ADAPTIVE VIDEO STREAMING OVER COGNITIVE RADIO NETWORKS

by

Ala Eldin Omer Mohamed

A Thesis presented to the Faculty of the American University of Sharjah College of Engineering In Partial Fulfillment of the Requirements for the Degree of Master of Science in Electrical Engineering

Sharjah, United Arab Emirates

January 2017
# Approval Signatures

We, the undersigned, approve the Master’s Thesis of Ala Eldin Omer Mohamed.


<table>
<thead>
<tr>
<th>Signature</th>
<th>Date of Signature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr. Mohamed Hassan</td>
<td>___________________________</td>
</tr>
<tr>
<td>Associate Professor, Department of Electrical Engineering</td>
<td>Thesis Advisor</td>
</tr>
<tr>
<td>Dr. Mohamed El-Tarhuni</td>
<td>___________________________</td>
</tr>
<tr>
<td>Professor, Department of Electrical Engineering</td>
<td>Thesis Advisor</td>
</tr>
<tr>
<td>Dr. Mahmoud H. Ismail</td>
<td>___________________________</td>
</tr>
<tr>
<td>Associate Professor, Department of Electrical Engineering</td>
<td>Thesis Committee Member</td>
</tr>
<tr>
<td>Dr. Mohamed Saad</td>
<td>___________________________</td>
</tr>
<tr>
<td>Associate Professor, Department of Electrical and Computer Engineering</td>
<td>University of Sharjah</td>
</tr>
<tr>
<td>Thesis Committee Member</td>
<td></td>
</tr>
<tr>
<td>Dr. Nasser Qaddoumi</td>
<td>___________________________</td>
</tr>
<tr>
<td>Head, Department of Electrical Engineering</td>
<td></td>
</tr>
<tr>
<td>Dr. Mohamed El-Tarhuni</td>
<td>___________________________</td>
</tr>
<tr>
<td>Associate Dean, College of Engineering</td>
<td></td>
</tr>
<tr>
<td>Dr. Richard Schoephoerster</td>
<td>___________________________</td>
</tr>
<tr>
<td>Dean, College of Engineering</td>
<td></td>
</tr>
<tr>
<td>Dr. Khaled Assaleh</td>
<td>___________________________</td>
</tr>
<tr>
<td>Interim Vice Provost for Research and Graduate Studies</td>
<td></td>
</tr>
</tbody>
</table>
Acknowledgement

I would like to sincerely thank my advisors Dr. Mohamed Hassan and Dr. Mohamed El-Tarhuni for providing knowledge, guidance, support, and motivation throughout my research stages. I’m deeply beholden for their great assistance, worthy discussion and suggestions. I learnt a lot from our weekly research meetings that enhanced my research capabilities, communication, presentation and reports writing skills. I am really proud that I have been their student.

I would like to thank the professors of the Electrical Engineering department who taught me the master level courses with mighty teaching methods and skills. I am really appreciate their dignified advices and motivation. Special thanks also to the Electrical Engineering department for granting me the graduate teaching assistance (GTA) for the entire period of the MSc degree.

I am thankful to the examiners committee of my thesis defence, Dr. Mahmoud H. Ismail and Dr. Mohamed Saad. I really appreciate their valuable comments and feedback. I am also deeply grateful to Dr. Husameldin Mukhtar from Khalifa University, for his support and help during the first stages of my research.

Finally, I would like to express my sincere gratitude to my family: my parents, brothers and sisters for their kindness, ultimate support and motivation throughout the years of study. They are really the spirit for any success in my life.
Dedication

To my parents, brothers and sisters...

To my teachers...

To my country...
Abstract

Several challenges face reliable video streaming over wireless networks due to the stringent requirements of high data rate, low error rate, and limited end-to-end delay. Cognitive radio (CR) networks offer a great advantage to unlicensed users (typically called secondary users) by allowing them to exploit the unused spectrum of licensed users (known as primary users) on an opportunistic basis. However, it is more challenging to deliver video services over dynamic CR channels that are available to secondary users not only intermittently but with all the challenges of wireless channels.

In this research, several frameworks are proposed to stream different scalable videos from a base station to multiple secondary users over a CR network. The objective of this study is to ensure that end users will enjoy continuous video playback with acceptable perceptual quality. To achieve such a goal, a channel allocation algorithm is introduced to adaptively assign the available radio channels to secondary users while taking into considerations the quality of their assigned channels as well as their buffer occupancies. In addition, different streaming algorithms are devised to ensure the delivery of scalable video frames, with base and enhancement layers, within the delay constraints with priority given to the base-layer frames to guarantee the continuity of video playback. Moreover, adaptive modulation is used based on the importance of transmitted video information and the channel state information (CSI) as fed-back by secondary users. Extensive simulations are performed using SimEvents simulator in MATLAB, the results of which show that the proposed schemes that integrate the devised channel allocation and streaming algorithms with adaptive modulation and scalable source coding techniques react to the variations in the channel conditions and the dynamics of the secondary users’ playback buffers in an acceptable fashion. This in turn resulted in efficient usage of the available CR resources as demonstrated in the achieved peak signal-to-noise ratio (PSNR) of the reconstructed video streams with no interruptions in the playback process. It has also been shown that scalable videos outperform their single-layer counterparts in terms of the achieved video quality. Finally, it is shown that joint adaptive modulation and channel coding results in improved bandwidth utilization, continuous playback and enhanced perceptual video quality at the secondary users end.

Keywords: Cognitive radio networks; video streaming; dynamic channel allocation; opportunistic access; Adaptive Modulation; Markov model.
Table of Contents

Abstract ................................................................. 6
List of Figures .......................................................... 9
List of Tables ........................................................... 12
List of Abbreviations ................................................. 13

Chapter 1. Introduction and Literature Review ....................... 14
1.1. Overview ......................................................... 14
1.2. Basics of Cognitive Radio Networks ............................ 15
  1.2.1. Cognitive radio networks architecture ...................... 17
  1.2.2. Spectrum sensing ............................................ 19
  1.2.3. Traffic in CR systems ...................................... 21
  1.2.4. Challenges in CR networks .................................. 22
1.3. Multimedia Streaming ........................................... 23
  1.3.1. Video compression .......................................... 25
  1.3.2. Multimedia QoS ............................................ 27
  1.3.3. Scalable video coding ...................................... 29
1.4. Related Work .................................................... 32
  1.4.1. Multimedia services over CR networks .................... 32
  1.4.2. Unicast transmission ....................................... 33
  1.4.3. Multicast transmission ..................................... 33
1.5. Thesis Objectives ............................................... 36
1.6. Research Contribution .......................................... 36
1.7. Thesis Organization .............................................. 37

Chapter 2. On-Demand Video Streaming over CRN .................... 38
2.1. Problem Formulation ............................................ 38
2.2. System Model ................................................. 39
2.3. Channel Model ................................................ 41
2.4. Adaptive Modulation ......................................... 43
2.5. Video Encoding ............................................... 46
2.6. Dynamic Resource Allocation Subsystem ....................... 46
  2.6.1. Channel allocation algorithms ............................ 47
  2.6.2. Video streaming algorithms .............................. 50
2.7. Proposed VoD Streaming Architectures ......................... 57
  2.7.1. CSI-based VoD architecture ............................. 57
  2.7.2. Joint buffer-and-CSI based VoD architecture .......... 58
List of Figures

Figure 1.1: Cognitive radio network architecture .................................................. 18
Figure 1.2: Classification of spectrum sensing techniques ..................................... 20
Figure 1.3: General architecture of multimedia streaming system over CR networks .............................................................. 24
Figure 1.4: Frames dependency in the G16B3 GoP structure .................................. 27
Figure 1.5: Performance of the perceived video quality at the end CR user ............ 29
Figure 1.6: PSNR quality provided by the base and enhancement layers of one GoP. ........................................................................ 31
Figure 1.7: Differences between the layered coding techniques ............................ 32
Figure 2.1: Proposed video streaming scenario ..................................................... 38
Figure 2.2: Architecture of the video streaming system over CR channels .......... 40
Figure 2.3: Two-state discrete-time Markov model for the n-th PU channel ............ 41
Figure 2.4: BER for Rayleigh flat-fading channel under different modulation schemes ............................................................................. 45
Figure 2.5: Joint based channel allocation algorithm flow chart ............................ 48
Figure 2.6: The arrangement of the base and enhancement frames in GoPs ............ 50
Figure 2.7: Streaming algorithm flow chart .......................................................... 54
Figure 2.8: Streaming mechanisms ....................................................................... 55
Figure 2.9: CSI-based VoD architecture at the BS ................................................ 58
Figure 2.10: BCSI-based VoD architecture at the BS ............................................ 60
Figure 2.11: Channels allocation for sending Star Wars I to all SUs in CSI-based architecture ................................................................. 62
Figure 2.12: PSNR of Star Wars I sequence as received by SU_1, SU_2, and SU_3, respectively, in CSI-based architecture .......................... 63
Figure 2.13: Channels allocation for sending different videos at target BER 10^{-5} ....................................................................................... 65
Figure 2.14: PSNR of the different video sequences as received by SU_1, SU_2, and SU_3, respectively, with target BER 10^{-5} ................... 66
Figure 2.15: PSNR of the different video sequences as received by SU_1, SU_2, and SU_3, respectively, with target BER 10^{-3} ................. 68
Figure 2.16: Channels allocation for sending different videos at target BER 10^{-3} ....................................................................................... 69
Figure 2.17: Number of idle channels detected every $T_{slot}$ duration .................. 70
Figure 2.18: Average PSNR of Star Wars I sequence as received by SU_1, SU_2, and SU_3, respectively, for different streaming techniques ....................................................................................... 72
Figure 2.19: Average PSNR of Star Wars I sequence as received by SU_1, SU_2, and SU_3, respectively, for different streaming techniques ............ 73
Figure 2.20: Average PSNR of different video sequences as received by SU_1, SU_2, and SU_3, respectively, for different streaming techniques ....................................................................................... 74
Figure 2.21: Average PSNR of different video sequences as received by SU_1, SU_2, and SU_3, respectively, for different streaming techniques ....................................................................................... 75
Figure 2.22: Average PSNR of Star Wars I sequence as received by SU_1, SU_2, and SU_3, respectively, for different allocation techniques using one-GoP based streaming........................................ 77

Figure 2.23: Average PSNR of different video sequences as received by SU_1, SU_2, and SU_3, respectively, for different allocation techniques using one-GoP based streaming........................................ 78

Figure 2.24: Average PSNR of Star Wars I video sequence as received by SU_1, SU_2, and SU_3, respectively, for different allocation techniques using one-GoP based streaming........................................ 80

Figure 2.25: Average PSNR of different video sequences as received by SU_1, SU_2, and SU_3, respectively, for different allocation techniques using one-GoP based streaming........................................ 80

Figure 2.26: Average PSNR of Star Wars I sequence as received by SU_1, SU_2, and SU_3, respectively, for different allocation techniques using half-GoP based streaming........................................ 81

Figure 2.27: Average PSNR of different video sequences as received by SU_1, SU_2, and SU_3, respectively, for different allocation techniques using half-GoP based streaming........................................ 82

Figure 2.28: Average PSNR of Star Wars I video sequence as received by SU_1, SU_2, and SU_3, respectively, for different allocation techniques using half-GoP based streaming........................................ 83

Figure 2.29: Average PSNR of different video sequences as received by SU_1, SU_2, and SU_3, respectively, for different allocation techniques using half-GoP based streaming........................................ 83

Figure 3.1: PSNR of different video sequences as received by SU_1, SU_2, and SU_3, respectively........................................................................... 91

Figure 3.2: Buffer occupancy status for SU_1, SU_2, and SU_3, respectively. .......... 92

Figure 3.3: Channels allocation for the SUs .......................................................... 93

Figure 3.4: Average PSNR of different video sequences as received by SU_1, SU_2, and SU_3, respectively.......................................................... 94

Figure 4.1: The joint adaptive VoD architecture at the BS................................... 96

Figure 4.2: Go-back-N ARQ approach................................................................. 98

Figure 4.3: Stop-and-Wait AQR approach......................................................... 98

Figure 4.4: Joint adaptive modulation and channel coding video streaming architecture........................................................................................................ 99

Figure 4.5: Dividing video frame into several packets with an overhead bits. ... 100

Figure 4.6: The probability of correctly receiving a frame within a critical time vs the maximum number of correctable bits for GBN and SW ARQ approaches. .......................................................... 105

Figure 4.7: Frame dependencies in the G16B15 GoP structure.......................... 107

Figure 4.8: Number of idle channels detected every T_{slot} duration.................. 108

Figure 4.9: PSNR of NBC News sequence as received by SU_1, SU_2, and SU_3, respectively, using fixed FEC with R_c=1/3................................. 109

Figure 4.10: Buffer occupancies for SU_1, SU_2, and SU_3, respectively, using fixed FEC with R_c=1/3................................................................. 109
Figure 4.11: PSNR of NBC News sequence as received by SU$_1$, SU$_2$, and SU$_3$, respectively, using fixed FEC with $R_c=1/2$.

Figure 4.12: Buffer occupancies for SU$_1$, SU$_2$, and SU$_3$, respectively, using fixed FEC with $R_c=1/2$.

Figure 4.13: PSNR of NBC News sequence as received by SU$_1$, SU$_2$, and SU$_3$, respectively, using fixed FEC with $R_c=2/3$.

Figure 4.14: Buffer occupancies for SU$_1$, SU$_2$, and SU$_3$, respectively, using fixed FEC with $R_c=2/3$.

Figure 4.15: PSNR of NBC News sequence as received by SU$_1$, SU$_2$, and SU$_3$, respectively, using fixed FEC with $R_c=3/4$.

Figure 4.16: Buffer occupancies for SU$_1$, SU$_2$, and SU$_3$, respectively, using fixed FEC with $R_c=3/4$.

Figure 4.17: PSNR of NBC News sequence as received by SU$_1$, SU$_2$, and SU$_3$, respectively, using adaptive FEC.

Figure 4.18: Buffer occupancies for SU$_1$, SU$_2$, and SU$_3$, respectively, using adaptive FEC.

Figure 4.19: Average overall PSNR of NBC News sequence as received by SU$_1$, SU$_2$, and SU$_3$, respectively, for adaptive and fixed FEC with different code rates.
List of Tables

Table 2.1: The effective channel bit rate for different modulation schemes............. 43
Table 2.2: The availability parameters for the primary channels in the network................................................................. 70
Table 2.3: Statistics of Star Wars I received by all SUs. ........................................... 73
Table 2.4: Statistics of Star Wars III, Star Wars I and Star Wars II received by SU₁, SU₂, and SU₃, respectively.............................................. 75
Table 2.5: Comparison Statistics of different allocation algorithms when Star Wars I was delivered to all SUs on one-GoP basis. ....................... 80
Table 2.6: Comparison Statistics of different allocation algorithms when different videos delivered to all SUs on one-GoP basis......................... 80
Table 2.7: Comparison Statistics of different allocation algorithms when Star Wars I was delivered to all SUs on half-GoP basis.............................. 83
Table 2.8: Comparison Statistics of different allocation algorithms when different videos delivered to all SUs on half-GoP basis......................... 83
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACK</td>
<td>Acknowledgment</td>
</tr>
<tr>
<td>ARQ</td>
<td>Automatic Repeat Request</td>
</tr>
<tr>
<td>AVC</td>
<td>Advanced Video Coding</td>
</tr>
<tr>
<td>AWGN</td>
<td>Additive White Gaussian Noise</td>
</tr>
<tr>
<td>BCSI-CA</td>
<td>Joint Buffer and Channel State Information based Channel Allocation</td>
</tr>
<tr>
<td>BL</td>
<td>Base Layer</td>
</tr>
<tr>
<td>BS</td>
<td>Base Station</td>
</tr>
<tr>
<td>CSI-CA</td>
<td>Channel State Information based Channel Allocation</td>
</tr>
<tr>
<td>CGS</td>
<td>Coarse Grained Scalability</td>
</tr>
<tr>
<td>CR</td>
<td>Cognitive Radio</td>
</tr>
<tr>
<td>CSI</td>
<td>Channel State Information</td>
</tr>
<tr>
<td>EL</td>
<td>Enhancement Layer</td>
</tr>
<tr>
<td>FEC</td>
<td>Forward Error Correction</td>
</tr>
<tr>
<td>FGS</td>
<td>Fine Grained Scalability</td>
</tr>
<tr>
<td>GoP</td>
<td>Group of Pictures</td>
</tr>
<tr>
<td>MGS</td>
<td>Medium Grained Scalability</td>
</tr>
<tr>
<td>PU</td>
<td>Primary user</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to Noise ratio</td>
</tr>
<tr>
<td>SU</td>
<td>Secondary user</td>
</tr>
<tr>
<td>SVC</td>
<td>Scalable Video Coding</td>
</tr>
<tr>
<td>TDMA</td>
<td>Time Division Multiple Access</td>
</tr>
<tr>
<td>VoD</td>
<td>Video on Demand</td>
</tr>
</tbody>
</table>
Chapter 1. Introduction and Literature Review

In this chapter, we provide a short introduction about the multimedia services over wireless channels and the encountered problems in this field. Then we discuss the fundamentals of cognitive radio networks alongside the motivation of using this technology in video streaming applications. We also present the techniques used in video encoding as well as the metrics used for performance evaluation. After that we discuss the related work in this field of research and the problem investigated in this study as well as the thesis contribution. Finally, the general organization of the thesis is presented.

1.1. Overview

Recent advances in wireless communication have led to a huge interest in providing many new services and applications over wireless devices. However, this interest is impeded by the shortage of radio spectrum, due to the legacy method of static assignment of the radio spectrum to licensed users. Such fixed assignment causes underutilization of the radio spectrum, which results in inefficient spectrum usage. Therefore, researchers have been motivated to find new wireless communication technologies to overcome the radio spectrum scarcity and meet the increasing demand for the wide range of wireless services [1].

Cognitive Radio (CR) networks have been introduced by researches to overcome the spectrum scarcity problem. It provides an efficient usage of the underutilized radio spectrum through dynamic spectrum access methodology. CR has different system models and cognitive functionality levels [2]-[3]. The first CR standard using TV bands is IEEE 802.22 Wireless Regional Area Networks (WRAN), in which the medium access is controlled by a base station (BS) [4]. Researches on CR technology have mainly concentrated on investigating efficient spectrum sensing and access techniques with significant achievements done and still continuing [2]-[3].

Multimedia streaming applications over wireless channels are expected to grow drastically in the upcoming years. Delivery of such multimedia applications to end users is facilitated by the recent advances in wireless networks, which have an edge on the wired networks in terms of coverage, installation cost, mobility and availability. However, video streaming over wireless channels is faced with great challenges. For instance, wireless channels are of dynamic nature that typically leads to transmission
errors. On the other hand, video streaming has unique transmission requirements of high bandwidth, strict end-to-end delays and jitter. The perceived video quality is highly affected by variations in the transmission delay. Generally, the arrival rate of video frames to the playback buffer should be equal to the playback rate to maintain continuous playback. In addition, the inherent dependencies between video frames typically cause error propagation, if one reference frame is lost for example, which results in decreasing the quality of the decoded video significantly.

Video streaming applications can take full advantage of the improved spectrum efficiency that CR provides; by exploiting the available opportunities for transmission to attain an acceptable quality level for the received videos with continuous playback. However, reliable video streaming over wireless and CR networks is faced with challenges of different types. First, independent of the underlying network, challenges that stem from the stringent quality-of-service (QoS) requirements on the end-to-end delays and jitter, high data rate and low error rate, as demanded by the nature of the multimedia streams. In addition, video applications have heterogeneous traffic nature that differs not only in terms of video frames sizes but in terms of their dependency. Hence, deadlines and priority of video frames should be carefully considered to limit error propagation due to the loss of a reference frame. Second, challenges introduced by the nature of wireless channels that are known for their time varying nature with fluctuating signal-to-noise ratio (SNR), multipath fading and interference. Third, radio resources in CR systems are available to unlicensed users in a random and intermittent way depending on the traffic activities of primary users. Therefore, there is a need for video streaming schemes that efficiently utilize the available resources in the CR network, while taking into consideration the characteristics of video streams to satisfy the requirements of the end users.

