{i=1}^{N} TI_i}. \quad (3.11) \]

Then the log likelihood function is given by

\[ L(\alpha) = N \ln \alpha - \alpha \sum_{i=1}^{N} TI_i. \quad (3.12) \]

As we want to maximize the function with respect to \( \alpha \). The first derivative and second derivative are as follows:

\[ \frac{dL}{d\alpha} = \frac{N}{\alpha} - \sum_{i=1}^{N} TI_i, \quad (3.13) \]

\[ \frac{d^2L}{d\alpha^2} = -\frac{N}{\alpha^2} < 0. \quad (3.14) \]

As the second order derivative is negative, the unique maximum likelihood estimator can be computed as follows:

\[ \frac{dL(\alpha_{ML})}{d\alpha} = 0, \]

\[ \frac{N}{\alpha_{ML}} - \sum_{i=1}^{N} TI_i = 0, \quad (3.15) \]

\[ \alpha_{ML} = \frac{N}{\sum_{i=1}^{N} TI_i} = \frac{1}{\overline{TI}_N}. \quad (3.16) \]

where \( \overline{TI}_N \) is the sample mean of idle durations.

3.4.2 Prediction Intervals

In [35], the authors consider the problem of predicting, on the basis of observations \( x_1, ..., x_n \) from an exponential distribution, a specified order statistic or average of a future sample of observations \( y_1, ..., y_k \). We use the work to construct a prediction interval for future observations of busy and idle durations.

Suppose \( x = (x_1, ..., x_n) \) is a sample of \( n \) observations of channel idle duration that follows exponential distribution. We also assume that \( y = (y_1, ..., y_k) \) is a sample of future observations.
Therefore, lower and upper prediction limits in the form of a 100\(\alpha\)% prediction interval \([L(x), U(x)]\) can be defined for a given statistic \(T(y)\) of the future sample.

\[
Pr[L(x) < T(y) < U(x)] = \alpha
\]

The probability density function of \(x_i\) which follows an exponential distribution is

\[
f(x_i, \theta) = \theta^{-1} e^{-x_i/\theta}
\]

and as it is shown before, \(T = x_1 + \ldots + x_n\) is sufficient for \(\theta\). Then \(Q = 2T/Q = 2n\bar{x}/\theta\) is a \(\chi^2(2n)\) variate. \(P_i = 2y_i/\theta\), \((i = 1, \ldots, k)\) are independent \(\chi^2(2)\) variate and \(2k\bar{y}/\theta\) is a \(\chi^2(2k)\) variate. Therefore, the variates \(z_i = y_i/\bar{x}(i = 1, \ldots, k)\) are correlated \(F(2, 2n)\) variates and \(z = \bar{y}/\bar{x}\) is a \(F(2k, 2n)\) variate. By using \(z_1, \ldots, z_k\), prediction intervals can be set for future observations \(y_1, \ldots, y_k\) based on observed samples \(x_1, \ldots, x_n\).

The joint density function of correlated \(F\) variates is:

\[
f(z_1, \ldots, z_k) = \frac{(n + k - 1)!}{(n - 1)!n^k} [1 + \sum_{i=1}^{k} z_i/n]^{-n-k}
\]

The above joint probability statements about \(z_1, \ldots, z_k\) yields joint prediction statements for \(y_1, \ldots, y_k\) based on the observed \(\bar{x}\).

Lawless et al. denote

\[
Pr_{r,k}(t) = Pr[\text{at least } r \text{ of } z_1, \ldots, z_k \geq t]
\]

where

\[
Pr_{r,k}^{(n)}(t) = \sum_{s=r}^{k} I_{s,k}^{(n)}(t) \quad 0 \leq r \leq k.
\]

and

\[
I_{r,k}^{(n)}(t) = \binom{k}{r} \sum_{i=0}^{k-r} \binom{k-r}{i} (-1)^i [1 + (r+i)t/n]^{-n}
\]

Then, for given \(r, k, n, t\) and observed \(\bar{x}\), the \(t\) value satisfying \(Pr_{r,k}^{(n)}(t) = \alpha\), can be extracted from the tables of \(t\) factors and will give us a certain one-sided prediction interval for samples from the exponential distribution.
3.4.3 Prediction Strategy

Based on the channel usage history, we obtain MLE of $\theta$, which is the weighted average of observed durations. Then we obtain a 90% lower limit on the prediction interval using Eq. 3.21 and values from $t$ table. In our system, we assume that there is a common control channel between clients and base-station where clients can communicate with base-station by sending control packets in times of failure. We may adopt a conservative approach in which we estimate the next idle duration to be equal to the lower limit of the its prediction interval. In this case, we may experience the situation where the channel stays idle longer than expected.