1.2. Basics of Cognitive Radio Networks

The current allocation policies assign the most useful radio frequency spectrum bands to a specific set of network operators to be used exclusively for a long-term over large geographic areas [5]. A given frequency band is allocated to only one licensed operator in specific locations, while the other operators cannot use that part of spectrum for their applications. Such allocation policies cause inefficient usage of the radio spectrum and result in spectrum underutilization. As a result, the radio frequency spectrum becomes a costly and heavily regulate resource. For instance, in Europe, the
3G spectrum auction yielded 35 billion US dollars in England and 46 billion US dollar in Germany [5]. However, only a limited part of the spectrum resources will be used regardless of the traffic conditions of the operators and the opportunities missed by other users/operators for short term usage when that spectrum is not fully utilized by the licensed operators.

Studies by the Spectrum Task Force (SPTF) of the Federal Communications Commission (FCC) [5] have shown that the licensed spectrum bands are not fully occupied in a given area and time; as stated by FCCs [5] that, the licensed frequency spectrum utilization over the time and geographical location varies from 15% to 85%. Also measurements done in major US cities indicated that for significant periods of time, many portions of the spectrum below 1 GHz remain unused. 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 [1]. Also, experiments conducted by a shared spectrum company state that, about 62% of the spectrum below 3 GHz is unused white space, even in the most congested area near downtown Washington, D.C., where the spectrum is intensively utilized [6].

As a result of spectrum scarcity and underutilization, cognitive radio (CR), proposed by Joseph Mitola in 1998 [7], emerged as a new paradigm to solve the inefficient spectrum usage and scarcity problems by enabling dynamic and opportunistic access for the underutilized portions of the spectrum. CR allows the unlicensed users to access unused portions of the spectrum owned by primary users. Cognitive radio is a novel technology that improves the spectrum utilization by opportunistically access the frequency band allocated to a licensed user when its transmission is detected to be inactive, or to share the spectrum with that licensed user without causing an interference. The licensed users’ inactive portions of the spectrum are referred to as “spectrum holes” or “white spaces”, which are defined as: “a band of frequencies assigned to a licensed user, but at a particular time and specific geographic location, the band is not being utilized by that user” [8]. Therefore, a cognitive radio can be defined as: “an intelligent wireless communication device, which senses its operational electromagnetic environment and can dynamically and autonomously adjust its radio operating parameters based on the interaction with the radio operational environment, and learn from the results of its actions to improve the communication performance” [8, 9].
According to the above definition, CRs have two main characteristics [10]: cognitive capability and re-configurability. The cognitive capabilities of a CR device, is the ability to sense the surrounding radio environment at any time and location, analyse the sensed information, and make the spectrum access decision according to the analysis results obtained. This capability cannot be simply realized by monitoring the power in some frequency bands, but more sophisticated techniques are required which allow to identify the unused portions of the spectrum at specific time and location. Consequently, the best spectrum bands and appropriate operating parameters can be selected. The other characteristic is the re-configurability, which is the ability of a CR to send and receive at different frequency bands by changing its operating parameters such as: (frequency, transmission power, modulation scheme, and communication protocol), based on the information obtained from spectrum analysis. Such CRs, when interconnected together (either through a cognitive base station (CBS) or directly in a peer-to-peer manner), form a cognitive radio network (CRN) in which the available spectrum is allocated among the users efficiently by using a smart spectrum management system, allowing both licensed and unlicensed users to communicate without interference.

1.2.1. Cognitive radio networks architecture. Figure 1.1 shows the components of a CR network which can be classified as primary and secondary network. The primary network is a network in which the users can access the licensed spectrum bands with exclusive prerogative. The common cellular and TV broadcast networks are examples of such primary networks. It consists of: (i) primary users (PUs) also called licensed users, who have a license to operate in a certain spectrum band, and (ii) the primary base stations, (e.g. the base station in a cellular system). Typically, the PUs and base stations do not have the CR capabilities to share the radio spectrum with other CR users [8].

In contrast, the secondary network is not licensed to operate in a certain frequency band, but it can access the licensed band of PUs in an opportunistic or negotiated manner. As illustrated in Figure 1.1, the secondary network can either be an infrastructure based or an ad-hoc network without infrastructure. The secondary network is composed of CR users, CR base station and spectrum broker. The CR user is also known as secondary user (SU), who has no license to operate in a certain frequency band. Such users should be provided with special functionalities to share the
licensed band with other CR users, however, in other networks that are infrastructure-based, the CR users have only the capability to sense particular parts of the shared spectrum band via local observations, and then report their sensing results to the CR base station without making decision regard the spectrum availability. The second element is the CR base station, which is a fixed infrastructure component supplied with the required CR capabilities. It plays different roles such as: providing connections to CR users who are located within its transmission domain while controlling their operations to access other networks. Moreover, the sensing operations, observations and analysis that are executed by the CR users are received and synchronized by the CR base station to make a final decision on the spectrum availability. The last component is the spectrum broker that is also known as scheduling server is a centric structure in the network that has the ability to share the available spectrum resources with other CR networks, while managing their allocation based on the sensing information provided by each network.

SUs have to avoid the interference with PUs in their licensed bands by vacating the spectrum band immediately and switch to other available bands when the PU is discovered and that what is known as “spectrum hand-off”. To access the unlicensed spectrum without causing harmful interference to PUs whose activity could be totally
random and independent to the CR networks, to achieve such a goal, the CRs should be equipped with some additional functionalities (compared to traditional wireless devices), in order to detect the spectrum availability and manage the spectrum efficiently [10], those four main functions are:

- **Spectrum sensing:** The ability to sense the spectrum at any time and location, to maximize the usage of available spectrum bands and to avoid the interference with PUs. CR users monitor the available unused spectrum bands in a timely manner by detecting the presence of PUs operating in their licensed band.

- **Spectrum management:** The ability to allocate the best available spectrum bands based on the collected information by the BS from the individual CR users.

- **Spectrum mobility:** The ability to vacate the currently used spectrum band in the presence of any PU, and move to the next best available spectrum band to continue the interrupted communication.

- **Spectrum sharing:** The ability to provide a fair and optimal spectrum allocation among multiple SUs who are trying to access the spectrum in order to prevent multiple users colliding in overlapping portions of the spectrum.

Through the previous mentioned CR functionalities, the CR networks allow users to actively monitor and exploit the temporally unused spectrum [8]. However, due to the dynamic nature of the channel availability and status, it is difficult to guarantee the QoS for CR users, and that is a significant challenge in CR networks.

### 1.2.2. Spectrum sensing

One of the most important functions required to establish an efficient CR network is the spectrum sensing, which is a task of having a knowledge about the spectrum usage through monitoring the PUs’ activity patterns in specific frequency bands [12],[13]. With CR technology, spectrum sensing has exceeded the conventional understanding of only measuring the radio frequency energy content over the spectrum, but rather to more complicated functions such as: determining the spectrum utilization attributes across time, space and frequency dimensions. Moreover, investigating the signal characteristics that is occupying the spectrum such as: carrier frequency, modulation techniques and bandwidth. Conversely, this broad concept of spectrum sensing requires more efficient algorithms and signal analysis approaches with an overhead computational complexity.
With spectrum sensing, SUs become able to adapt to the radio environment by exploiting the spectrum opportunities for transmission without causing an interference to PUs. This can be done through a real-time wideband spectrum sensing capability to detect weak primary signals in a wide spectrum range. After the spectrum holes have been identified, then SUs can utilize them for their transmission. In order to reduce the SUs’ connection loss and the PUs’ interference; the SUs need to avoid the channels with high occupancy probability within a given time period. Thus, SUs should be able to evaluate the availability of the channel(s) through predicting the traffic pattern of PUs.

To obtain an efficient spectrum sensing performance, different techniques have been introduced by the research community [7, 14]. Those techniques can be classified into three categories as shown in Figure 1.2: primary transmitter detection, primary receiver co-operative detection and interference-based detection. All spectrum sensing approaches aim to decide between the two hypotheses, defined as:

$$y(n) = \begin{cases} v(n), & H_0 \\ s(n)h(n) + v(n), & H_1 \end{cases}$$

where $y(n)$ is the signal received by the SU at the $n^{th}$ time sample, $s(n)$ is the transmitted signal of the PU, $v(n) \sim N(0, \sigma_v^2)$ is the additive white Gaussian noise (AWGN) with zero mean and variance $\sigma_v^2$, $h(n)$ is the amplitude gain of the channel. $H_0$ is the null hypothesis, which indicates that the channel is free to use with no PU existence over the spectrum, while, $H_1$ is the hypothesis that indicates the PU is online utilizing the channel.

![Figure 1.2: Classification of spectrum sensing techniques.](image-url)
Errors occur in spectrum sensing due to the false alarms and missed detections. The first one occurs when the sensed spectrum is available but $H_1$ is decided. Consequently, it leads to low spectrum utilization as the SUs miss the available transmission opportunities. The second error occurs when the sensed spectrum is allocated by PUs but $H_0$ is decided, which leads to undesirable harmful interference to PUs. Therefore, the performance evaluation of any spectrum sensing is conducted through estimating the probability of false alarm $P_F$ and the probability of detection $P_D$ which are related over the Receiver Operating Characteristic (ROC) curves.

The implementation of different transmitter detection techniques for sensing the frequency spectrum has been introduced in [14], with a detailed comparison between these techniques in terms of detection performance under different SNR conditions, sensing time required, ease of implementation and detection sensitivity. Those detection techniques are: Single-threshold Energy detection (ED), Matched filter detection (MFD), Cyclostationary feature detection (CFD). Moreover, an adaptive double threshold Energy detection based along with an improved Energy detection approaches are introduced to improve the accuracy performance of the conventional Energy detection.

1.2.3. Traffic in CR systems. The radio spectrum in CR networks is a shared resource between primary and secondary users with a higher priority to PUs in utilizing the channels. Hence, SUs have to vacate the channel immediately whenever the PU is detected on the spectrum to avert the interference possibility. Generally, the channels ON/OFF intervals should be estimated by the centralized BS by predicting the traffic behaviour of PUs, so that SUs can use only the highly probable available channels and avoid the channels with high probabilities of occupancy.

The PUs follow various traffic models. TV channels for example, have continuous traffic characteristics, because TV programs are predetermined in time and the off-air intervals will continue for a while once the program ends and the channel turns into another state. Studies showed that, the PU traffic behaviour occupies the channel for fixed amount of time (certain number of slots), after which the channel becomes idle. Such traffic patterns, can be modelled as a two-state Markov chain with two different states, 0 (idle or available channel) and 1 (busy or occupied channel) [15]. Transitions between states occur at the beginning of time slots with certain probabilities.
that form the probability transition matrix. Switching between states occurs independently in each channel, therefore SUs can start transmission on the available opportunities. Another type of PUs traffic is in cellular systems in which the traffic and the channels state are time variant. In such dynamic traffics, Markov model will not be efficient and the Poisson process can be used instead. The arrival of data packets can be modelled as a Poisson process with service time that vary exponentially [16].

For SUs, the availability of the channel relies on the PUs traffic patterns, therefore they should avoid the possible interference by vacating the channel in a timely manner once the PU arrives. Spectrum pooling (i.e. free portions of radio spectrum at each time) concept [17] can be used to overcome the possibility of connection loss and shortly find a replacement for the lost channel to continue their transmissions. The spectrum that is allocated to SUs is an aggregation of free portions from different PUs’ spectrum. Therefore, the allocated SUs will not lose the transmission link completely if a PU arrives at any time. The performance of the SUs depends on the traffic pattern of the PUs and the type of data patterns they send on the network. The traffic pattern can be either deterministic (e.g. TV channels) or stochastic (dynamic) (e.g. Cellular networks) pattern [8]. PUs traffic can be categorized into voice data and video data. Video data has variable bit rate traffic with different characteristics addressed in several models such as, Auto-Regressive stochastic model [18], Markov Modulated Fluid Flow (MMFF) [18] and Markov renewal model [19].

1.2.4. Challenges in CR networks. CR networks are faced with several challenges as follows:

1.2.4.1. Spectrum management. The CR functionalities are facing a great challenge due to the fact that CRs coexist with the primary networks. Thus, the spectrum sensing module should be able to know: where, how long to sense and the interference limits allowed. After that a decision is made by the CR through analyzing the information received from each sensing stage. Then, the management system should be aware of how to access and allocate the CR users on the available spectrum bands and releasing those bands once the PUs appear for transmission. Here, the spectrum mobility mechanism should know how to stop the CR transmission, where to move to other available bands and change the configuration parameters accordingly. All of these
issues done to avoid the PUs interference while improving the throughput of the CR [1].

1.2.4.2. Dynamic resource allocation. The main goal here to avoid harmful interference to PUs while exploiting the available resources (i.e., power and spectrum) efficiently. The power control is an essential factor in resource allocation for CR networks that operate in the underlay mode, where primary and secondary users can simultaneously access the spectrum band, while it is not essential in the overlay mode, where primary and CR users do not access the spectrum band at the same time [1]. Therefore, rather than traditional schemes used in legacy wireless networks, there is a need for updated MAC protocols that consider PUs activities in different spectrum bands to provide fair and optimal resource allocation among CR users and coordinate their access to the shared radio frequency resources, while reducing the interference to PUs. The MAC protocol is either performed by CR base station in centralized CR networks or independently by CR users in distributed CR networks. The need for such resource allocation methods is based on:

- The CRN architecture (i.e., centralized or distributed CRN)
- The type of operation mode of the CRN (e.g., overlay mode or underlay mode).

1.2.4.3. Security aspects. All known wireless network attacks and threads are challenging CR networks [20] such as jamming data and channel control signals in the radio frequency, fake MAC frame transmission and eavesdropping [21]-[22]. Along with these significant attacks, CR networks are also prone to new security threats according to their unique characteristics, such as emulation attacks on PUs (PUEA), spectrum sensing data distortion, beacon distortion and cross-layer attacks [23], [24], [25] and [26].

1.3. Multimedia Streaming

Multimedia transmissions over wired and wireless networks have increased rapidly in the last decade. Different wireless services are targeted by the users nowadays such as multimedia messaging, interactive voice communications (VoIP), live TV broadcasting, multimedia streaming on-demand (e.g. YouTube) and video conferencing. Wireless networks provide the multimedia services at a low-cost and flexible infrastructure but with limited QoS guarantees and resources that vary dynamically according to the network conditions. However, transmission of multimedia applications
has special characteristics different from those required by conventional data communication [27]. Such as large bandwidth, time sensitivity and stringent QoS requirements. For example, video frames should be available at the playback decoder on the receiver side on or before their specified playback time for undistorted video playback. Each frame should be received before its deadline, otherwise it is useless and will be discarded. Video streaming systems aim to deliver the video sessions frames within certain end-to-end delay to insure the continuity of the video playback and thereby, satisfy the clients’ requirements without negative effects. General architecture of the multimedia streaming system over CR networks is shown in Figure 1.3. It consists in general of a video encoder and decoder. A playback buffer is employed to compensate the difference in delays of the arriving packets due to the network conditions, where $f_r$ is the rate of successfully decoded frames and $f_p$ is the playback rate.

![General architecture of multimedia streaming system over CR networks.](image)

The media streaming can be categorized into two groups:

- **Live media streaming**, (e.g. TV broadcasting).
- **On-demand streaming**, also known as video-on-demand (VoD), (e.g. YouTube).

The on-demand streaming requires interactions between the communication end-points [28], rather than delivering the video packets one way from the source to destination in the live media streaming, where all users concerned with the service receive the same
video contents with similar playback process. Another difference is that, video-on-demand users are allowed to select different video contents from massive media database and interactively control the playback process. In addition to the diverse characteristics of media services, the QoS requirements for the two media services are also different. For live streaming services, the start-up delay should be minimized by keeping the playback buffer small, so that the information is delivered instantly from the source. In on-demand media streaming on the other hand, the major requirement is the smooth playback with no interruptions and the users can tolerate the long start-up delay unlike the live streaming.

Multimedia transmissions over CR networks encounter several challenges due to the unique nature of multimedia applications along with the attributes of CR networks. The CR resources have constraints such as the available spectrum bandwidth, transmission power, transmission data rate and time slots allocation. In CR systems, the sensing activity and data transmission occur separately at different times. Therefore, the time slots allocation and packets scheduling should be appropriately carried out to ensure an effective utilization of the available spectrum and deliver the video packets with the optimum possible quality. Another challenge is the dynamic conditions of the network that vary over time due to shadowing, interference, and fading. According to the channel condition, the system should compromise between data rates used and the corresponding bit error rates. Videos introduce their own characteristics such as heterogeneous traffic due to the dependency between different frames. For instance, if some important frames are lost then the remaining dependent frames become useless and will be discarded. Therefore, the frames should be carefully considered according to their sizes, deadlines, priority, and dependency.

Users can tolerate the varying network conditions by considerable buffering. Factors such as playback rate, network throughput and users’ buffer occupancy, determine the smoothness of video playback. VoD users are free to select the playback position. Therefore, when the user moves to a new playback position then new frames should be placed into that buffer immediately as the previous contents are no longer needed.

1.3.1. Video compression. There is an enormous amount of data contained in any digital video stream that consists of a sequence of images (frames) ordered in a
timely manner. Video frames are displayed at a specific rate called the nominal playback rate; so that the human eye can perceive the motions in the video stream in continuous perspective. Therefore, video streaming applications require high reliable channels with high data rate, and thus demand too much bandwidth compared to other data communication applications. On the other hand, wireless channels have time varying nature and restricted bandwidth. Therefore, compression and source coding techniques are needed to overcome the limitation of data rates [29]. Commonly, those techniques are used to reduce the size of the transmitted video information, for example some techniques remove the spatial and temporal redundancies in the video data. However, by removing that redundancy then the compressed video becomes more sensitive to channel and propagation errors. There are inherent interdependencies between the video frames, whereby losing a significant frame will result in losing a set of consecutive frames that mainly depend on the first one. Those important frames are encoded independently and required for encoding the consecutive dependent frames.

There are three main steps to achieve a complete video compression, which are: motion compensation, transform coding, and quantization and binary encoding. Most video codec (e.g. MPEG-2, MPEG-4) implement motion compensation technique for video compression. In this technique the video temporal redundancy is removed to reduce the size of the transmitted video data, because there is similarity between the adjacent frames in any video sequence. Therefore, considerable compression can be attained by encoding the differences between the frames only. Each video sequence is encoded into three types of frames, I, P, and B. The most important frames are the ones with I type, which are ‘intra-coded’ and independent of other frames but they are important for encoding/decoding other frames. I-frames with low compression levels are inserted at constant time intervals to limit the propagation errors. The other two types (P and B) are ‘inter-coded’ and dependent on the prior and/or future encoded frames respectively. The P-frame is coded ‘predictively’ depending on the prior I or P-frame, while the B-frame is ‘predicted bi-directionally’ depending on both the prior and future I or P-frame with the highest level of compression. In some standards, even the B-frames are used as source of prediction (e.g. H.264/SVC) [30]. The group of picture (GoP) is defined as the number of frames between two successive I-frames. Figure 1.4 shows the dependencies between the video frames in the G16B3 structure, in which the
GoP consists of 16 total frames starting with I and three B-frames between each successive I or P-frame (IBBBPBBBPBBBPBBB).

![Figure 1.4: Frames dependency in the G16B3 GoP structure.](image)

Discrete cosine transform (DCT) and discrete wavelet transform (DWT) are examples of transform coding in which the video frames are transformed from the spatial (pixel) domain to frequency domain. For example, in the widely used DCT technique, the video frames are transformed to the frequency domain with separable coefficients that vary in importance. The compression is done by considering the most significant DCT coefficients that fall in low frequencies. As a result, the spatial redundancy in the video frames has been reduced by averaging the areas with the same color using the DCT technique [31]. The video compression process is completed with the quantization and binary coding, in which a set of discrete integer values is used to approximate the transform coefficients achieved by the last stage ‘transform coding’. These integers are represented in bits and some coding techniques are employed to reduce the redundancy that present in the bit stream (e.g. differential coding, Huffman coding … etc.) [31].

1.3.2. Multimedia QoS. The important QoS issue in multimedia is considered from the following perspectives:

1.3.2.1. Network perspective QoS. A set of metrics should be maintained at a specific level to satisfy the QoS requirements for the end users. These metrics include the data rate (throughput), the total time at the network for (processing, queuing and transmission), which is designated as the end-to-end delay, the variation in delay known as the jitter, the transmission reliability indicated by the bit error rate (BER) and the packet loss rate. Those metrics determine the given network resources for the video service and thus provide a great help in allocating the network resources efficiently.
However, these metrics are not measuring the video quality received by the end users. A common measure for the presented media service is the user satisfaction, therefore, the end users specify their QoS requirements that are considered in the network resource allocation.

1.3.2.2. **Viewer perspective QoS.** The evaluation of the presented QoS from the end-user perspective is specified from two aspects: media image quality and media playback quality. The resolution distortion of the video images received by the end user is used to evaluate the media image quality. This distortion mainly occurs due to the loss of packets in the transmission. Two methods are used to measure the resolution distortion: subjective and objective quality measures. In the subjective approach, a group of viewers compare between the source and received videos and then evaluate the media quality on a scale from 1 to 5 called mean opinion score (MOS) as shown in Figure 1.5, on which 5 indicates the best quality and 1 is the worst. This method provides reliable assessment for the received video quality but requires a large group of viewers with costly viewing equipment. The objective quality metric measures the quality of the received video images using the peak signal-to-noise-ratio (PSNR) computed as:

\[
PSNR = 10 \log_{10} \left( \frac{1}{PQ} \sum_{i=0}^{K-1} \sum_{j=0}^{L-1} \frac{255^2}{[v(i,j) - w(i,j)]^2} \right),
\]

where \(v(i,j)\) is the original version of video frame, \(w(i,j)\) is the distorted version of that frame, \(P\) and \(Q\) are the number of rows and columns in the input image (i.e. the size of the frame) and 255 is the peak pixel value.

The PSNR is an objective metric that is used to quantify the spatial aspect of the video quality. One of the drawbacks of this metric is that, it requires a former knowledge of the original video stream at the receiver side to be compared with the received video as a reference.

The media playback quality can be a metric that assess the smoothness of video playback as well as the start-up delay. The smoothness of the video playback is evaluated by the playback frozen frequency as the video pauses during the playback process due to the buffer starvation. While the start-up delay is the time interval between the instant at which the user initiates a request for a video sequence up to the
instant it starts the playback. The video playback frozen frequency should be as low as possible to allow uninterrupted media, also the start-up delay should be small especially for live programs. Users require different quality metrics according to their applications, for live streaming programs, users prefer minimum start-up delay while a smooth playback is preferred for VoD applications.

1.3.3. Scalable video coding. A great challenge encounters the utilization of CR networks to provide the best video quality under the time-varying capacity and conditions for the wireless channels. Therefore, it is required to adapt the source rate according to the variations of the channel bandwidth to ensure a continuous video playback [32]. For instance, the source rate should be reduced when the available bandwidth decreases to avoid the buffer starvation which causes interruptions in the video playback at the cost of degradation in the video quality. To achieve such a control for the source rate, several scalable video coding (SVC) techniques have been developed to provide scalability to heterogeneous network channels. In these techniques, the video sequence is encoded once at the transmitter side and can be decoded in several ways at the receiver side based on the channel conditions and the users’ buffer capabilities. There are two classes for SVC, layered coding and multiple description coding (MDC). The video sequence in the layered coding is encoded into one base layer (BL) and multiple enhancement layers (ELs). The BL provides the minimum basic quality which can be further improved by receiving more ELs.
However, the ELs cannot be decoded without receiving the BLs. Figures 2.7(a), 2.7(b) and 2.7(c), show the PSNR quality provided by the base and enhancement layers in the first GoP for three different scalable videos, namely, Sony Demo, Star Wars I and Star Wars II.

Scalable video coders are of different implementations, this includes coarse-grained scalable (CGS), medium grained scalable (MGS) and fine-grained scalable (FGS). H.264/SVC is the standard for scalable video coding [30], this standard adds scalability to the advanced video coding standard H.264/AVC [33]. The hierarchical B frame prediction was introduced in H.264/SVC and H.264/AVC coding standards, while the MPEG-4 and its preceding standards were employing the classical B frame prediction that prohibits the prediction of B frames from other frames of the same type.

Several scalability approaches are supported by the H.264/SVC such as, temporal scalability, spatial scalability and quality SNR scalability [34]. A general illustration for those techniques is shown in Figure 1.7. A hierarchical structure is constructed between the frames of one GoP in the temporal (frame rate) scalability, according to this structure, the frames in lower layers are used as references to predict the higher layers frames. While in the spatial scalability, the motion compensation is used in each layer to predict the dependent frames. Therefore, spatial layer \( n \) for example in one frame is predicted from the \( n^{th} \) spatial layer of other frames and its lower spatial layers within the same frame [34]. Up to eight quality layers with similar spatial resolution are provided by the CGS in H.264/SVC, which enhance the frames PSNR quality [34]. The FGS is introduced to overcome the poor rate-distortion performance of the CGS. The later FGS encoding scheme is uniquely characterized by truncating the bit stream of the EL at any position with valid remaining bits for decoding at high computational cost. Therefore, a middle technique for coding between CGS and FGS has been designed by H.264/SVC standard called medium grain scalability (MGS) which divides the remaining DCT coefficients of a CGS layer into a number of MGS layers. For example, the EL can be partitioned up to 16 layers in the MGS coding.

Multiple description coding (MDC) is another coding technique in which the video sequence is encoded into a number of bit-streams (descriptions) that can be decoded separately. In MDC, the descriptions are independent; so that the video quality is improved by receiving more descriptions [29].
(a) Sony Demo scalable video.

(b) Star Wars I scalable video.

(c) Star Wars II scalable video.

Figure 1.6: PSNR quality provided by the base and enhancement layers of one GoP.
1.4. Related Work

1.4.1. Multimedia services over CR networks. Multimedia streaming services over CR networks have been recently investigated by the research community. Generally, from users’ point of view, the QoS at the application layer is the most significant requirement in multimedia transmission. Some of the works and efforts done in CR research focused mainly on maximizing the throughput of the SUs in their design approaches. Though, maximizing the SUs’ throughput does not guarantee the required QoS at the application layer for multimedia applications as shown by recent studies in [35]. Moreover, cross-layer multimedia transmission has been widely investigated in the literature, in which several optimization parameters are jointly considered across the various Open System Interconnection (OSI) layers. The important QoS issue has also been considered in [36], [37], [38] and [39]; in which they focused on metrics such as throughput and delay. In [36], the authors proposed 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. While in [37], the capacity of the secondary users in a cognitive radio system was investigated against the distributed and dynamic primary user activity that is detected using a two-switch model for two different schemes, frequency hopping and coding. Cross-layer MAC protocols are proposed in [38], that integrate several functions at the PHY and MAC layers to allow for spectrum utilization with a limited interference to primary users. Secondary users’ throughput and delay

Figure 1.7: Differences between the layered coding techniques [31].
performances were investigated for different network cases. In [39], a cognitive MAC protocol is proposed to establish a CR ad-hoc 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. The authors in [40] proposed an adaptive technique for transmitting the video over Ultra Wideband cognitive network using MAC-based cross layer framework, that takes into consideration the QoS requirements and the time-varying channel conditions for optimum resource allocation. In general, their framework is composed of several modules for sensing, resource allocation, video traffic and quantization scale. The PSNR and job failure rate (JFR) were used to evaluate the performance of the proposed scheme. More works done on video streaming over CR networks considering both unicast and multicast transmission are explained next.

1.4.2. **Unicast transmission.** With unicast transmission, each SU in the CR network who might want to view a specific video will receive a devoted video stream from the server at the CR base station. Contrasted with multicast transmission, where all SUs on the network receive the same streamed video. On the unicast video transmission research efforts; a CR network was assumed in [41], with one licensed user and ‘N’ cognitive users who access the spectrum opportunistically when it becomes idle to transmit video streams to a receiver. The problem of spectrum allocation was formulated as an auction game with three distributed allocation algorithms proposed. In [42], a cross layer scheme with spectrum sharing and joint routing is introduced to enhance the overall throughput, subject to the video sequence delay and sensitivity constraints using prioritized queuing model. The traffic and dynamic network conditions were used to estimate the expected delay. While the authors in [43] introduced another cross layer approach that takes into consideration the queue status as well as the channel quality for the users to allocate the available resources while satisfying their real-time requirements. Real-time video streaming has been investigated in [44], where a resource allocation algorithm was proposed for an OFDMA uplink case to achieve the SUs real-time requirements. A scheduling algorithm that is aware of the channel condition and the queue status was employed to optimally assign the uplink resources to the best SUs using a certain utility function.

1.4.3. **Multicast transmission.** A wide range of research efforts have considered video multicasting over CR networks. The QoS was considered in [45] for
live video streaming over centralized CR network, where multiple video sequences are transmitted to multiple groups of SUs. A mixed integer non-linear programing (MINLP) optimization problem was formulated to enhance the overall quality of the streamed videos with minimum distortion. The enhancement layer of the Fine Grained Scalable (FGS) video sequence is partitioned into multiple sub-layers with different data rates and modulation-coding schemes using three efficient algorithms, namely, sequential fixing, greedy algorithm and equal allocation. Then, the video packets are scheduled on the available time slots using an algorithm for tile scheduling. The SUs in each group receive different quality-based videos according to their capabilities of decoding the transmitted encoded videos.

In the multi-hop multicasting protocol proposed in [46], SUs communicate in a cooperative manner to mitigate the losses and control the error using adaptive network coding. The SUs’ throughput and the interference to licensed PUs were used as performance metrics. In [47], the video streaming problem over multi-hop CR network is formulated as a MINLP optimization problem that aims to guarantee the proportional fairness among SUs and optimize the video quality with a sequential fixing algorithm. Several researches also focused on channel adaptive multimedia streaming over CR network to enhance the application layer QoS and reduce the video distortion. A streaming video communication system that considers multiple users with different data rates was investigated in [48], in which a Markovian game problem with dynamic switching control was formulated with improvements in the resultant video quality.

The authors in [49] proposed a framework for video transmission over CR networks in which the channel was modelled using a two-state Markov model. The objective was to optimize the bit rate of the enhancement layer of each frame under the available bandwidth budget constraint in order to achieve the maximum possible video quality with minimum fluctuations. The resource allocation optimization problem was solved by applying a dynamic programming algorithm under three levels of optimization: frame, GoP and scene. The work in [50] introduced a new scalable video coding framework, namely, MPEG-4 Fine Granular Scalability (FGS), with a detailed analysis and comparisons with other coding tools. The delay and throughput were used as metrics for QoS. In [51] and [52], it was proposed to design single layer techniques to guarantee QoS for multimedia transmission over CR networks, where the PUs were protected from the interference caused by the SUs using MAC scheme in [52], while in
the PUs arrival was modelled as a Poisson random process with a quality-based metric for selecting the required sub-channels for SUs from the available spectrum. The spectrum resources were optimized considering the activity of the PU and Fountain codes were used to recover from the interference occurs in CR links. A video multicast service in CR networks that considers SVC, was addressed in [53] by formulating an optimization problem that targets the improvement of video quality while guaranteeing proportional fairness among SUs and maintaining the interference to the PUs below a pre-set threshold. Video layering and bandwidth allocation problems were considered in [54]. A utility function was modelled for all receivers and then optimized using decomposition-based algorithms to allocate the bandwidth between the video sessions and layers.

The availability of the PU’ channel was modeled as an on-off Markov process in [55], then an optimization framework was formulated for the channel allocation based on the playback buffer storages of different SUs and the available bandwidth capacity. The objective was to maximize the overall throughput of the network while ensuring smooth video playback for the end users. A central CR base station transmits on-demand videos from a massive video server to a number of SUs on multiple non-overlapping FDMA channels, by exploiting the available idle time intervals to achieve the best video quality in terms of the smoothness of video playback rather than the known objective quality metric PSNR. In [56] and [57], feedback from the radio channel was used to minimize the video distortion of the SUs for real-time non scalable video transmission. Finally, [58] and [59] conduct similar studies for multi-hop CR and Femto-cell CR respectively.

The work in [60] addressed the issue of the opportunistic spectrum access by introducing a flexible scheme for sensing the PUs’ spectrum in order to maintain longer transmission periods for SUs. Then, a utility-based scheme for SVC video transmission was developed to improve the perceived video quality by assigning utilities to the network layer units (NALUs) of each scalable video sequence according to their influence on the video quality, then maximizing the total utility of the transmitted video. SVC was also used in [61] to encode the video sequence into base and enhancement layers to provide a continuous uninterrupted video transmission to SUs by sending the base and enhancement layers together in an overlay mode, in which both primary and secondary users can share the spectrum simultaneously, while sending the base layer
only in the underlay mode where only low power and data rate can be used for transmission to avoid the interference to the PUs. An I-frame (intra-coded frame) insertion technique was introduced to protect the base layer from the packet loss to achieve an acceptable level of video quality.

1.5. Thesis Objectives

Driven by the developing interest in multimedia services over wireless networks, and the opportunities that CR networks can offer to facilitate the delivery of such services, we focus on the problem of allocating the available resources in the CR network to provide a video streaming service that satisfies the SUs’ requirements. Moreover, we focus on the scheduling of the video sessions on the allocated channels by taking advantage of the recent advances in video coding, namely, SVC techniques that allow for adaptive source rate control according to the heterogeneous network conditions and the available CR resources. The objective of this work is to develop algorithms that would maintain fair allocation of CR network resources for streaming multiple video sessions to SUs, so that they can enjoy continuous video playback with acceptable perceptual quality.

1.6. Research Contribution

The contributions of this research work can be summarized as follows:

- Propose channel allocation algorithms that adaptively assign the available channels to SUs based on their updated feedback information to satisfy their real-time requirements.

- Introduce video streaming algorithms that benefit from the SVC techniques to schedule the frames from different coded layers on the allocated slots subject to the constraints that guarantee a specific QoS to the end SUs. The developed algorithms are characterized by their reduced complexity as they do not require any reconfiguration for the codec parameters.

- Propose several video streaming architectures that focus on allocating the spectrum opportunistically to SUs in a CR network, while integrating the proposed channel assignment and streaming algorithms with adaptive modulation and scalable source coding mechanisms to offer video services, such as real-time one way video streaming and on demand video streaming, with continuous playback and acceptable perceptual quality.
Propose a scheme that integrates adaptive modulation and channel coding along with error control and scalable source coding techniques, to improve the perceptual quality of the received videos while maintaining a continuous playback. A probabilistic-based streaming algorithm that is aware of the buffer occupancies of the SUs is employed to control the source rate of the scheduled frames on the allocated slots to meet a predefined target probability for receiving the delivered frames correctly by their deadlines.

1.7. Thesis Organization

The rest of the thesis is organized as follows: Chapter 2 introduces the proposed on-demand video streaming architectures alongside the utilized algorithms. Chapter 3 presents the implementation of a real-time video streaming architecture as well as the performance evaluation for the employed streaming algorithms. Chapter 4 presents a new streaming scheme that uses joint adaptive modulation and channel coding technique integrated with error and source rate control mechanisms. Finally, Chapter 6 concludes the thesis and outlines the future work.
Chapter 2. On-Demand Video Streaming over CRN Architectures

In this chapter, we formulate the problem of streaming scalable video sequences from a base station (BS) to multiple secondary users (SUs) over a CR network. The BS solves a scheduling problem for streaming the video sequence for each SU at the beginning of each available time slot to determine the number of the enhancement layers (ELs) frames to be transmitted along with their base layer (BL) frames, under the maximum bit budget and delay deadline constraints. Different architectures and algorithms for on-demand video streaming over CR network are proposed in this chapter. Finally, the performance evaluations of the proposed streaming and allocation algorithms are presented.

2.1. Problem Formulation

We consider an infrastructure-based CR network with BS that transmits multiple scalable videos to multiple SUs. The video sequences are encoded using the H.264/SVC encoder and saved onto a server located at the BS as shown in Figure 2.1. Our work aims to ensure a continuous playback and improve the quality of streamed videos to the SUs by exploiting the available opportunities for transmission over the PUs’ channels.

Figure 2.1: Proposed video streaming scenario.
In this model, there are $N$ channels licensed to $N$ PUs and can be opportunistically accessed by one or more of the $M$ SUs as long as the channels are idle. We model the availability of the PUs’ channels using a discrete-time two-state Markov model. An overlay mode of transmission was assumed, where the primary and secondary users cannot simultaneously access the CR spectrum. Therefore, once a channel is declared as idle by the BS, it will remain available until the end of a certain period of time denoted as $T_{\text{slot}}$ with no interference to the PUs.

The available channels are allocated to SUs based on their feedback information, which include their buffer occupancies as well as the measured SNR. Once the channel is declared to be idle for a specific time period, the BS then schedules the video frames from different coded layers under the constraints of maximum bit budget ($B_{\text{max}}$) for each channel and the delay deadline, using different streaming algorithms to determine the number of the ELs to be scheduled along with the BL frames. The BL frames are always included in the stream to be sent, because they contain the basic set of quality information that ensure the continuity of the video playback. Moreover, the ELs frames cannot be decoded without the BL frames as previously discussed.

We assess the performance of the proposed video streaming schemes using the PSNR objective metric discussed in Chapter 1; that is typically used to quantify the spatial aspect of the video quality. We also use the average PSNR video quality given by

$$\bar{Q} = \frac{1}{N_f} \sum_{i=1}^{N_f} Q_i,$$

(2.1)

where $Q_i$ is the received PSNR quality for frame index $i$, and $N_f$ is the total number of frames in the received video sequence.

### 2.2. System Model

Figure 2.1 shows a CR base station that multicasts the encoded video sequences to multiple SUs during the idle intervals on subset of the $N$ channels. The radio channels are assumed to undergo Rayleigh flat-fading with AWGN. We assume that the availability of each time slot follows a discrete-time two-state Markov model with a transition probability matrix $P_n$, where $n$ is the channel index. The duration of an idle time slot ($T_{\text{slot}}$) assumed to be fixed. Each video sequence is encoded using a scalable
video encoder (H.264/SVC) into one BL and multiple ELs with either Fine Grained Scalability (FGS), Coarse Grained Scalability (CGS) or Medium Grained Scalability (MGS). These scalability coding techniques allow the BS to adapt to the dynamic network conditions and the intermittent availability of the PUs’ channels to deliver the video sequences with the highest possible quality and minimum discontinuities. Video frames form different coded layers are sent on the available time slots over the allocated channels for each SU as selected by the BS according to proposed channel allocation and streaming algorithms that will be discussed later in this chapter.

![Figure 2.2: Architecture of the video streaming system over CR channels.](image)

The architecture of the proposed video streaming system over CR channels is shown in Figure 2.2. Generally, it consists of a video transmitter, PUs’ channels and video receivers. The raw video sequence for each SU is encoded at the transmitter side into one BL and multiple ELs using the H.264-SVC video encoder. Then, the encoded frames from different layers are stored into a transmission buffer for each SU. One or more of the PUs’ channels can be used to transmit the video frames of one SU as decided by the BS according to a devised channel allocation algorithm. Adaptive modulation is employed on the allocated channels based on the feedback channel state information (CSI) provided by the SUs. The last part in the transmitter block is the streaming controller that implements the proposed streaming algorithms to schedule the video frames on the allocated slots. On the SU receiver side, the received frames are
demodulated and checked for errors, then the coded layers for each received frame are applied to the video decoder. Finally, the frames are forwarded to the playback buffer before being displayed. The SU receiver provides an updated buffer occupancy status and CSI to the transmitter on a reverse reliable channel in regular time periods of $T_{\text{slot}}$ duration.

2.3. Channel Model

A discrete-time two-state Markov chain is used to model the availability of the PUs’ channels at each time slot of duration $T_{\text{slot}}$. We assume that there are $N$ PUs in the CR network (and hence $N$ channels). Figure 2.3 shows the state transition diagram for the ON/OFF model for each PU channel $n$, ($n = 1, 2, \ldots, N$) with two distinct states. State “0” (OFF state) indicates an idle channel that is available for the SUs transmission (i.e. PU is offline), while state “1” (ON state) indicates a busy channel by the PU transmission (i.e. PU is online). The $N$ PUs’ channels change their states independently with the same or different transition matrices according to the PUs activities.