If the future interval last shorter than expected, and clients experience packet loss, they notify base-station. The clients will send feedback of lost packets to streaming server based on a randomized scheme, in a way one percent of the clients will send feedback. As the server gets notified, it will update the estimator and estimate the next busy and idle duration. It’s important to mention that in the worst case, the quality of missed video frames will be not be worse than the base layer.
Chapter 4

Performance Evaluation

In this chapter, we first explain the setup of our simulation and define several performance metrics used in the evaluation. We then present the evaluation results of our proposed Cognitive Radio Streaming framework under three optimization levels and also compare it with the streaming over traditional wireless networks without taking advantage of opportunistic spectrum access.

4.1 Simulation Setup

We implemented an event-driven simulator in Java to evaluate the performance of our Cognitive Radio Streaming framework under three optimization cases as mentioned in Chapter 3. The simulated network consists of a server that streams multiple video sequences to clients over a secondary channel based on dynamic spectrum access policy. The secondary channel’s bandwidth is equally divided into sub-channels to accommodate the transmission of each video. The server uses the CRS-Optimal-Quality algorithm in Figure 3.3 in order to transfer the optimal number of enhancement layer frames of a video sequence.

In our simulator, the availability of the secondary channel is derived from a two-state Markov model which switches between idle and busy states. The arrival pattern of channel into each state is a Poisson process and the times spent in each state are negative-exponentially distributed with a mean inter-arrival time of $\mu$. In our simulation, we chose the mean duration of idle states to be 2 minutes.

In order to generate the video traffic, we use multiple scalable video traces available at the Video Trace Library at Arizona State University [49][55]. In particular, the results of our simulation are based on three video streams: Starwars, NBC News, and Sony Demo. The traces are encoded
into four PSNR scalable layers using medium grain scalability (MGS) feature of H264/SVC coding standard. Each video has a frame rate of 30 fps. The frame resolution is $352 \times 288$ and each group-of-picture (GoP) has 16 frames with the structure of G16B15.

We summarize the information of the encoding, data rates and quality of each layer of different video files in Table 4.1.

<table>
<thead>
<tr>
<th></th>
<th>Starwars</th>
<th>Sony</th>
<th>NBC News</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame Rate (fps)</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Average Frame Size (bytes)</td>
<td>1098</td>
<td>2090</td>
<td>2987</td>
</tr>
<tr>
<td>Average PSNR (dB)</td>
<td>42.3</td>
<td>37.6</td>
<td>37.8</td>
</tr>
<tr>
<td>Average Bit Rate (Kbps)</td>
<td>255</td>
<td>700</td>
<td>489</td>
</tr>
</tbody>
</table>

The client receives segments of enhancement layer video frames on secondary channel during the idle intervals. The segment size may be varied depending on the length of idle interval. Each frame in the segment is assigned the optimal number of layers and its quality is optimized by Figure 3.3. In order to evaluate the streaming mechanism of our CRS framework, we consider the following metrics: (i) quality of individual frames in PSNR, (ii) average overall quality of received segments in PSNR, (iii) coefficient of quality variation in each segment, and (iii) normalized quality (min. quality/ avg. quality) of each segment.

### 4.2 Simulation Results

In the following subsections, we present the results of our evaluation based on the mentioned metrics.

#### 4.2.1 Video Frame Quality

In our first experiment, we compare the performance of CRS framework versus the streaming case where no secondary channel is present. Figure 4.1 shows the average GoP quality for the Starwars, News and Sony video sequences respectively. The comparison is made over a period of 10 minutes of each video clip. The results show that we can achieve up to 6 db improvement in quality by using our algorithm to have opportunistic access to secondary channel.
Figure 4.1: Quality comparison.
4.2.2 Comparison between Aggregation Schemes

We compare the proposed two aggregation methods, Frame-based and GoP based, by plotting the average frame quality of each received segment in Figure 4.2.

As the automatic segmentation of scene boundaries, where frames are aggregated into scenes with similar visual and motion content, is highly dependent on the individual video content and is difficult to obtain and the scene segmentation for MGS video traces used in our simulation was not provided, we did not include the Scene-based approach in our comparisons.