![Two-state discrete-time Markov model for the n-th PU channel.](image)

The transition matrix that describe the availability behaviour of the $n^{th}$ channel is given by

$$P_n = \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix} = \begin{bmatrix} 1 - \alpha_n & \alpha_n \\ \beta_n & 1 - \beta_n \end{bmatrix},$$

(2.2)

where $\alpha_n$ is the transition probability from the idle state to busy state and $\beta_n$ is the probability of transition from the busy state to idle state ($0 \leq \alpha_n, \beta_n \leq 1$). Based on the historical accumulated observations, the limiting (i.e. steady state) probability of finding channel $n$ as idle (i.e. OFF state), is $\pi_{0,n} = \beta_n/(\beta_n + \alpha_n)$ [62].
Generally, if all the $N$ channels follow the same transition matrix, $P_n = P$ for all $n$, and hence have the same steady-state probability $\pi_0$, then the channels availability can be modeled as independent and identically distributed (i. i. d) Bernoulli trials. The probability of finding $k$ idle channels can be estimated using the binominal probability distribution as follows:

$$Pr(N_{idle} = k) = \binom{N}{k} \pi_0^k (1 - \pi_0)^{N-k}. \quad (2.3)$$

While, if the $N$ channels have different availability patterns that follow different transition matrices, $P_n$, and hence have different limiting probabilities, $\pi_{0,n}$, where $n = 1, 2, \ldots N$, in this case, the channels availability can be modeled as independent and non-identically distributed (i. n. i. d) Bernoulli trials. If the PUs’ channels have non-homogeneous idle probabilities (i.e. some probabilities are too small or too large and the others are moderate values), then the probability of finding number of idle channels (i.e. random variable, $N_{idle}$) can be approximated using the Poisson-Normal approximation introduced in [62].

The PUs’ channels in our proposed system are assumed to be Rayleigh flat-fading channels. It is also assumed that the channel status does not change during one transmission slot with slowly varying SNR over multiple successive time slots. The error-free channel bit rate or capacity $C$ for a given SNR is calculated by Equation (2.4) for AWGN channel

$$C = B \log_2 \left(1 + \frac{E_s}{N_o}\right) = B \log_2 \left(1 + \frac{kE_b}{N_o}\right) \quad \text{(bits/s)}. \quad (2.4)$$

While for Rayleigh flat-fading channels, the instantaneous capacity is given by

$$C = B \log_2 \left(1 + \frac{E_s|\delta|^2}{N_o}\right) = B \log_2 \left(1 + \frac{kE_b|\delta|^2}{N_o}\right) \quad \text{(bits/s)}, \quad (2.5)$$

where $B$ is the channel bandwidth, $E_s$ is the energy per symbol, $E_b$ is the energy per bit, $k$ is the number of bits per symbol, $N_o$ is the two sided noise power spectral density, and $\delta$ is a random variable ($\delta \geq 0$) that represents the channel gain and follows a Rayleigh distribution with a second moment $E[\delta^2] = \omega$ and probability density function (PDF) given by [63]:
\[ P_\delta(\delta) = \frac{2\delta}{\omega} \exp\left(\frac{-\delta^2}{\omega}\right). \] (2.6)

The SNR fed back from the SUs is exponentially distributed with the following PDF [64]:

\[ P_\gamma = \frac{1}{\gamma'} \exp\left(\frac{\gamma}{\gamma'}\right). \] (2.7)

where \( \gamma \) represents the *instantaneous* value of the received SNR per symbol \( \triangleq \delta^2 k E_b / N_0 \), while \( \gamma' \) is the *average* value of the received SNR per symbol given by:

\[ \gamma' = \bar{\delta}^2 k E_b / N_0. \]

2.4. Adaptive Modulation

Adaptive modulation is used to maximize the spectral efficiency and enable the transmission of the video information at higher data rates, and hence achieve better perceptual quality at the SUs end. However, it is difficult to achieve such an objective unless a control is exerted over these adaptive modulation techniques to compromise between the data rates and the corresponding BER [65]. In the proposed streaming system, the modulation level that is used for transmission depends mainly on the fed back CSI (i.e. measured SNR values of different channels), as well as the target BER. Generally, a lower modulation level is used when the channel condition is bad (e.g. binary phase shift keying (BPSK)). While, a higher modulation level is used when the channel condition is good, e.g. 64-quadrature amplitude modulation (64-QAM). At the end, choosing any of the modulation levels according to the channel condition will result in a certain BER and an effective bit rate \( C_e \) as given by Equation (2.8), when no channel coding is used. Table 2.1 shows the effective bit rates for different modulation schemes used in our models, where \( C \) is the error-free channel capacity.

\[ C_e = C \log_2 L. \] (2.8)

<table>
<thead>
<tr>
<th>Modulation scheme</th>
<th>BPSK ((L = 2))</th>
<th>4-QAM ((L = 4))</th>
<th>16-QAM ((L = 16))</th>
<th>64-QAM ((L = 64))</th>
<th>256-QAM ((L = 256))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effective data rate ( (C_e) )</td>
<td>( C )</td>
<td>( 2C )</td>
<td>( 4C )</td>
<td>( 6C )</td>
<td>( 8C )</td>
</tr>
</tbody>
</table>

Table 2.1: The effective channel bit rate for different modulation schemes.
Increasing the modulation level comes at the cost of the BER for a given SNR, because increasing the modulation level $L$, results in more bits per symbol $k = \log_2 L$, which in turn increases the corresponding BER as shown in Figure 2.4. In our streaming systems, we use the following modulation schemes: BPSK, 4-QAM, 16-QAM, 32-QAM, 64-QAM and 256-QAM. The selection of a modulation level is based on the channel condition (in terms of the reported SNRs from the SUs) and the ability to satisfy the target BER ‘$\psi$’ decided for transmission. The target BER ‘$\psi$’ is either set to a fixed value or according to the importance of transmitted layer of each frame. After the target BER has been identified, the highest modulation level that achieves the closest BER value to the target BER is used for sending that video frame on the allocated channel(s).

Over a Rayleigh flat fading channel, Figure 2.4 shows that to achieve low target BER (e.g. $10^{-5}$), a lower modulation level (BPSK) should be employed up to a high value of the channel SNR of about 45.7 dB. While the same modulation scheme (i.e. BPSK) can be used for lower SNR values up to about 35.7 and 25.7 dB, when the target BER is relaxed to $10^{-4}$ and $10^{-3}$, respectively. Therefore, when the target BER is set to low value, we send the video frames at lower data rates, thus less errors will be encountered in the delivered frames. On the other hand, high data rate is achieved when the target BER is set to high value, which will result in introducing more errors in received frames.

The pseudocode for this adaptive modulation algorithm is given in Algorithm (1). The general expression for the BER over Rayleigh flat fading channels for any $L$-QAM modulation is given by Equation (2.9) [63, 64]:

$$P_b \approx \frac{2}{\log_2 L} \left(1 - \frac{1}{\sqrt{L}}\right) \left(1 - \sqrt{\frac{1.5 \gamma' \log_2 L}{L - 1 + 1.5 \gamma' \log_2 L}}\right).$$  

Therefore, the BER for the modulation schemes (BPSK, 4-QAM, 16-QAM, 64-QAM and 256-QAM), used for transmission over a Rayleigh flat-fading channels are given by the following set of equations, assuming coherent detection:
\[ P_b = \begin{cases} \frac{1}{2} \left( 1 - \frac{\gamma'}{1 + \gamma'} \right), & L = 2, 4 \\ \frac{3}{8} \left( 1 - \frac{1}{\sqrt{1 + \frac{5}{2} \gamma'}} \right), & L = 16 \\ \frac{7}{24} \left( 1 - \frac{1}{\sqrt{1 + \frac{7}{4} \gamma'}} \right), & L = 64 \\ \frac{15}{64} \left( 1 - \frac{1}{\sqrt{1 + \frac{85}{4} \gamma'}} \right), & L = 256 \end{cases} \] (2.10)

Figure 2.4: BER for Rayleigh flat-fading channel under different modulation schemes.

Algorithm 1: Adaptive modulation assignment algorithm.

**Divide**: The source encoder output into \( l \)-layers of unequal priorities // e.g. one BL and two ELs

**for** \( i = 1, 2, ..., l \) // for each layer of a frame

**do**

- Determine the priority \( \rho_i \) of the \( i \)-th layer;
- Determine the desired target BER \( \psi_i \) for the \( i \)-th layer;
- Find the SNR thresholds that achieve the target BER \( \psi_i \);
- Get an estimate of the received SNR \( \gamma' \);
- Compare \( \gamma' \) against the SNR thresholds;
- Select the corresponding modulation level \( L_i \) using the lookup tables;

**end**
2.5. Video Encoding

As previously mentioned in Chapter 1, depending on the GoP structure, each video frame is either independent intra-coded frame (I), dependent forward predictive coded frame (P) that can be predicted from the previous I or P-frame, or dependent bi-directional predictive coded frame (B), which is predicted from the previous as well as the next I or P-frame. The MPEG-4 and previous MPEG standards employ the classical B-frame prediction, in which the B-frame is predicted from the previous as well as the next I or P-frame using motion compensation encoding. In this classical approach, the B-frame cannot be referenced for another B-frame. While in recent video encoding standards (e.g. H.264/SVC), the B-frame can be used as a reference to predict other B-frames in the hierarchical B-frame prediction [34]. The two approaches differ mainly in the encoding order of the frames. A given frame can be encoded only after all its reference frames are encoded successfully. This order of encoding is the same as the order in which we transmit the frames on different channels [34].

The video sequences in our streaming systems are encoded using the H.264/SVC standard that employs the hierarchical B-frame prediction. The input to the encoder is the video sequence with frames arranged in the order of display, which is the original order of capturing the frames when shooting the video, also used for playing back the video for display. The frames are encoded for transmission in different order known as encoding or transmission order based on the frame dependencies in the motion compensation prediction. For example, in the hierarchical B-frame prediction in G16B3 structure used in our experiments with the pattern I₁, B₂, B₃, B₄, P₅, … etc. the frames are encoded for transmission in the order I₁, P₅, B₃, B₂, B₄, … etc., which is different from the encoding order in the classical B frame prediction in other video encoding standards.

2.6. Dynamic Resource Allocation Subsystem

The proposed video streaming system aims to improve the quality of the streamed scalable video sequences while maintaining the continuity of the video playback at the SUs end. To achieve such goal, two algorithms are introduced, namely, channel allocation and streaming algorithms. The channel allocation algorithm aims to satisfy the needs of the SUs when allocating the available channels. While the streaming algorithm is employed to schedule the video session frames within the slots of the allocated channels based on certain constrains as we will see next.
2.6.1. Channel allocation algorithms. The available channels identified by the BS are allocated to SUs based on their feedback information, received regularly every $T_{slot}$ seconds on a reliable error-free reverse channel, which includes their buffer occupancies as well as the measured SNR of different channels. To elaborate, the BS allocates the available channels to SUs by comparing their buffer occupancies with a predefined occupancy threshold $\Delta_{th}$ to determine the urgency of sending the video frames to each SU. The threshold $\Delta_{th}$ is set to a relatively large value ($\approx T_{slot} f_p$) to overcome the problem of intermittent spectrum availability. As shown in Figure 2.5, the algorithm is summarized for the following two cases:

1) Case 1: all SUs have their buffer occupancies larger than the threshold ($\Delta_i > \Delta_{ch}$, $\forall i = 1, 2, ..., M$), (i.e. enough frames to be played back), or all their buffer occupancies are below or equal the threshold $\Delta_i \leq \Delta_{ch}$, $\forall i = 1, 2, ..., M$ (i.e. starvation mode). In this special case the BS allocates the available channels among all the SUs equally in a Round Robin manner, starting with the SU of the lowest occupancy.

2) Case 2: some SUs are underflowing with buffer occupancies below the threshold ($\Delta_i \leq \Delta_{ch}$, $i = 1, 2, ..., M_u$) where $M_u < M$, which threatens the continuity of the video playback. In this case the BS allocates the available channels equally between those underflowing users only to avoid their buffer starvation, and hence assuring the continuity of their videos playback, while ignoring the other SUs who have enough frames in their buffers to be played back.

In both cases, the BS takes into account the instantaneous buffer occupancies of the SUs, ($\Delta_i$, $i = 1, 2, ..., M$), as well as the reported SNR of the different channels as seen by the SUs in the CR network, ($\text{SNR}_i$, $i = 1, 2, ..., M$), where $\text{SNR}_i = [\text{SNR}_{i,1}, \text{SNR}_{i,2}, ..., \text{SNR}_{i,N}]$ is a vector of the SNRs measured on the $N$ channels. Let $u$ indicate the index of the channel to be allocated to a certain SU and $M_u$ be the number of SUs who are competing on this channel (i.e. $M_u = M$ for case (1) or $M_u = M_a$ for case (2)). The $M_u$ SUs send the following SNRs to the BS as they measure on the $u^{th}$ channel, ($\text{SNR}_{u,1}, \text{SNR}_{u,2}, ..., \text{SNR}_{u,M_a}$). Assume that the $j^{th}$ SU has the largest SNR reported for the $u^{th}$ channel, hence, $\text{SNR}_{u,j} = \max \{\text{SNR}_{u,1}, \text{SNR}_{u,2}, ..., \text{SNR}_{u,M_a}\}$. Therefore, the BS allocates channel $u$ to SU$_j$ who can achieve the highest quality when
using this channel compared to other SUs. If two or more SUs have the same level of quality on a channel, then the BS allocates that channel to one of them randomly (i.e. flipping a coin). The BS continues repeating the same allocation procedure until all the available PUs’ channels are fully allocated to the SUs. The detailed description of this allocation algorithm is given as a pseudocode in Algorithm (4). On the other hand, the allocation of the available channel(s) can be decided based on the buffer occupancy status of the SUs only. In this case the SUs with lower occupancies are allocated first to avoid their buffers starvation, however, they may not achieve a good quality on the allocated channel(s). While the allocation that is based only on reported SNRs may assign the channel(s) with good quality to SUs, though, no attention is paid towards their buffer storages, and that may cause buffer starvation events for some SUs and threaten the continuity of their videos playback. The pseudocodes for the allocation techniques that are based only on the buffer status or the SNR are explained in Algorithms (2) and (3), respectively.

Figure 2.5: Joint based Channel allocation algorithm flow chart.
Algorithm 2: Buffer based channel allocation algorithm.

**Input:** \( N_{idle}, \Delta_i, i = 1, 2, ..., M \).

**Output:** \( Ch_{s1}, Ch_{s2}, ..., Ch_{sM} \) // Vector of channels allocated to each SU.

Set \( \Delta_{th} = x \); // Buffer Occupancy threshold.

if \( \Delta_i > \Delta_{th} \) or \( \Delta_i \leq \Delta_{th}, \forall i \) // same case to all SUs (\( M_a = M \)).

then

sort \( \Delta_1, \Delta_2, ..., \Delta_M \); // arrange the occupancies in ascending order.
Allocate the channels equally to all SUs on Round Robin basis according to their ascending order;
evaluate \( (N_{s1}, N_{s2}, ..., N_{sM}) \);

else

\( \Delta_i \leq \Delta_{th}, i = 1, 2, ..., M_u \) // only some SUs are underflowing (\( M_a = M_u \)).

sort \( \Delta_1, \Delta_2, ..., \Delta_{M_u} \); // arrange the occupancies in ascending order.
Allocate the channels equally to underflowing SUs only on Round Robin basis according to their ascending order;
evaluate \( (N_{s1}, N_{s2}, ..., N_{sM_u}) \);

end

Algorithm 3: SNR based channel allocation algorithm.

**Input:** \( N_{idle}, SNR_i, i = 1, 2, ..., M \).

**Output:** \( Ch_{s1}, Ch_{s2}, ..., Ch_{sM} \) // Vector of channels allocated to each SU.

Allocate the channels equally to all SUs on Round Robin basis;
evaluate \( (N_{s1}, N_{s2}, ..., N_{sM}) \);

for \( n = 1, 2, ..., N_{idle} \) // for each available channel

\( SNR_{n,j} = \max(SNR_{n,1}, SNR_{n,2}, ..., SNR_{n,M}) \);
Allocate channel \( n \) to the SU \( j \) with max SNR;
\( N_{s_j} \leftarrow N_{s_j} - 1 \); // the remaining allocated channels decided for SU \( j \).

end

Algorithm 4: Buffer-and-SNR based channel allocation algorithm.

**Input:** \( N_{idle}, (\Delta_i \text{ and } SNR_i), i = 1, 2, ..., M \).

**Output:** \( Ch_{s1}, Ch_{s2}, ..., Ch_{sM} \) // Vector of channels allocated to each SU.

Set \( \Delta_{th} = x \); // Buffer Occupancy threshold

if \( \Delta_i > \Delta_{th} \) or \( \Delta_i \leq \Delta_{th}, \forall i \) // same case to all SUs (\( M_a = M \)).

then

sort \( \Delta_1, \Delta_2, ..., \Delta_M \); // arrange the occupancies in ascending order.
Allocate the channels equally to all SUs on Round Robin basis according to their ascending order;
evaluate \( (N_{s1}, N_{s2}, ..., N_{sM}) \);

for \( n = 1, 2, ..., N_{idle} \) // for each available channel

\( SNR_{n,j} = \max(SNR_{n,1}, SNR_{n,2}, ..., SNR_{n,M}) \);
Allocate channel \( n \) to the SU \( j \) with max SNR;
\( N_{s_j} \leftarrow N_{s_j} - 1 \); // the remaining allocated channels decided for SU \( j \).

end

elseif \( \Delta_i \leq \Delta_{th}, i = 1, 2, ..., M_u \) // only some SUs are underflowing (\( M_a = M_u \)).
sort ($\Delta_1, \Delta_2, \ldots, \Delta_{M_u}$);  // arrange the occupancies in ascending order.
Allocate the channels equally to underflowing SUs only on Round Robin basis according to their ascending order;
evaluate ($N_{s_1}, N_{s_2}, \ldots, N_{s_{M_u}}$);
for $n = 1, 2, \ldots, N_{idle}$  // for each available channel
do
  $\text{SNR}_{n,j} = \max(\text{SNR}_{n,1}, \text{SNR}_{n,2}, \ldots, \text{SNR}_{n,M_u})$;
  Allocate channel $n$ to the SU$_j$ with max SNR;
  $N_{s_j} \leftarrow N_{s_j} - 1$; // the remaining allocated channels decided for SU$_j$.
end

2.6.2. Video streaming algorithms. The raw video sequences in our proposed streaming system are encoded using the H.264/SVC encoding standard into one BL and a number of ELs. The BL provides the basic accepted quality level, whilst the quality level can be improved by including more ELs as shown in Figure 2.6. The EL frames cannot be decoded without their corresponding BL frames. Therefore, the BL frames should be correctly received before their display deadline and the deadline of the dependent BL of other frames as well as the ELs that belong to their frames, otherwise, the dependent frames and their corresponding ELs become useless even if they are correctly received and decoded by their display times. Moreover, due to the inherited dependency between the frames of one GoP; if the reference frames are discarded, then the dependent frames should not be transmitted as they cannot be decoded without the reference frames on the receiver side. The streaming algorithms should satisfy the above dependency and priority characteristics when sending the video frames to ensure the continuity of the video playback at the maximum possible quality with efficient resources utilization.

![Figure 2.6: The arrangement of the base and enhancement frames in GoPs.](image-url)
Once a channel or more are assigned to a SU, the BS then performs the scheduling of the video frames from all layers (i.e. BL, EL₁, EL₂, ..., etc.) under $B_{\text{max}}$ constraint, which is the maximum number of bits that can be transmitted on an available slot given by Equation (2.11), while satisfying the delay deadline constraints.

$$B_{\text{max}} = T_{\text{slot}} C_e,$$  (2.11)

where $T_{\text{slot}}$ represents the duration of the available slot in seconds and $C_e$ is the effective channel bit rate with no channel coding calculated using Equation (2.8).

The objective of these streaming algorithms is to schedule the ELs of the video frames along with their essential BLs to enhance the quality of the reconstructed videos at the SUs end. A scheduled frame has the following probability of being correctly received

$$Pr_c = \sum_{m=0}^{b_{\text{max}}} \binom{S_f}{m} p_b^m (1 - p_b)^{S_f - m},$$  (2.12)

where $S_f$ is the scheduled frame size in bits, $b_{\text{max}}$ is the maximum number of bits that can be corrected on each received frame and $p_b$ is the BER of the channel on which the frame has been scheduled given by Equation (2.10). If a frame is received in error, then it will be discarded by the error checker in the receiver with all its dependent frames from the same layer and higher ELs.

Next, we briefly describe the proposed streaming mechanisms for scheduling the scalable video frames on the allocated channel(s): one-GoP, 3/4-GoP, 1/2-GoP, 1/4-GoP and frame-based streaming mechanisms. To start with the one-GoP streaming approach, as shown in Figures 2.7 and 2.8(a), the algorithm starts from the first GoP taking frame by frame from the BL. Let $S_{f_{Bi}}$ be the size of the BL of frame $i$ in bits, when one or more channels are granted to certain SU, then the total number of bits the SU will transmit from the BL of the first GoP is given by $W = \sum_{i=1}^{16} S_{f_{Bi}}$ (assuming that the full GoP size is 16 frames for example). The streaming algorithm compares $W$ with $B_{\text{max}}$ given by Equation (2.11) supported by the allocated channel(s). If $W$ is less than $B_{\text{max}}$, then all frames from the BL of the current GoP will be transmitted to guarantee the continuity of the video playback. Otherwise, the algorithm transmits the maximum possible number of BLs then waits for the following available slot(s) with a new bit budget. Only after all the BL frames of the currently transmitted GoP have been
scheduled on the available slot(s), if there is still room for more transmission, then the algorithm assigns the ELs frames from the current GoP after checking their sizes against what remains of bit budget, while taking into considerations the delay deadline constraints to ensure they are delivered by the display time of their corresponding BLs ($\varphi_{B_i}$). Otherwise, the ELs will waste the available resources and may cause interruptions in the video playback. Moreover, the quality that the EL frame introduces is compared with a predefined quality threshold ($\Delta_q$). If the enhancement in quality is equal to or above the threshold, then the EL of that frame will be transmitted otherwise the attention is moved to the EL of the next frame in the current GoP. Before the transmission of the ELs, the streaming algorithm also checks that the needed base/enhancement layers are correctly received, so that it guarantees the decoding of the current transmission. If the EL satisfies that criteria, it will be scheduled on the slot and the algorithm moves to the next EL(s) of the next frame. If a frame is discarded, then all frames that depend on it will also be discarded from the same layer as well as the higher EL(s). The algorithm moves to the next GoP after it finishes the first GoP and the process repeats until the end of the video sequence.

The 3/4-GoP, 1/2-GoP and 1/4-GoP streaming techniques, shown in Figures 2.8(b), 2.8(c) and 2.8(d), respectively, have similar mechanism as the one-GoP streaming method but on different partitions of the GoP, which we will call “streaming unit” that consists of $S_{size}$ frames. The BL frames of the first unit are scheduled for transmission as the load size $W = \sum_{f=1}^{S_{size}} S_{f_{Bi}}$ is less than $B_{max}$, then the algorithm considers the frames belong to the first enhancement layer (EL1), a given EL frame is scheduled for transmission if its size has a space on what remains of $B_{max}$ and satisfies the different constraints on the delay deadline and the introduced quality thresholds previously mentioned. Otherwise, the frame is discarded with all its dependent frames from the same EL as well as the higher ELs (i.e. EL2, EL3, ... , etc.), and the next EL frame is considered from the streaming unit under considerations. The algorithm continues to process the ELs selection on the same manner and repeats the process for the next streaming units.

The last approach introduced for streaming the video frames as shown in Figure 2.8(e), is based on the idea of scheduling the different layers of a frame, starting from the BL then EL1, EL2, ... etc. And then the attention is moved to the different layers of
the next frame till the end of the video sequence. Whenever the scheduled load $W$ exceeds the maximum budget $B_{\text{max}}$, the streaming controller waits for a new budget on the next allocated channel(s) for that user. The ELs selection is based on satisfying the constraints stated previously as given in Equation (2.13).

For all previous streaming mechanisms, the streaming controller at the BS solves a problem to schedule the video frames from different coded layers that belong to the video session to be delivered to each SU at the beginning of the available time slot(s). The objective is to maximize the quality of the delivered video while guaranteeing the continuity of the video playback at the SUs. The frames from all layers are scheduled on the allocated slot(s) under the following constraints:

$$\begin{align*}
W &= \sum_{x=1}^{X} S_{f_{Bx}} + \sum_{y=1}^{Y} S_{f_{E1y}} + \sum_{z=1}^{Z} S_{f_{E2z}} \leq B_{\text{max}} \tag{2.13} \\
Tr_{E1y} &\leq \varphi_{y} \quad \text{and} \quad Q_{E1y} \geq \Delta_{q1} \quad , \quad y = 1, 2, \ldots, Y \\
Tr_{E2z} &\leq \varphi_{z} \quad \text{and} \quad Q_{E2z} \geq \Delta_{q2} \quad , \quad z = 1, 2, \ldots, Z
\end{align*}$$

where $W$ is the total size of the scheduled frames from all layers in bits, $S_{f_{Bx}}, S_{f_{E1y}}$ and $S_{f_{E2z}}$ are the sizes in bits of frame $x$, $y$ and $z$ from the BL, EL$1$ and EL$2$, respectively. $X$, $Y$ and $Z$ indicate the total number of scheduled frames from the BL, EL$1$ and EL$2$, respectively, where $(X \geq Y \geq Z)$, $\varphi_{y}$ is the display deadline of frame indexed $y$, $Q_{Ejy}$ is the quality introduced by frame indexed $y$ from EL$j$, $\Delta_{qj}$ is the quality threshold for EL$j$ frames and $Tr$ is the estimated transmission time of the scheduled frame given by:

$$T_{r} = \frac{S_{f}}{C \log_{2}(L)}, \quad \tag{2.14}$$

where $L$ is the assigned modulation level for the scheduled frame on the allocated channel based on the criteria mentioned previously in Section 2.4.
Figure 2.7: Streaming algorithm flow chart.

(a) One-GoP based streaming.
(b) 3/4-GoP based streaming.

(c) 1/2-GoP based streaming.

(d) 1/4-GoP based streaming.

(e) Frame based streaming.

Figure 2.8: Streaming mechanisms.
Algorithm 6: Streaming algorithm.

**Input:** $C_{hsi}, T_{slot}, L$, Frames’ sizes, dependency and display deadline; // The allocated channel(s) and video frames specifications for SU$_i$.

**Output:** Number of frames to be scheduled from different coded layers.

**Initialize:** $W = 0$; // initializing the slot load counter for each allocated channel.

Set $C_e = C \log_2(L)$; // effective channel bit-rate.

Set $B_{\max} = C_e T_{\text{slot}}$; // max bit budget of the slot.

Set $S_{\text{size}} \in \{1, 4, 8, 12, 16\}$; // streaming unit size.

while $W \leq B_{\max}$ // compare the current slot load size against the max bit budget of the slot

do

$W \leftarrow W + \sum_{i=1}^{S_{\text{size}}} S_{f_{EI_i}}$; // scheduling the BL frames of the current streaming unit.

Set $i = 1$; // initializing the frame index for the current streaming unit.

while $W \leq B_{\max}$ and $i \leq S_{\text{size}}$ // scheduling the EL$_1$ frames of the current streaming unit

do

$T_{r_{EI_i}} = S_{f_{EI_i}} / C_e$; // the corresponding transmission time.

if $T_{r_{EI_i}} \leq \varphi_{BI_i}$ and $Q_{EI_i} \geq \Delta_{q1}$ // satisfying the constraints

then

$W \leftarrow W + S_{f_{EI_i}}$; // schedule the EL$_1$ frame

else

Discard the frame and all its dependent frames from the same layer and higher ELs;

end

$i \leftarrow i + 1$; // move to the next EL$_1$ frame

end

while $W \leq B_{\max}$ and $i \leq S_{\text{size}}$ // scheduling the EL$_2$ frames of the current streaming unit

do

$T_{r_{EI_i}} = S_{f_{EI_i}} / C_e$; // the corresponding transmission time

if $T_{r_{EI_i}} \leq \varphi_{BI_i}$ and $Q_{EI_i} \geq \Delta_{q2}$ // satisfying the constraints

then

$W \leftarrow W + S_{f_{EI_i}}$; // schedule the EL$_2$ frame

else

Discard the frame and all its dependent frames from the same layer and higher ELs;

end

$i \leftarrow i + 1$; // move to the next EL$_2$ frame.

end

Move to the next streaming unit;

end

Move to the next streaming unit;

until last video frame;
2.7. Proposed VoD Streaming Architectures

In this section, we will introduce the implementation of two different architectures for video streaming over CR network. Both architectures provide VoD service over a CR network and differ mainly in the algorithms used for allocating the available channels and scheduling the video frames. Moreover, different criteria for adaptive modulation are employed in each architecture.

The proposed adaptive video streaming architectures aim to enhance the perceptual quality of the received videos at the SUs end with no or less interruption in the playback. Therefore, all architectures opportunistically exploit the available channel(s) that are sensed as idle by the BS at regular intervals of time. In addition, the BS receives feedback information on a reverse reliable channel that includes the buffer occupancies of the SUs and/or the estimated SNRs of the different channels as seen by those users. Using this fed-back information, the BS is able to allocate the available channel(s) to the SUs according to the proposed channel allocation algorithms previously introduced in Section 2.6.1. In addition, the architectures also employ different streaming algorithms to schedule the video frames on the allocated time slots with a bit-rate and adaptive modulation that are decided based on the channel state and a target BER that is set according to different criteria in each architecture. Scalable video sequences are considered in the proposed architectures; to reduce the impact of the dynamic nature of the CR network by taking the advantage of the scalability feature of the video frames that enables a quick codec reconfiguration. Therefore, the employed algorithms perform the allocation of the resources to stream the video session for each SU at the beginning of the allocated time slot(s).

2.7.1. CSI-based VoD architecture. In this model we assume that the PUs’ channels have independent and identical availability behavior that follows a Markovian model with a certain transition matrix. Figure 2.9 shows the complete subsystem at the BS used in this proposed architecture. Firstly, the available channels detected by the BS are allocated to the SUs equally based on their CSI feedback only as described earlier in Algorithm (3). Then, the modulation level is selected from BPSK, 4-QAM, 16-QAM, 32-QAM, and 64-QAM, on each allocated channel based on the reported CSI to meet a predefined target BER. The same modulation level will be used to send the encoded frames from all layers scheduled on that allocated slot of a channel regardless of its impact on the quality of the reconstructed video sequence.
In the streaming controller block, based on $B_{\text{max}}$ of the allocated channel(s) and the delay deadline and quality constrains, the one-GoP based streaming algorithm discussed in Section 2.6.2, selects the BL frames of one-GoP along with a number of their corresponding ELs from the encoded video sequence stored in the transmission buffer for each SU to be sent on the slots of the allocated channel(s). If there is still a room for more transmission on the allocated slot(s) (i.e. $W < B_{\text{max}}$), then the streaming algorithm schedules frames from the different coded layers (i.e. BL, EL$_1$, EL$_2$ ... etc.), of the next GoP in the same manner. Otherwise, the streaming controller waits for new allocated slots to continue scheduling the frames of each SU’ video.

2.7.2. Joint buffer-and-CSI based VoD architecture. In this architecture, we assume that the PUs have different activity patterns on the spectrum, unlike the hypothesis assumed in the CSI-based architecture for the PUs’ channels availability.
This architecture aims to improve the continuity of the video playback at the SUs end while maintaining the perceptual quality at an acceptable level. To achieve such a goal, the channel allocation algorithm explained in Algorithm (4), that is based jointly on the current buffer occupancies of the SUs as well as their CSI feedback, is employed along with a streaming algorithm that is implemented on different partitions of the GoP as previously explained in Algorithm (5).

As shown in Figure 2.10, first, the BS continuously senses the PUs’ spectrum to detect any available channels every $T_{\text{slot}}$ duration. Then, it allocates the available channels to the SUs based on their feedback information, which includes their buffer occupancies $\Delta_i, i = 1, 2, \ldots, M$ and the CSI of the channels as seen by each SU, $\text{SNR}_i, i = 1, 2, \ldots, M$, where $\text{SNR}_i = [\text{SNR}_{i,1}, \text{SNR}_{i,2}, \ldots, \text{SNR}_{i,N}]$ is a vector of the measured SNRs on different channels. The streaming controller at the BS employs the streaming algorithm explained in Algorithm (5), to schedule the video frames within the allocated slots to meet their deadlines while achieving the highest possible quality with no interruptions in the playback. Furthermore, adaptive modulation is employed on the allocated channels as explained in Algorithm (1), to improve the spectral efficiency by enabling the transmission of the video information at higher rates and hence achieve better perceptual quality at the SUs end. A certain level of modulation among BPSK, 4-QAM, 16-QAM, 64-QAM and 256-QAM, is assigned by the BS on the allocated channels based on the fed-back CSI to meet a predefined target BER $\psi$, that is decided according to the sensitivity of the transmitted video information. The BL frames are sent with the modulation level that achieves a low target BER (e.g. $10^{-5}$), to ensure their correct reception and hence guaranteeing the continuity of the playback. While, the ELs are transmitted with higher modulation levels that might have a higher target BER (e.g. $10^{-3}$), to enable their transmission with higher rates, so that they can catch up the display deadline of their corresponding BLs. However, more errors will present on those ELs compared with their BLs counterparts because the resultant BER increases as we enhance the modulation level for a given SNR as we previously discussed.
2.8. Results and Analysis

We used the SimEvents discrete-event simulator in MATLAB to implement the proposed video streaming systems over CR networks. In all our simulations, the BS transmits three video sequences that are CGS encoded to $M = 3$ SUs denoted by SU$_1$, SU$_2$, and SU$_3$, over $N = 10$ PUs’ channels that are opportunistically accessed according to the discrete-time two-state Markov chain model. We used the H.264/SVC CGS encoded trace files for the different video sequences used in the simulations [34, 66, 67]. Each of the CGS video sequences consists of 241 total number of frames and is encoded into one BL and two ELs with 352 x 288 resolution and playback rate of $f_p = 30$ frames per second. The GoP structure of the three layers is IBBBBPBBB and is abbreviated as G16B3, with hierarchical dependency between the frames of each layer where B frames can be used as references to other B frames. The PUs’ channels are assumed to be Rayleigh flat-fading channels with
exponentially distributed SNRs. The error-free channel bit rate of each of the PUs’
channels is $C = 256$ Kbps. Next, we present the performance evaluation for the two
implemented VoD architectures in this chapter.

2.8.1. CSI-based VoD architecture. In this section, we present the results for
the first implemented video streaming architecture, in which the BS assigns the
available channels equally on Round Robin manner based on the reported SNR from
the SUs only. Moreover, the modulation level is adapted on the allocated channel to
achieve a target BER that is set to a fixed value regardless of the sensitivity of the
transmitted layer of the scheduled frame (i.e. fixed $L$ throughout the transmission on
the allocated slot). We used the one-GoP based algorithm for streaming three CGS
video sequences, namely, Sony Demo, Star Wars I and Star Wars II, which are actually
scenes truncated from the original complete videos. We start the video playback after
receiving the first GoP. We assume that the PUs’ channels are independent and identical
changing their states following the transition matrix $P = \begin{bmatrix} 0.3 & 0.7 \\ 0.4 & 0.6 \end{bmatrix}$ with
limiting probability $\pi_0 = 0.36$.

2.8.1.1. Transmission of the same video to all SUs. We first test our system
by sending the same video called ‘Star Wars I’ to all SUs in the CR network, while
setting the target BER to $10^{-5}$. The number of channels allocated to each SU on each
transmission round is shown in Figure 2.11. Figure 2.12(a), 2.12(b) and 2.12(c) show
the PSNR video quality received by SU$_1$, SU$_2$ and SU$_3$ respectively, compared with
maximum quality achieved when all the coded layers are successfully received up to
the second EL. We notice that the received video sequences show the same PSNR
quality performance in average with slight degradation in quality on the first half of the
display (about 38 dB), due to the frames discarded from EL$_2$. Therefore, the quality on
this period is achieved by displaying EL$_1$ frames and sometimes the maximum quality
of about 48 dB is achieved, whenever EL$_2$ frames are displayed. Some instants in the
next half of the display attained more degradation in quality, since both EL$_1$ and EL$_2$
frames are discarded or received in error, achieving about 27 dB PSNR on average by
displaying the BL frames. There is no interruption throughout the display since all BL
frames are received correctly.
Figure 2.11: Channels allocation for sending Star Wars I to all SUs in CSI-based architecture.

(a) Received by SU\textsubscript{1}.

(b) Received by SU\textsubscript{2}.
Figure 2.12: PSNR of Star Wars I sequence as received by SU₁, SU₂, and SU₃, respectively, in CSI-based architecture.

2.8.1.2. Single-layer video against SVC performance. In this part of the experiments, we investigate the behaviour of the proposed streaming architecture when different scalable video sequences are requested by the different SUs in the CR network. Then, we compared the quality performance of the perceived videos against the corresponding single-layer video sequences. Unlike the scalable video sequences, single-layer video sequences are encoded into one BL only with no ELs. Same transmission parameters are considered in this test (i.e. same transition matrix and steady state probability).

The number of the channels allocated to each SU at the beginning of each time slot is shown in Figure 2.13. SU₁ is not allocated more channels after \( t = 2 \) seconds, since its video frames are scheduled on the available slots on the first two seconds. Sony Demo video has the smallest frame sizes compared to the other two scalable videos ‘Star Wars I’ and ‘Star Wars II’. Therefore, the BS continues allocating SU₂ and SU₃ to schedule their complete set of video frames.

Figures 2.14(a), 2.14(b) and 2.14(c) show the resultant PSNR at a target BER \( 10^{-5} \), when the Sony Demo, Star Wars I and Star Wars II video sequences are transmitted to SU₁, SU₂ and SU₃, respectively. Sony Demo video received by SU₁ is shown in Figure 2.14(a). Clearly, the PSNR quality of the single-layer video achieved
throughout the playback is about 35 dB on average with only one frame received with error. While the corresponding scalable version achieved a maximum quality of about 59 dB most often, with slight degradation in the quality to about 47 dB due to one EL2 of an I-frame received in error, causing the error to propagate to all its dependent consecutive frames from the same layer. The quality of about 47 dB is achieved by displaying the corresponding EL1 frames in this period. The slight degradation in the quality at the beginning of the display is due to 16 frames discarded from EL2 by the transmitter. One BL frame was received in error causing a single moment of interruption. The scalable video received by SU2 is shown in Figure 2.14(b), the quality fluctuates between the maximum of about 47 dB achieved by displaying the EL2 of the corresponding frames and the average PSNR quality of about 38 dB at some intervals during the display presented by EL1 frames since 162 frames are discarded from EL2 by the streaming algorithm. The quality drops to the minimum of about 27 dB achieved by the BL frames for short period of time at the beginning of the first and second half of the display because of 37 frames discarded from both EL1 and EL2. One BL frame was corrupted during the transmission causing one moment of interruption. It’s clear that SU2 has much poor PSNR compared to SU1, due to the large frame sizes of SU2’s high motion video that require large bandwidth allocation. ‘Star Wars I’ single-layer video received by SU2 shows PSNR quality performance with vast difference compared to the quality achieved by the scalable video throughout the display. No interruption in the display for both videos since no BL frame was received in error.

Star Wars II video sequence delivered to SU3 shown in Figure 2.14(c), achieved the maximum PSNR quality of about 50 dB throughout the first half of the display as a result of scheduling all frames up to EL2, with small quality regression at the video start-up due to number of frames discarded from EL2. In the second half of the display, the video quality drops to an average value of about 32 dB and the reason for that degradation is the frames discarded from EL1 (2 frames) and EL2 (43 frames) in addition to 6 frames received in error from EL1 as well as 5 frames from EL2. The errors in EL1 and EL2 frames are propagated to their dependent frames resulting in the minimum acceptable quality showed up by the BL during these intervals. Two frames from the BL were corrupted resulting in a discontinuity as shown in the plot. The corresponding single-layer video sequence achieves an average PSNR quality of about 34 dB during
the display, which is lower than the minimum quality achieved by its corresponding scalable video even when only the BL frames are presented.

Figure 2.13: Channels allocation for sending different videos at target BER $10^{-5}$.

(a) Received by SU$_1$. 
Figure 2.14: PSNR of the different video sequences as received by SU₁, SU₂ and SU₃, respectively, with target BER $10^{-5}$.