As it’s observed the average quality of received segments are higher in the frame-based approach as it results in a finer optimization. As the duration that the secondary channel stays idle varies from one interval to another, the average quality of segments may also fluctuate. The quality fluctuation among segments is dependent on the secondary channel’s bandwidth as well as the characteristics of video frames.

4.2.3 Quality Variation

As mentioned earlier, the average image quality in PSNR alone can not be an indication of the actual overall quality. High variation in quality between successive frames lowers the users’ perceived quality. The quality fluctuation among segments may be inevitable due to the dynamic natures of secondary spectrum, however; the quality variation among successive frames within a segment should be kept minimum. Therefore, we take into account two other metrics: Coefficient of Variation \(Cov_Q(C)\), and Normalized Quality \(Q_{min}/\overline{Q}\). which are defined as follows:

\[
Cov_Q(C) = \frac{\sigma_Q(C)}{\overline{Q}(C)},
\]

(4.1)

where \(\overline{Q}(C)\) is the mean, and \(\sigma_Q(C)\) is the standard deviation of image quality which are computed as:

\[
\overline{Q}(C) = \frac{1}{N} \sum_{n=1}^{N} Q_n(C),
\]

(4.2)

\[
\sigma_Q(C) = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} [Q_n(C) - \overline{Q}(C)]^2}
\]

(4.3)

The challenge for our streaming algorithm is to minimize \(Cov_Q(C)\) while having \(Q_{min}/\overline{Q}\) ratio close to 1.
We plot the coefficient of variation among successive frames in the received segments of three video sequences in Figure 4.3. We note that the coefficient of quality variation in the received segments is lower in Frame-based approach than in GoP-based approach. The reason can be explained as the latter approach does not consider the type of frames (e.g. I, B, P frames) and assigns the same number of layers to all frames within a GoP. As the increase in quality gained from assigning the same number of layers varies for different frame types, the quality variation tends to be higher in GoP-based approach comparing to frame-by-frame optimization.
Figure 4.3: Coefficient of variation of each segment.

We plot the $Q_{\text{min}}/Q$ ratio in Figure 4.4. The diagram shows a slight difference in $Q_{\text{min}}/Q$ values for the two mentioned approaches. This value is slightly higher in GoP-based approach. As noted in Figure 4.2, the average segment quality is lower in GoP-based optimization, while the minimum frame quality stays more or less the same in two approaches. This leads to higher $Q_{\text{min}}/Q$ ratio in GoP-based approach, but as it can be seen the ratio in frame-based approach is above 0.8, which is close to 1.
Figure 4.4: Min. quality/ Avg. quality ratio.
Chapter 5

Conclusions and Future Work

5.1 Conclusions

Limited bandwidth is the major obstacle to delivering high quality multimedia services. In this thesis, we investigated the problem of transmitting multimedia content over cognitive radio networks, which is a newly emerged wireless communication paradigm that resolves the spectrum underutilization problem by being aware of the environment and having the ability to change its transmission parameters accordingly. We introduced a secondary channel usage pattern based on Markov model and estimated the duration of future busy and idle intervals. On the basis of this scheme, we formulated the streaming optimization problem under the constraint of the available bandwidth budget so that the optimal number of enhancement layer bits are assigned to each frame. We extended this algorithm for three different optimization levels: frame, GOP. We then evaluated our algorithm through extensive trace-driven simulation. The evaluation results confirmed that the extra bandwidth gain from dynamic spectrum access improved the received video quality and increased bandwidth utilization.

In our evaluation study, we also compared the performance of cognitive radio streaming framework under the two optimization cases mentioned above. This comparison was based on the perceived overall quality of the received video which is determined by the individual frame quality and the quality variation among frames. We have found that the frame-based approach leads to higher video frame quality as well as smoother video in terms of quality fluctuation. However, this approach comes with high computational costs. On the other hand, GoP-based approach reduces the computational complexity but it results in slightly higher quality fluctuation and lower average video quality.
5.2 Future Work

The work in this thesis can be extended in multiple directions. We can consider other secondary channel occupancy models to match the behavior of various primary users depending on the data they transmit. The framework can also be extended so that the opportunistic access to the secondary channel is based on various multiple access methods (e.g. OFDMA).

Another possible extension to our work is improvement of dynamic spectrum access scheme by having access to more than one secondary channel. In this situation, the base-station can switch to any available secondary channel in the presence of a primary user. Such dynamic secondary channel assignments can be further enhanced by providing a pool of available channels in which the channels are prioritized based on their availabilities in the future.
Bibliography