Next, we perform the same comparison between the single-layer video sequences against the scalable ones when the target BER is set to higher value of $10^{-3}$. The received video sequences by the three SUs in the CR network are shown in Figures 2.15(a), 2.15(b) and 2.15(c) respectively. Figures 2.15(a) shows that, the video received by SU₁ has discontinuities at different instants of the display since 8 frames from the BL were corrupted during the transmission. Moreover, the PSNR quality drops from 60 dB at the display start-up to an average of about 38 dB at some intervals during the
display due to the frames received in errors from EL₁ (13 frames) and EL₂ (6 frames), with some of which are reference frames causing the errors to propagate to other consecutive dependent frames. The single-layer video ‘Sony Demo’ achieves an average quality of about 38 dB during the playback similar to that attained by the BL frames of the scalable video counterpart. Discontinuities are also introduced at different moments of the display by missing 10 BL frames. We notice that more frames received in error comparing to the case when the target BER was set to $10^{-5}$.

The scalable video received by SU₂ shows better quality performance in Figure 2.15(b), fluctuating between the maximum PSNR quality of about 48 dB by displaying the EL₂ frames, and sometimes degraded slightly to about 37 dB achieved by the correctly received EL₁ frames. While their corresponding EL₂ frames (96 frames) are discarded by the transmitter or lost in error at the receiver. The video starts with low quality of about 28 dB as 17 frames are discarded by the streaming algorithm from both EL₁ and EL₂. ‘Star Wars II’ single-layer video achieves 26 dB PSNR throughout the display with discontinuities at the end of the show caused by 8 BL frames in error. ‘Star Wars II’ scalable video delivered to SU₃ is shown in Figure 2.15(c). The maximum PSNR quality is attained at most times of the display with slight degradation in the quality for short periods during the display as 37 frames are missed from the EL₂. One BL frame is received in error causing a single instant of interruption in the playback. The corresponding single-layer video sequence is received correctly, but again its quality performance that is too much less than the scalable counterpart.

The channels allocation decisions set by the BS at the beginning of each time slot are shown in Figure 2.16. The streaming algorithm completed scheduling the video frames of the SU₁ after 3 seconds, as this video ‘Sony Demo’ has small size frames compared to the other two videos. And therefore, the available channels are allocated to SU₂ and SU₃ only till the seventh second, when SU₃ received its complete video sequence. While the available channel detected at the eighth second is allocated to SU₂, whose video frames have the largest sizes compared to all other scalable video sequences in the CR streaming system.
Figure 2.15: PSNR of the different video sequences as received by SU₁, SU₂ and SU₃, respectively, with target BER $10^{-3}$. 
2.8.2. BCSI-based VoD architecture. In this section, we evaluate the performance of the on-demand video streaming architecture proposed in Section 2.7.2, for different streaming mechanisms. Furthermore, we compare the performance of the allocation algorithm that jointly considers the buffer occupancy and CSI, against the allocation algorithm that is based only on the reported SNRs or the occupancy of the SUs. We used the H.264/SVC CGS encoded sequences, namely, Star Wars I (high motion), Star Wars II (moderate motion) and Star Wars III (low motion). The PUs’ channels are assumed to be Rayleigh flat-fading channels with AWGN and exponentially distributed SNRs. For BL frames, the target BER $\psi$ was set to $10^{-4}$, while for EL1 and EL2 frames was set to $10^{-3}$. The modulation scheme that achieves the closest BER to $\psi$ is selected from BPSK, 4-QAM, 16-QAM, 64-QAM and 256-QAM, and used to send the video frame on the allocated channel(s). Without loss of generality, we assume the video playback starts after $T_{\text{start}} = 0.5$ seconds for all SUs, to ensure a fair comparison between the applied streaming algorithms.

The base station is updated with the information from the SUs on a reverse reliable channel every time slot duration $T_{\text{slot}} = 1$ seconds. We assume that the PUs’ channels are independent with non-identical availability behaviour following different transition matrices $P_n, n = 1, 2, ..., 10$, and steady-state probabilities $\pi_{0,n}, n = 1, 2, ..., 10$, in the range $(0.2 – 0.5)$ which indicates a busy to moderate availability behaviour. Without loss of generality, Table 2.2 shows sample of the transition and

Figure 2.16: Channels allocation for sending different videos at target BER $10^{-3}$. 

![Channels allocation for sending different videos at target BER $10^{-3}$]
steady-state probabilities for the ten PUs’ channels. The buffer occupancy threshold used by the channel allocation algorithm was set to $\Delta_{th} = 25$ to minimize the buffer starvation events for the SUs when no channels are available for allocation. The results are averaged over 100 runs in simulations. The available channels $N_{idle}$ at different instants are shown in Figure 2.17.

2.8.2.1. **Performance evaluation of video streaming mechanisms.** Firstly, we compare the performance of the proposed streaming mechanisms for delivering the different types of the CGS scalable videos to the SUs. We start by sending the high motion video “Star Wars I” that is characterized by its large frame sizes to all SUs. Figures 2.18(a), 2.18(b) and 2.18(c), compare the average quality (in terms of PSNR) received by SU$_1$, SU$_2$ and SU$_3$ respectively, for the different algorithms.

### Table 2.2: The availability parameters for the primary channels in the network.

<table>
<thead>
<tr>
<th>Channel index (n)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transition matrix ($P_n$)</td>
<td>$[0.3 \ 0.7]$</td>
<td>$[0.4 \ 0.6]$</td>
<td>$[0.2 \ 0.8]$</td>
<td>$[0.25 \ 0.75]$</td>
<td>$[0.35 \ 0.65]$</td>
</tr>
<tr>
<td>Idle probability ($\pi_{0,n}$)</td>
<td>0.3636</td>
<td>0.2727</td>
<td>0.375</td>
<td>0.235</td>
<td>0.333</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Channel index (n)</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transition matrix ($P_n$)</td>
<td>$[0.33 \ 0.67]$</td>
<td>$[0.24 \ 0.76]$</td>
<td>$[0.28 \ 0.72]$</td>
<td>$[0.41 \ 0.59]$</td>
<td>$[0.23 \ 0.77]$</td>
</tr>
<tr>
<td>Idle probability ($\pi_{0,n}$)</td>
<td>0.264</td>
<td>0.38</td>
<td>0.352</td>
<td>0.325</td>
<td>0.545</td>
</tr>
</tbody>
</table>

Figure 2.17: Number of idle channels detected every $T_{slot}$ duration.
In the first three seconds when enough number of idle channels \((N_{idle})\) are available for allocation, the streaming approaches achieve very close average PSNR quality for that streamed video in the first quarter of the display for all SUs as shown in Figure 2.18. Otherwise, the one-GoP, 3/4-GoP and 1/2-GoP streaming algorithms provide good PSNR quality for all SUs with no discontinuity in the playback. For that received by SU\(_1\), on average, the quality starts with 45 dB then degraded slowly to about 36 dB and further to 30 dB at the end of the first half of the display, as a considerable number of frames from both ELs have been discarded by the algorithms when the delay deadline constraints were violated. The performance of the other mechanisms 1/4-GoP and frame by frame streaming for this video type achieved lower PSNR quality in comparison, as more ELs are discarded causing the quality to degrade more. Moreover, 1/4-GoP and frame based streamed videos have discontinuities when the average PSNR went to low values of 19 dB and 10 dB as shown in Figure 2.18(a). Same observations can be noticed on the videos received by SU\(_2\) and SU\(_3\) in Figures 2.18(b) and 2.18(c), respectively. The degradation in quality at different instants of this video when it was streamed using the one-GoP, 1/2-GoP, 3/4-GoP, 1/4-GoP and 1/16-GoP mechanisms was mainly caused by the EL\(_1\) and/or EL\(_2\) frames that are either discarded by the algorithm or corrupted as detailed in the statistics provided in Table 2.3, averaged over 100 runs. Figure 2.19 shows a summarized comparison between the different streaming approaches in terms of the received PSNR quality averaged over the total number of frames. For all the SUs, it is evident that the half-GoP achieved the highest performance followed by the one-GoP and 3/4-GoP which show close performances. The 1/4-GoP and frame based approaches on the other hand, have lower average PSNR qualities in comparison.

The second set of results have been obtained for streaming different scalable videos to the SUs in the CR network. In details, Star Wars III to SU\(_1\), Star Wars I to SU\(_2\) and Star Wars II to SU\(_3\). Obviously, the one-GoP, half-GoP and 3/4-GoP streaming approaches achieved comparable performances as shown in Figure 2.20. For the video received by SU\(_2\) shown in Figure 2.20(b), the three approaches exhibit close performances with slight superiority to the one-GoP and half-GoP based algorithms. However, the videos received by SU\(_1\) and SU\(_3\), manifest the advantage of those algorithms over the 1/4-GoP and frame based streaming algorithms. As the quality degraded to lower values at different instants for the video received by SU\(_1\) as shown
Figure 2.18: Average PSNR of Star Wars I sequence as received by SU₁, SU₂ and SU₃, respectively, for different streaming techniques.
in Figure 2.20(a). While for that received by SU_3 “Star Wars II”, the quality went to low levels of about 30 dB at the end of the display as Figure 2.20(c) shows. The degradation in quality at different instants of those videos when they were streamed using the one-GoP, 1/2-GoP, 3/4-GoP, 1/4-GoP and 1/16-GoP mechanisms, was mainly due to the EL_1 and EL_2 frames that are either discarded by the algorithm or received in errors which propagated to their dependent frames from all layers. Table 2.4 shows this statistics in details. We notice that, large number of frames were discarded from EL_1 and EL_2 of SU_2’ video “Star Wars I”, compared with the frames discarded from the same layers for the other two videos because of its large frames sizes.

![Figure 2.19: Average overall PSNR of Star Wars I sequence as received by SU_1, SU_2 and SU_3, respectively, for different streaming techniques.](image)

<table>
<thead>
<tr>
<th>Streaming Algorithm</th>
<th>SU_1</th>
<th>SU_2</th>
<th>SU_3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Discarded</td>
<td>In errors</td>
<td>Discarded</td>
</tr>
<tr>
<td>EL_1</td>
<td>18</td>
<td>168</td>
<td>20</td>
</tr>
<tr>
<td>EL_2</td>
<td>164</td>
<td>3</td>
<td>150</td>
</tr>
<tr>
<td>BL</td>
<td>28</td>
<td>180</td>
<td>3</td>
</tr>
<tr>
<td>EL_1</td>
<td>37</td>
<td>177</td>
<td>3</td>
</tr>
<tr>
<td>EL_2</td>
<td>53</td>
<td>175</td>
<td>2</td>
</tr>
<tr>
<td>BL</td>
<td>2</td>
<td>9</td>
<td>1</td>
</tr>
</tbody>
</table>

When different video sequences with different characteristics (i.e. Star Wars III is a low motion with small frame sizes, Star Wars I is a high motion with large frame sizes and Star Wars II is a moderate motion with moderate frame sizes) were delivered to the SUs, we notice that as Figure 2.21 shows for all the videos received by the SUs,
Figure 2.20: Average PSNR of different video sequences as received by SU$_1$, SU$_2$ and SU$_3$, respectively, for different streaming techniques.
the streaming approaches show comparable performances with slight superiority for the half-GoP, one-GoP and 3/4-GoP streaming approaches over the other two approaches. Nevertheless, the 1/4-GoP and frame based approaches achieved a very close average PSNR quality because the received quality of the individual frames fluctuates between high and low values due to the discontinuities in the received video sequence (i.e. zero PSNR quality) in some trials of the 100 runs.

![Figure 2.21: Average overall PSNR of different video sequences as received by SU₁, SU₂ and SU₃, respectively, for different streaming techniques.](image)

Table 2.4: Statistics of Star Wars III, Star Wars I and Star Wars II received by SU₁, SU₂ and SU₃, respectively.

<table>
<thead>
<tr>
<th>Streaming Algorithm</th>
<th>SU₁ Discarded</th>
<th>SU₁ In errors</th>
<th>SU₂ Discarded</th>
<th>SU₂ In errors</th>
<th>SU₃ Discarded</th>
<th>SU₃ In errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EL₁</td>
<td>EL₂</td>
<td>BL</td>
<td>EL₁</td>
<td>EL₂</td>
<td>BL</td>
</tr>
<tr>
<td>One-GoP based</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>6</td>
<td>135</td>
<td>5</td>
</tr>
<tr>
<td>1/2-GoP based</td>
<td>2</td>
<td>13</td>
<td>7</td>
<td>6</td>
<td>132</td>
<td>3</td>
</tr>
<tr>
<td>3/4-GoP based</td>
<td>1</td>
<td>20</td>
<td>7</td>
<td>8</td>
<td>133</td>
<td>3</td>
</tr>
<tr>
<td>1/4-GoP based</td>
<td>3</td>
<td>11</td>
<td>8</td>
<td>17</td>
<td>131</td>
<td>4</td>
</tr>
<tr>
<td>Frame based</td>
<td>3</td>
<td>12</td>
<td>7</td>
<td>24</td>
<td>131</td>
<td>3</td>
</tr>
</tbody>
</table>

From the previous two set of results we can conclude that, the one-GoP and 3/4-GoP based streaming mechanisms achieve high PSNR quality performance for all SUs, however, further improvements in the PSNR quality of the received video streams can be attained by implementing the streaming algorithm on half-GoP basis. The latter approach realizes the best harmonization between scheduling a suitable number of
frames from the BL, that guarantee the continuity of the playback, and including a set of frames from the ELs under the constraints stated by the streaming algorithm to improve the quality to higher levels. The other two mechanisms for streaming (i.e., 1/4-GoP and frame based), show lower PSNR quality performance in comparison, with too much fluctuation between high and low PSNR levels across the display. Furthermore, they exhibit too much interruptions in the video playback as a few number of BL frames are scheduled on the allocated channels followed by the ELs frames of the streaming unit under consideration. Those ELs frames are characterized by their large sizes and consequently they take longer transmission time than their corresponding BL frames. As a result, scheduling of such frames on small streaming units basis will cause the perceived quality to fluctuate from one frame display to another. Moreover the BL frames from the next streaming unit may miss their display deadline and hence resulted into discontinuities of the video display at the end SUs. Such implementation for the streaming algorithm will result into inefficient utilization of the available spectrum and resources and will not meet the real-time requirements of the end SUs.

2.8.2.2. Performance evaluation of channel allocation algorithms. As previously mentioned, the BS in our proposed streaming system allocates the available channels every $T_{\text{slot}} = 1$ seconds to the SUs using the proposed channel allocation algorithm that is aware of the channel conditions as well as the occupancies for all SUs in the CR network. Next, we compare the performance of this joint buffer and SNR based allocation against the allocation algorithms that are based only on the channel quality or the SUs’ occupancies. We used the two streaming approaches, half-GoP and one-GoP based, to deliver “Star Wars I” to all SUs for one scenario. Moreover, results are also presented when different videos were streamed out to all SUs in the network as a second scenario. Figures 2.22 and 2.23, compare the average PSNR performance for the three allocation algorithms using the one-GoP based streaming algorithm. Clearly, the buffer based allocation has the worst PSNR performance throughout the display. As based on the playback buffer status only, the SUs are allocated the available channels randomly regardless of the transmission quality the SUs will experience on them (i.e., good or bad condition). As a result, a considerable number of frames from the ELs were received in errors and discarded by the algorithm to satisfy the constraints, as indicated in Table 2.5 and Table 2.6. The channel based allocation algorithm provides a close PSNR performance (on average) to the joint based allocation, as the
Figure 2.22: Average PSNR of Star Wars I sequence as received by SU₁, SU₂ and SU₃, respectively, for different allocation techniques using one-GoP based streaming.
Figure 2.23: Average PSNR of different video sequences as received by SU1, SU2 and SU3, respectively, for different allocation techniques using one-GoP based streaming.
algorithm allocates the available channels equally to the SUs on Round Robin basis with attentions to assign good quality channel(s) to those allocated users. However, its performance sometimes degrades too much for the videos received by some SUs compared to the other allocation approaches as shown in Figures 2.22(a), 2.22(b) and 2.22(c), when the quality fell back to 15 dB at the end of the display for both SU_1 and SU_2, while it went below 5 dB in the middle of the video display for SU_3. As long as the algorithm ignores the buffer status for the SUs, and their urgent needs to receive more BL frames to ensure the continuity of the playback when they are starving below the threshold, the available channels are allocated randomly between the SUs. As a result, some SUs will achieve good PSNR quality with no interruptions in their display, whilst the others are unsatisfied with the too much discontinuities in their video playback.

Figures 2.24 and 2.25, show the overall PSNR quality received by each SU averaged over the number of frames for the two previous scenarios. Clearly, the joint based allocation achieved better performance than the other allocation techniques, however, for some SUs the SNR based allocation sometimes achieves such high performance similar to the joint based allocation, when those users are allocated good quality channel(s) at the cost of interrupting the display for the starving SUs in the CR network when they are not allocated new channels. The buffer occupancy based allocation has the worst performance in comparison.

The two scenarios of video streaming have been implemented again with the half-GoP based streaming algorithm. Similar observations can be noticed evidently on Figures 2.26 and 2.27 for the PSNR received by each SU averaged over 100 runs, while Figures 2.28 and 2.29, show the overall PSNR quality averaged over the number of frames, they also indicate that the half-GoP based streaming approach outperforms its one-GoP based counterpart as it achieves higher average PSNR quality for all SUs in comparison, and that in essence for all the employed allocation algorithms. Tables 2.7 and 2.8 present a detailed statistics for comparison between the allocation algorithms in terms of the number of frames discarded and corrupted in each case for all SUs in the CR network.
Figure 2.24: Average PSNR of Star Wars I video sequence as received by SU₁, SU₂ and SU₃, respectively, for different allocation techniques using one-GoP based streaming.

Table 2.5: Comparison Statistics of different allocation algorithms when Star Wars I was delivered to all SUs on one-GoP basis.

<table>
<thead>
<tr>
<th>Allocation Algorithm</th>
<th>SU₁</th>
<th>SU₂</th>
<th>SU₃</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Discarded</td>
<td>In errors</td>
<td>Discarded</td>
</tr>
<tr>
<td>Buffer and SNR based</td>
<td>EL₁</td>
<td>17</td>
<td>169</td>
</tr>
<tr>
<td>Buffer based</td>
<td>EL₁</td>
<td>17</td>
<td>185</td>
</tr>
<tr>
<td>SNR based</td>
<td>EL₁</td>
<td>17</td>
<td>184</td>
</tr>
</tbody>
</table>

Figure 2.25: Average PSNR of different video sequences as received by SU₁, SU₂ and SU₃, respectively, for different allocation techniques using one-GoP based streaming.

Table 2.6: Comparison Statistics of different allocation algorithms when different videos delivered to all SUs on one-GoP basis.

<table>
<thead>
<tr>
<th>Allocation Algorithm</th>
<th>SU₁</th>
<th>SU₂</th>
<th>SU₃</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Discarded</td>
<td>In errors</td>
<td>Discarded</td>
</tr>
<tr>
<td>Buffer and SNR based</td>
<td>EL₁</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>Buffer based</td>
<td>EL₁</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>SNR based</td>
<td>EL₁</td>
<td>0</td>
<td>14</td>
</tr>
</tbody>
</table>
Figure 2.26: Average PSNR of Star Wars I sequence as received by SU₁, SU₂ and SU₃, respectively, for different allocation techniques using half-GoP based streaming.
Figure 2.27: Average PSNR of different video sequences as received by SU₁, SU₂ and SU₃, respectively, for different allocation techniques using half-GoP based streaming.
Figure 2.28: Average PSNR of Star Wars I video sequence as received by SU₁, SU₂ and SU₃, respectively, for different allocation techniques using half-GoP based streaming.

Table 2.7: Comparison Statistics of different allocation algorithms when Star Wars I was delivered to all SUs on half-GoP basis.

<table>
<thead>
<tr>
<th>Allocation Algorithm</th>
<th>SU₁ Discard</th>
<th>SU₂ Discard</th>
<th>SU₃ Discard</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EL₁</td>
<td>EL₂</td>
<td>BL</td>
</tr>
<tr>
<td>Buffer and SNR based</td>
<td>13</td>
<td>158</td>
<td>3</td>
</tr>
<tr>
<td>Buffer based</td>
<td>12</td>
<td>172</td>
<td>6</td>
</tr>
<tr>
<td>SNR based</td>
<td>13</td>
<td>169</td>
<td>5</td>
</tr>
</tbody>
</table>

Figure 2.29: Average PSNR of different video sequences as received by SU₁, SU₂ and SU₃, respectively, for different allocation techniques using half-GoP based streaming.

Table 2.8: Comparison Statistics of different allocation algorithms when different videos delivered to all SUs on half-GoP basis.

<table>
<thead>
<tr>
<th>Allocation Algorithm</th>
<th>SU₁ Discard</th>
<th>SU₂ Discard</th>
<th>SU₃ Discard</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EL₁</td>
<td>EL₂</td>
<td>BL</td>
</tr>
<tr>
<td>Buffer and SNR based</td>
<td>2</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>Buffer based</td>
<td>3</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>SNR based</td>
<td>2</td>
<td>12</td>
<td>2</td>
</tr>
</tbody>
</table>
Chapter 3. Real-time Video Streaming Over CRN Architecture

In this chapter, an occupancy-based channel-aware architecture is proposed for real-time streaming of scalable video sequences to multiple SUs in a CR network. We will present the implementation of this architecture along with the employed algorithms for real-time video streaming. Finally, the achieved results and analysis will be provided.


In this architecture, the BS allocates the available channels to SUs using Algorithm (4) previously presented in Chapter 2, that is aware of the buffer occupancy of the SUs as well as the quality status of the available channels to meet the real-time requirements of the SUs on the CRN. Moreover, two streaming algorithms are devised to schedule the scalable video frames on the allocated channel slots. The first algorithm controls the source rate of each encoded frame to satisfy the different constraints. While the second one, splits the video sequence into half-GoP basis and starts scheduling the BL frames first to guarantee the continuity of video playback and then schedules the ELs frames in such a way to satisfy the transmission and delay deadline constraints. Both algorithms use the CSI as fed-back by the SUs to adapt the modulation level on the allocated channels as previously presented by the adaptive modulation algorithm introduced in Section 2.4.

The complete streaming system at the BS will have the same architecture shown previously in Figure 2.10 without employing the transmission buffers. As in real-time video streaming the captured video frames are encoded on the fly and instantly scheduled for transmission. Therefore, the employed algorithms for real-time streaming have strict transmission and delay deadline constraints that differ from the ones employed for on-demand video streaming discussed previously in Chapter 2.

The video frames are encoded using the H.264/SVC encoder into one BL and multiple ELs at a nominal rate of $f_n = 30$ frames/second, then the target BER is set to a predefined value $\psi$ depending on the layer of each encoded frame (i.e. BL, EL$_1$ or EL$_2$). After that, the BS selects a certain level of modulation, $L \in \{\text{BPSK}, \ 4\text{-QAM}, \ 16\text{-QAM}, \ 64\text{-QAM}, \ 256\text{-QAM}\}$, that achieves the closest BER to the targeted one, and use it to send that frame on the allocated slots.
3.2. Real-time Streaming Algorithms

The streaming algorithms used for this real-time streaming architecture have to satisfy the video characteristics in dependency, priority and sensitivity, as well as satisfying the delay deadline constraints. Moreover, for real-time one way video communication, the video frames should be delivered within another delay constraint denoted by $T_d \approx 300 \text{ ms}$, which is an upper bound approximation of the total delay encountered throughout the transmission including, the transmitter buffer queuing time $T_{Q_{TX}}$, the propagation time $T_{prop}$ and the receiver buffer queuing time $T_{Q_{RX}}$ [44]. Therefore, the streaming algorithms should satisfy the delay deadline and transmission limit constraints when scheduling video frames for transmission to ensure the continuity of the video playback at the maximum possible quality.

3.2.1. Frame-based streaming algorithm. The first streaming algorithm is frame based, in which the streaming controller at the BS takes the encoded frames one by one including all coded layers, and implements a source rate control algorithm to decide on the number of encoded ELs to be transmitted along with the BL of each frame, based on the transmission and delay deadline constraints. To achieve the goal of minimizing the discontinuity in the playback, we define a parameter known as critical time $T_c$, which is the available time for the scheduled frame to be correctly received in the playback buffer. This parameter can be estimated based on the current buffer occupancy for the allocated user $\Delta_i$ as follows [32]

$$T_c = \begin{cases} \frac{\Delta_i - \Delta_{th}}{f_p} & \Delta_i > \Delta_{th} \\ \frac{1}{f_p} & \Delta_i = \Delta_{th} \\ \frac{1}{f_p (\Delta_{th} - \Delta_i)} & \Delta_i < \Delta_{th} \end{cases} \quad (3.1)$$

where $\Delta_{th}$ is a predefined buffer threshold and $f_p$ is the nominal playback rate. The frame based source rate control algorithm compares both the critical time $T_c$ and the transmission limit $T_d$ constraints against the estimated transmission time for the scheduled frame (i.e. $T_r = S_f/C \log_2(L)$), where $L$ is the assigned modulation level for the scheduled frame on the allocated channel and $S_f$ is the rate controlled frame size in bits. The pseudocode in Algorithm (6) summarizes the frame based source rate control
algorithm. The BS runs this algorithm at the beginning of the allocated slot(s) to decide on the frame sizes to be scheduled for transmission.

Algorithm 6: Frame based source rate control streaming algorithm.

Input: \( Ch_{si}, N_{frames}, N_{layers}, T_{slot}, f_p, \Delta_{th}, L, \) and \( \Delta_i, i = 1, 2, ..., M \).
Initialize: \( W = 0 \); // initializing the slot load counter for each allocated channel.
Initialize: \( i = 1 \); // initializing the frame index.
Set \( B_{max} = T_{slot} C \log_2(L) \); // the maximum bit budget of the slot.
While \( W < B_{max} \) and \( i \leq N_{frames} \) // compare the current load size against \( B_{max} \) do

Initialize: \( S_{f_i} = \sum_{j=0}^{N_{layers}} S_{f_{i,j}} \); // initialize the frame size up to the total number of the ELs.
Set \( T_r = S_{f_i} / C \log_2(L) \); // evaluate the corresponding transmission time.
Set \( k = N_{layers} \); // number of coded layers.
Determine \( T_c \) by comparing \( \Delta_i \) against \( \Delta_{th} \); // critical time parameter.
While \( T_r > (T_c \text{ and } T_d) \) // compare the transmission time against the constraints.
do

\( S_{f_i} \leftarrow S_{f_i} - S_{f_{i,k}} \); // reduce the frame size accordingly by discarding the \( k^{th} \) EL
Discard the dependent frames from the \( k^{th} \) EL;
\( k \leftarrow k - 1 \); // move to the next lower EL.
\( T_r = S_{f_i} / C \log_2(L) \); // a new evaluated transmission time to be compared against the constraints.
end

\( W \leftarrow W + S_{f_i} \); // schedule the frame.
\( i \leftarrow i + 1 \); // move to the next frame.
end

Move to the next allocated slot(s);

The summation index \( j \) indicates the coded layer (i.e. \( j = 0 \) for the BL, \( j = 1 \) for the EL\(_1\), ... etc.) and \( S_{f_{i,j}} \) is the size of layer \( j \) of frame \( i \) in bits.

Once the frame sizes have been decided by the BS according to the employed streaming algorithm, then they are scheduled for transmission on the allocated channels under the following constraint

\[
W = \sum_{i=1}^{V} S_{f_i} \leq B_{max}, \tag{3.2}
\]

where \( W \) is the total size of the scheduled frames in bits, \( V \) indicates the total number of scheduled frames and \( B_{max} \) is the maximum bit budget of the allocated channel given
by Equation (2.11).

### 3.2.2. Half-GoP based streaming algorithm.

For the second proposed streaming algorithm, the BS schedules the video frames from the different layers on half-GoP basis starting from the BL to ensure the continuity of the playback and then go through the ELs of the frames that belong to the half-GoP under considerations. Let $S_{f_{B,i}}$ be the size of the BL of frame $i$ in bits, when one or more channels are granted to a certain SU, then, the total number of bits that user will transmit from the BL of the first half-GoP is given by $W = \sum_{i=1}^{8} S_{f_{B,i}}$ (assuming that the full GoP size is 16 frames). The streaming system compares $W$ with $B_{max}$ of the allocated channel(s). If that sum is less than $B_{max}$, then all frames from the BL of the current half-GoP will be transmitted to guarantee the continuity of the video playback. Otherwise, the algorithm transmits the maximum possible number of BLs then waits for the following available slot(s) with a new bit budget. After all the BL frames of the currently transmitted half-GoP have been scheduled on the available slot(s), if there is still space for more transmission, then the algorithm assigns more of the ELs after checking their sizes against the unused portion of $B_{max}$, while taking into considerations the upper bound on the transmission limit $T_d$ and the delay deadline $\varphi_B$ constraints. If the EL frame satisfies that criteria, then it will be scheduled on the slot and the algorithm moves to the next ELs of next frame in the current half GoP. If a frame is discarded, then all frames that depends on it from the same layer as well as the higher layers will also be discarded. The algorithm moves to the next half-GoP after it finishes the first half-GoP and the process repeats until the end of the video sequence. The pseudocode in Algorithm (7) summarizes the half-GoP based streaming algorithm, the BS runs this algorithm at the beginning of each allocated slot to schedule a number of frames from different layers that provide the best received PSNR quality, while guaranteeing the continuity of the playback.

**Algorithm 7**: Half-GoP based streaming algorithm.

**Input**: $C_{si}$, $N_{frames}$, $N_{layers}$, $T_{slot}$, $L$, Frames’ sizes, dependency and display deadline; // The allocated channel(s) and video frames specifications for SU $i$.

**Output**: Number of frames to be scheduled from different layers.

**Initialize**: $W = 0$; // initializing the slot load counter for each allocated channel

**Set**: $B_{max} = T_{slot} C \log_2(L)$; // max bit budget of the slot

**while** $W \leq B_{max}$ // compare the current slot load size against the max bit budget of the slot

---

87
\[ W \leftarrow W + \sum_{i=1}^{8} S_{f_{B,i}}; \] // scheduling the BL frames of the half-GoP (assuming the full GoP has 16 frames).
\[ \text{Set } i = 1; \] // initializing the frame index in the half-GoP.
\[ \text{while } W \leq B_{\text{max}} \text{ and } i \leq 8 \] // scheduling the EL frames of the half-GoP
\[ \text{do} \]
\[ T_{r_{E,i}} = S_{f_{E,i}} / C_e; \] // the corresponding transmission time.
\[ \text{if } T_{r_{E,i}} \leq (\phi_{B,i} \text{ and } T_d) \] // satisfying the constraints.
\[ \text{then} \]
\[ W \leftarrow W + S_{f_{E,i}}; \] // schedule the EL frame
\[ \text{else} \]
Discard the frame and all its dependent frames from the same layer and higher ELs;
\[ \text{end} \]
\[ i \leftarrow i + 1; \] // move to the next EL frame.
\[ \text{end} \]
Move to the next half-GoP;
\[ \text{end} \]
Move to the next allocated slot(s);
\[ \text{until } \text{last video frame}; \]

### 3.3. Results and Analysis

In this section, we present the simulation results achieved for the implemented real-time video streaming architecture. The threshold used by the joint based channel allocation algorithm was set to \( \Delta_{th} = 25 \) frames to reduce the buffer starvation events for the SUs when they are not allocated new channels for transmission. And for the frame based streaming algorithm, by using simulation and trials for different values of occupancy, the threshold that identify the critical time \( T_c \) was set to \( \Delta_{th} = 10 \) frames as it achieves better performance. The scalable video sequences received by SU\(_1\), SU\(_2\) and SU\(_3\) are shown in Figure 3.1, for both streaming algorithms. Star Wars III received by SU\(_1\) is shown in Figure 3.1(a), the PSNR quality for the video delivered on frame based mechanism was relatively low (about 31 dB on average) at the beginning, with a moment of discontinuity as one BL frame missed its display time, then it jumped to the maximum of about 50 dB for a while before it degraded twice to about 37 dB on average. The degradations in quality at the different positions of this video are mainly due to 16 and 18 frames discarded from EL\(_1\) and EL\(_2\) respectively, along with 11 and 5 frames received in errors from the both layers EL\(_1\) and EL\(_2\), two moments of interruptions occurred as 1 BL frame missed its display time while another one was received in error. When the same video was delivered on half-GoP basis, it fluctuated...
between the maximum PSNR quality of about 48 dB and 39 dB when the display has just started, then it dropped to about 33 dB for short period before jumping to the maximum PSNR of 50 dB. The quality fluctuated for short periods throughout the display as a result of 6 and 15 frames discarded by the algorithm from EL_1 and EL_2 respectively. Moreover, 19 EL_1 frames were received in errors as well as 11 frames from EL_2. An event of starvation occurred in the buffer for a short period (≈ 0.1 seconds) as shown in Figure 3.2(a), when this user has not been allocated any channel at time \( t = 7 \) seconds. As SU_1 ’ occupancy was above the threshold at that moment, then the 3 idle channels are allocated to SU_2 and SU_3 on Round Robin manner as Figure 3.3(b) explained. As a result 2 BL frames missed their display deadline causing a short period of interruption.

Figure 3.1(b) shows the quality performance for Star Wars I scalable video received by SU_2. The video delivered on frame by frame basis, started the display with fluctuating the PSNR quality between 27 dB and about 38 dB for short period before it stabilized to about 37 dB on average (achieved by displaying the EL_1 frames), for a while and sometimes went down to 27 dB at the end of the display, when both ELs are discarded by the algorithm. The quality degradations throughout the display are caused by discarding 42 and 231 frames from EL_1 and EL_2 respectively, 3 BL frames missed their corresponding display times resulting into a discontinuity at the end of the display. On the other hand, Star Wars I delivered using the half-GoP streaming achieved better PSNR quality on average. The quality fluctuated between the maximum of 48 dB and 37 dB by playing the EL_1 frames when 190 frames discarded from the corresponding EL_2 with 2 frames in errors. Sometimes the quality degraded further to about 27 dB on average as a result of discarding 5 frames from EL_1 as well as 26 corrupted frames from the same layer with errors propagated to their dependent frames.

Star Wars II video received by SU_3 is shown in Figure 3.1(c). The scalable video delivered on half-GoP basis achieved the maximum PSNR quality of about 50 dB most of the display time. However, sometimes the discarded frames by the algorithm from the ELs (20 from EL_1 and 95 from EL_2), along with the 9 corrupted frames from EL_1 and 12 from EL_2 regressed the quality to a lower level of 32 dB on average. The display was interrupted for short period (≈ 0.2 seconds), as clearly indicated in Figure 3.2(c), when the buffer starvation occurred at time \( t = 3.8 \) seconds, causing moments of
interruption during the display. On the other hand, the video delivered using the frame based algorithm, started with about 42 dB PSNR, then it went up to the maximum of 50 dB for short period before it degraded slowly to 42 dB again. Figure 3.2(c) shows empty buffer at $t = 3.4$ seconds, as no channel was allocated to $SU_3$ at the 4th second as shown in Figure 3.3(c). Although, the occupancy was below the threshold for this user ($\approx 15$ frames), but there were only two idle channels that are allocated to $SU_1$ and $SU_2$ as they are starving more with ($\Delta_i = 13$) frames for both. And therefore, the display was interrupted for about 0.6 seconds before it is shown up again with 33 dB quality on average and then jumped to the maximum PSNR of 49 dB. A number of BL frames missed their deadlines causing the display to stop again for short period. The quality increased slowly to 30 dB and then to 39 dB before the display is disconnected as the buffer was starving again at $t = 8$ seconds as shown in Figure 3.3(c), with the corresponding BL frames arrived after their display deadline. The quality fluctuations throughout this video was caused by 96 and 142 discarded frames from $EL_1$ and $EL_2$ respectively, as well as 9 and 8 corrupted frames from the same layers.

Clearly we can notice that, sometimes when no channel(s) are available to be allocated to certain $SU(s)$, then the occupancy of their buffers starts starving at a certain time point. The half-GoP streaming algorithm has an edge over the frame streaming technique, as this starvation period (when occurred) continues for longer period when the frame based streaming algorithm is used to deliver the video sequence. And that is mainly due to the fact that the buffer has no enough BL frames to be played back during this time. While a set of BL frames is delivered before the ELs when the half-GoP based algorithm is employed, which helps in maintaining the continuity of the playback for the upcoming moments, and thus minimizing the buffer starvation period when this SU is not allocated new channel(s) by the BS for any reasons. To verify the previous achieved results, the simulated models for the proposed streaming system have been run over 100 times using different seeds in MATLAB and the corresponding PSNR qualities received by each SU were averaged out. Figure 3.4(a), 3.4(b) and 3.4(c) show this average PSNR for the videos received by $SU_1$, $SU_2$ and $SU_3$, respectively.
Star Wars III received by SU
Star Wars I received by SU
Star Wars II received by SU

Figure 3.1: PSNR of different video sequences as received by SU, SU, and SU, respectively.
Figure 3.2: Buffer occupancy status for SU₁, SU₂ and SU₃, respectively.
(a) Idle channels.

(b) Half-GoP based streaming.

(c) Frame based streaming.

Figure 3.3: Channels allocation for the SUs.
Figure 3.4: Average PSNR of different video sequences as received by SU₁, SU₂ and SU₃, respectively.
Chapter 4. Joint Adaptive Modulation and Channel Coding Scalable Video Streaming over CRNs

In this chapter, the proposed video streaming system has been implemented using a multi-level adaptive technique, which employs adaptive modulation and channel coding along with adaptive source rate control. The objective of this streaming scheme is to achieve better quality while maintaining a continuous video playback at the SUs end by reducing the buffer starvation events. Moreover, forward error control mechanisms are integrated with our proposed streaming system to facilitate correct reception of the frames by their deadlines.

4.1. The Proposed Architecture

Similar to the proposed architectures in Chapter 2, we assume that the encoded videos are pre-stored on a server located at the BS. We also assume that, there are $M$ SUs who can opportunistically access the CR spectrum which consists of $N$ channels that are licensed to $N$ PUs. This accessibility is valid as long as the channels are idle with no PU(s) activity. Additionally we assume that, the PUs have different activity patterns on the spectrum, but they cannot access the CR spectrum simultaneously with the SUs (i.e. overlay mode of transmission). Therefore, once a channel is declared idle by the BS, it will remain available until the end of a certain period of time denoted as $T_{slot}$ with no interference to the PUs.

The proposed scheme adapts the modulation level, the amount of channel coding and the number of retransmissions while benefiting from the scalability features of the video encoder to maintain the continuity of the video playback at the SUs end, while achieving an acceptable perceptual quality. To do so, the scheme opportunistically exploits the available idle channels and allocates these channels to the SUs based on their reported buffer occupancies using the channel allocation algorithm explained in Algorithm (2). The scheme also makes use of probabilistic-based streaming algorithm that controls the bit rate of each scheduled frame to schedule a number of ELs to be transmitted with the BL of a video frame while satisfying a lower bound on the probability of correctly receiving these layers by their deadlines as we will discuss later in this chapter.

As shown in Figure 4.1, the available channels detected by the BS every $T_{slot}$ duration are allocated to the SUs based on their feedback information, which includes
their buffer storages, $\Delta_i, i = 1, 2, ..., M$. Adaptive modulation and channel coding are employed on the allocated channels depending on their conditions to maximize the spectral efficiency and enable the representation of the video information at higher effective rates, and hence achieve better perceptual quality at the SUs end. The proposed streaming algorithm adaptively assigns the modulation level on the allocated channel(s) for each scheduled frame to achieve the highest correct reception probability, while satisfying a target probability threshold $\theta_{th}$. In particular, the BS assigns a certain level of modulation among (BPSK, 4-QAM, 16-QAM, 64-QAM and 256-QAM) on the allocated channels based on the CSI that is assumed to be changed per video frame transmission (i.e. slowly varying SNR). And furthermore, the selected modulation level should achieve the highest probability of correctly receiving a scheduled frame as we will detail later. The streaming controller at the BS employs the introduced source rate control streaming algorithm to schedule the video frames within the allocated slots to meet their deadlines, while achieving the highest possible quality with no interruptions in the playback. The buffer occupancy feedback of the SUs is used to determine the critical time constraint used by the streaming algorithm.

Figure 4.1: The joint adaptive VoD architecture at the BS.
In the transmitter side, we packetize each frame into number of packets that depends on the employed forward error correction (FEC) scheme for channel coding. Moreover, the transmitter is requested to reschedule the packets received in errors using the automatic repeat requests techniques (ARQ). These ARQ techniques are not typically employed in video transmission systems because of the strict end-to-end delay requested by the nature of video frames. However, we propose to employ these techniques in our system along with the FEC to improve the performance as long as we are able to meet the delay deadline constraints. More details about the error control mechanisms in general are introduced in the next section.

4.2. Error Control Mechanisms

Regardless of the design of any transmission system, errors are always introduced during transmissions over wireless channels that result in changing one or more bits in the received video frames. Therefore, video streaming systems employ different error control mechanisms to cope with data transmission errors such as: error resilient coding, error concealment, error correction codes also known as forward error correction (FEC) and automatic repeat request (ARQ) techniques [68].

4.2.1. Automatic repeat request (ARQ) techniques. This technique is basically used in the transport and data link control protocols and is based on the interaction between the transmitter and the receiver in the streaming system. According to the principle of this technique, the receiver detects the introduced errors in the received frames/packets using an employed error detection technique such as: parity check and cyclic redundancy check (CRC). Then, it updates the transmitter about the status of the received frame/packet, whether it is corrupted or not by sending positive or negative acknowledgment feedback on a reverse channel. Moreover, if the transmitter does not receive the feedback messages (i.e. positive acknowledgment (ACK) or negative acknowledgment (NACK)) within a certain amount of time known as “Round Trip Time (RTT)”, then the transmitter will retransmit that packet/frame again.

We will implement two different ARQ approaches in our streaming system, namely, Go-back-N (GBN) and Stop-and-Wait (SW) [29]. Figures 4.2 and 4.3 explain the mechanism for each ARQ approach. In the GBN approach, the transmitter keeps sending the video packets in a continuous manner, while the receiver discards the
packet if it is received in error and sends NACK to the transmitter. Moreover, the receiver keeps discarding all the next received packets after the corrupted one till that packet initially discarded is correctly received. As the same time, once the transmitter receives the NACK, all those discarded packets are rescheduled for transmission. The SW approach on the other hand, has a less complexity as the transmitter schedules the packet on the allocated channel and then waits for ACK from the receiver as shown in Figure 4.3. This approach seems to be less efficient than the GBN approach as the transmitter wastes the available idle time in CR resources while waiting for a feedback from the receiver side.

4.2.2. Forward error correction (FEC) techniques. Another error control mechanism is the FEC, which is basically a channel coding approach that adds redundancy bits to the original bitstream of the scheduled packet to form a relative structure. Therefore, when the original content of a received packet is changed during the transmission then that structure will change accordingly, and hence allows the
possibility of detecting or correcting that error. The FEC is designed to detect as well as to correct the introduced errors, thus avoiding the need for retransmission, unlike the ARQ techniques. There are several examples of such FEC techniques which could be either fixed or adaptive according to the channel(s) condition, such as: convolutional codes, Hamming codes and Reed-Solomon codes. Generally, the throughput could be improved using these FEC techniques, however, the introduced redundancy bits will result in transmission latency which could be minified by controlling the source rate to include the added FEC bits at the cost of delivering video sequences with lower quality [32].

4.3. System Model

As shown in Figure 4.4, the architecture of our proposed video streaming system over CR network consists in general of: video transmitter, PUs’ channels and video receivers at the SUs end. The raw video sequences are encoded at the transmitter side into one BL and multiple ELs using the H.264-SVC encoder. Then, the encoded frames from different layers are stored into transmission buffer for each SU.

![Joint adaptive modulation and channel coding video streaming architecture.](image)

The availability of the $N$ PUs channels at each time slot of fixed duration $T_{slot}$ follows the discrete-time two-state Markov chain model. We assume that the PUs’ channels follow a Rayleigh flat fading channel model, the condition of which does not
change during one frame transmission with slowly varying SNR over multiple successive frames. Adaptive modulation is employed on the allocated channels according to an integrated criteria that is addressed by a proposed probabilistic-based streaming algorithm to schedule the video frames on the allocated slots using several modulation levels: BPSK, QPSK, 16-QAM, 64-QAM and 256-QAM.

In addition, FEC convolutional codes are implemented for channel coding with either fixed or adaptive code rates. Assume for example, convolutional coding with a code rate \( R_c = k/n \), where \( k \) is the number of data or information bits, \( n \) is the total number of bits in the transmitted block, and \( n - k \) is the number of added FEC bits [69]. The coded BER will be related to the code rate of the FEC and the employed modulation level \( L \). Therefore, the BER for different modulation schemes used for transmission over channels with Rayleigh flat-fading and AWGN, can be obtained by substituting \( \gamma_c' = \gamma'/R_c \), (where \( \gamma' \) is the average value of the SNR per symbol received from SUs given by: \( \gamma' \triangleq \delta^2 kE_b/N_0 \)), in Equation (2.10) presented in Chapter 2 for no coding case [63, 65].

Packetization is employed to divide the entire content of each frame into a number of packets of known sizes as illustrated in Figure 4.5. The number of packets required to include an entire video frame can be estimated using

\[
N_p = \left\lceil \frac{S_f}{S_p - S_{FEC}} \right\rceil, \tag{4.1}
\]

where \( S_f \) is the complete frame size in bits, \( S_p \) is the known packet size in bits and \( S_{FEC} \) is the number of FEC bits attached to each packet.

![Figure 4.5: Dividing video frame into several packets with an overhead bits.](image-url)
The error control on the receiver side detects the errors in the received packets, and updates the transmitter with a feedback on a reverse reliable channel about the status of each received packet. Consequently, the transmitter employs the ARQ techniques to retransmit any scheduled packet a number of times until it is received successfully with a positive ACK from the receiver side as we previously mentioned in Section 4.2.

The number of retransmissions for a packet with errors until it is successfully delivered follows a geometric random variable $N_{rtx}$ with the following means for different ARQ approaches [68]. For the GBN approach

$$E[N_{rtx}] = \frac{Pr_c + N_{RTT}(1 - Pr_c)}{Pr_c},$$

(4.2)

And for the SW approach

$$E[N_{rtx}] = \frac{N_{RTT}}{Pr_c},$$

(4.3)

where $Pr_c$ is the probability of correctly receiving a packet and $N_{RTT}$ is the number of packets that can be delivered within the RTT.

Next, we will derive the probability of correctly receiving a packet, if the channel is available for the SU usage then

$$Pr_c = \sum_{m=0}^{b_{max}} \binom{S_p}{m} P_b^m (1 - P_b)^{S_p - m},$$

(4.4)

where $b_{max}$ is the maximum number of bits that can be corrected by the implemented FEC approach, $S_p$ is the decided packet size in bits and $P_b$ is the channel BER. The maximum number of corrected bits $b_{max}$ depends on the FEC scheme used for transmission. For example, for fixed FEC $b_{max}$ is defined with a fixed value (e.g. 50 bits), while for adaptive FEC, the value of $b_{max}$ can be estimated using Equation (4.5) according to the packet size $S_p$ and $P_b$ (i.e. the channel condition)

$$b_{max} \approx \left[ P_b S_p + 3 \sqrt{P_b S_p (1 - P_b)} \right].$$

(4.5)

If channel coding is applied then, the effective channel bit rate will be formulated as

$$C_e = Pr_c \frac{S_L}{S_p} C \log_2 L,$$

(4.6)
where $C$ is the error-free channel bit rate and $S_L$ is the size of the payload data excluding the FEC bits.

Next, we will formulate the transmission efficiency for the employed ARQ techniques. For the GBN ARQ protocol, the efficiency can be obtained as [68]

$$\xi_{GBN} = \frac{P_r}{P_r + N_{RTT}(1 - P_r)} S_L \log_2 L.$$  \hfill (4.7)

While this efficiency for the SW ARQ protocol is given by

$$\xi_{SW} = \frac{P_r}{N_{RTT} S_p} \log_2 L,$$  \hfill (4.8)

where $N_{RTT}$ is the number of packets that can be delivered within the round trip time interval $T_{RTT}$, and can be estimated using

$$N_{RTT} = \frac{T_{RTT} C \log_2 L}{S_p} + 1.$$  \hfill (4.9)

The transmission efficiency for the GBN and SW ARQ approaches for different modulation schemes (4-QAM, 16-QAM, 64-QAM and 256-QAM), under three different cases of channel coding: no implemented FEC, fixed FEC with $R_c = 3/4$ and for adaptive FEC are compared in [31]. The comparisons show that the GBN approach is more efficient than SW ARQ, as it achieved better transmission efficiency under all values of $E_s/N_o$. And furthermore, the performance of SW approach is affected too much when the RTT goes to high value.

To meet the objectives of our proposed streaming system of maintaining a continuous video playback at the SUs end and reducing the probability of SUs’ buffer starvation, the proposed source rate control streaming algorithm uses the buffer occupancy feedback information to define the critical time parameter $T_c$ given by Equation (3.1) that is previously introduced in Chapter 3. Moreover, an adaptive playback technique could also be implemented on the receiver side of the SUs. According to this technique, the playback rate $f_p$ is controlled in such a way to lessen the buffer starvation probability. The adapted playback rate $f_p$ is decided based on the current buffer occupancy $\Delta_i$ as well as the predefined buffer threshold $\Delta_{th}$ as follow

$$f_p = \begin{cases} 0.75 \, f_n, & \Delta_i \leq \Delta_{th} \\ f_n, & \Delta_i > \Delta_{th} \end{cases},$$  \hfill (4.10)
where $f_n$ denotes the nominal playback rate (encoding rate), $\Delta_{th}$ is set to a relatively large value to overcome the problem of the intermittent spectrum availability. However, that threshold should not keep large too much to avert effecting the end-to-end delay. To keep video playback variations not remarkable by the end users; the adaptive playback rate $f_p$ is kept within the range of $\pm 25\%$ of the nominal encoding rate [70].

Next we define a random variable $T_p$ which represents the transmission time required for successful packet delivery. The exponential distribution can be assumed to approximate this random variable with the following mean [68, 71]

$$E[T_p] = \frac{1}{\lambda} = \frac{S_L}{\xi C},$$  \hspace{1cm} (4.11)

where $\xi$ is the transmission efficiency previously introduced. Based on the ARQ approaches that we are using, the parameter $\lambda$ can be formulated as follows [68].

For the GBN approach

$$\lambda_{GBN} = \frac{S_p}{C \log_2 L} + \left( \frac{S_p}{C \log_2 L + T_{RTT}} \right) \frac{(1 - Pr_c)}{Pr_c}. \hspace{1cm} (4.12)$$

And for the SW approach

$$\lambda_{SW} = \frac{Pr_c}{\left( \frac{S_p}{C \log_2 L + T_{RTT}} \right)}. \hspace{1cm} (4.13)$$

Another random variable denoted as $T_f$ is introduced to denote the transmission time required for the entire frame successful delivery. And since the frame is a collection of several consecutive packets with exponentially distributed transmission time $T_p$ (i.e. for successful delivery), then $T_f$ can be approximated as m-Erlang distributed random variable with the shape parameter $N_p$ (positive integer) and the scale parameter $\lambda$ that depends on the ARQ approach used for retransmission. Therefore, the probability of receiving the complete frame successfully with no errors within the critical time $T_c$ specified in Equation (3.1) will be

$$Pr(T_f \leq T_c) = 1 - e^{-\lambda T_c} \sum_{k=0}^{N_p-1} \frac{(\lambda T_c)^k}{k!}. \hspace{1cm} (4.14)$$
Figure 4.6 shows the effect of the channel coding in terms of the maximum number of correctable bits $b_{max}$ on the probability of correctly receiving a frame within the critical time $T_c$ which we denoted as $Pr(T_f \leq T_c)$, for different modulation schemes assuming Rayleigh flat-fading channel. The plots compare that effect under two values of the error-free channel bit rate (i.e. $C = 256$ and 512 Kbps) for GBN and SW ARQ approaches. The plots were obtained for the following set of parameters: frame size $S_f = 8699$ bytes which is the average frame size of “NBC News” video sequence, packet size $S_p = 2272$ bytes which is the maximum transmission unit (MTU) according to IEEE 802.11 standard, average $E_s/N_o = 5$ dB, round trip time $T_{RTT} = 10$ ms and the critical time $T_c = 133 = 7/30$ ms assumed for the case of having 7 frames in the playback buffer of a SU with a playback rate $f_p = 30$ frames per second.

According to Figure 4.6, we can clearly notice that for a given modulation scheme, the probability is increasing as the number of correctable bits by the employed FEC is increasing (i.e. better performance) up to a certain value of $b_{max}$, after which the probability goes down. Although, the probability of correctly receiving a frame can be enhanced by the ability of the FEC technique to correct more bits in the received corrupted packets, which requires to add more FEC bits in the transmitted packets. Nevertheless, the corresponding payload size included in each scheduled packet will be small and the number of packets required to include the entire video frame is increasing correspondingly, resulting in an extra time required to deliver the entire video frame that violates the critical time. And that is mainly why the probability degraded after a certain value of $b_{max}$. For low SNR values which is the case we assumed (i.e. $E_s/N_o = 5$ dB), more FEC bits are included in the transmitted packets to overcome the errors introduced by increasing the modulation level. Figures 4.6(c) and 4.6(d), show the effect of the channel coding ($b_{max}$) on that probability for the two ARQ approaches GBN and SW, when the error free channel bit-rate was increased from 256 to 512 Kbps. Obviously we can notice that, the probability jumped to higher values (i.e. better performance) in comparison for the corresponding maximum number of correctable bits. Moreover, the probability started to degrade slightly after it reach its highest value. Similar to the previous case, the GBN approach performs better than the SW approach.
4.4. Joint Probabilistic-and-Occupancy based Streaming Algorithm

Each SU is allocated a number of available channel(s) by the BS according to the fed-back buffer occupancy information using the buffer based allocation algorithm given in Algorithm (2). Then, the BS starts to schedule the video frames for each SU on the slots of the allocated channel(s) using a joint probabilistic-and-occupancy based streaming algorithm that is explained as a pseudocode in Algorithm (8). According to this streaming algorithm, the BS decides on the size for each scheduled frame by comparing its transmission time $T_r$ against the critical time $T_c$ parameter given by Equation (3.1), which describes the updated status of SU’ buffer with an estimated delivery time. Additionally, the BS also computes the probability of correctly receiving a frame before its delay deadline, and compare this probability against a threshold $\theta_{th}$ that we set to a certain high value. If the frame size in the above computations satisfies the corresponding constraints, then the BS schedules that frame with the total number
of enhancement layers up to the $E_L_{N_{layers}}$, where $N_{layers}$ is the total number of ELs. Otherwise, the BS discards one EL and re-computes the transmission time and the correct reception probability, and compare them against the constraints and so on till all constraints are satisfied. At the end, the frame is scheduled with its BL to ensure the continuity of the playback, while the number of the ELs is depending on satisfying the constraints previously stated. Additionally, the frame may be scheduled with its BL only, if including any of the ELs results in violating the constraints, and that may consequently lead to interrupting the display. Moreover, the correct reception probability (CRP) is evaluated for all the modulation levels in the streaming system, then the BS employs the one which achieves the highest probability on the allocated channel(s).

Algorithm 8: Joint probabilistic-and-occupancy based streaming algorithm.

**Input:** $Ch_{si}$, $N_{frames}$, $N_{layers}$, $T_{prop}$, $T_{slot}$, $f_p$, $Δ_{th}$, ($Δ_i$, $i = 1, 2, ..., M$) and Frames’ sizes, dependency and display deadline; // Vector of allocated channel(s) and video frames specifications for SU$_i$.

**Output:** Number of coded layers for each frame to be scheduled.

**Initialize:** $W = 0$;

**Initialize:** $i = 1$; // initializing the frame index.

**Set** $B_{max} = T_{slot}C_e$; // maximum bit budget for the allocated channel.

**Set** $θ_{th} = x$ ; // correct reception probability threshold

**While** $W ≤ B_{max}$ and $i ≤ N_{frames}$ // compare the current load size against $B_{max}$.

**do**

**Initialize:** $S_{f_i} = \sum_{j=0}^{N_{layers}} S_{f_{i,j}}$; // include all enhancement layers in the frame size.

**Set** $k = N_{layers}$ ;

**Set** $T_{RTT} = 2T_{prop}$; // Round trip time

Determine $T_c$ by comparing $Δ_i$ against $Δ_{th}$; // critical time parameter in Equation (39).

**While** $T_r > T_c$ and $Pr(T_f ≤ T_c) < θ_{th}$ // sating the constraints.

**do**

$S_{f_i} ← S_{f_{i}} - S_{f_{i,k}}$; // Discard one enhancement layer (i.e. $E_L_k$).

$k ← k - 1$; // move to the next lower enhancement layer.

evaluate $Pr(T_f ≤ T_c)$ for each modulation level $L \in [2, 4, 16, 64, 256]$;

$Pr_m = \max (Pr_2, Pr_4, Pr_{16}, Pr_{64}, Pr_{256})$;

Select the modulation level that achieves the highest probability $L = m$;

$T_r = S_{f_i}/C \log_2(L)$; // a new transmission time to be compared

**end**

$W ← W + S_{f_i}$; // the updated slot load.

016
\[ i \leftarrow i + 1; \] // move to the next frame.

Move to the next allocated slot(s);

### 4.5. Results and Analysis

In this section we study the performance of the proposed joint adaptive streaming scheme. We assume that the BS transmits the MGS encoded video “NBC News” to the three SUs in the CR network. The trace file of the video is generated using JSVM (9.15) video encoder with high (level 2.2) and scalable high (level 3.1) encoding type. The video sequence consists of 256 total number of frames and is encoded into one BL and 6 MGS sub layers (i.e. ELs) with 352 x 288 resolution and playback rate of \( f_p = 30 \) frames per second. The GoP structure of the seven layers is IBBBBBBBBBBBBBBBB and is abbreviated as G16B15, with hierarchical dependency between the frames where B frames can be used as references to other B frames as shown in Figure 4.7. The PUs’ channels are assumed to be slowly varying Rayleigh flat-fading channels with AWGN and exponentially distributed SNR with an average of \( E_s/N_o = 18 \) dB. We used the GBN ARQ approach for its higher efficiency compared to SW ARQ approach as we previously discussed. We also employed adaptive modulation along with either adaptive or fixed FEC using different code rates \( (R_c = 1/3, 1/2, 2/3 \text{ and } 3/4) \), and compare their performances in terms of the received PSNR quality by each SU and the exhibited discontinuities in the playback. The error-free channel bit rate of each of the PUs’ channels is \( C = 256 \text{ Kbps} \). The video playback starts after \( \Delta_{\text{preroll}} = 15 \) frames are received correctly in the playback buffer.

![Figure 4.7: Frame dependencies in the G16B15 GoP structure.](image)

The BS is updated every \( T_{\text{slot}} = 0.5 \) second with information about the buffer occupancies of the SUs on a reverse reliable channel. We assume that the PUs’ channels are of independent and non-identical activities with different transition matrices \( P_n, n = \)
to change their availability state and with steady-state probabilities \( \pi_{0,n}, n = 1, 2, \ldots, 10 \) in the range (0.2 - 0.5), to investigate busy to moderate activity levels. The buffer occupancy threshold used by the channel allocation algorithm was set to \( \Delta_{th} = 25 \) frames to reduce the buffer starvation events for the SUs when no channels are available for allocation. Another occupancy threshold used to determine the critical time \( T_c \) in the streaming algorithm was set to \( \Delta_{th} = 10 \) frames, using simulations and trials for different values. Finally, the bound on the probability of correct reception \( \theta_{th} \) was set to 0.9. The available channels denoted by \( N_{idle} \) at different instants of time are shown in Figure 4.8.

First, we have implemented the proposed streaming system using fixed FEC with \( R_c = 1/3 \). Figure 4.9 shows the PSNR quality of the “NBC News” video sequence as received by each SU in the CR network. For the video received by SU\(_1\), the display started with the maximum PSNR of about 41 dB, then it fluctuated between 31 and 40 dB on average up to the show instant of frame indexed 120. The quality went up to the maximum again for the rest of the display before it degraded slowly from 38 dB to 32 dB at the display end. For SU\(_2\), the video started with the maximum PSNR quality of about 41 dB for a short period (= 0.7 seconds for displaying 20 frames), the PSNR quality then fluctuated between the maximum (of about 42 dB on average) and lower values throughout the display. The video received by SU\(_3\) exhibited quality performance similar to that received by SU\(_2\). In addition, the buffers for SU\(_2\) and SU\(_3\) were starving at times \( t \approx 3 \) seconds and \( t \approx 4.5 \) seconds, respectively, as shown in
Figure 4.10, which resulted in short periods of interruptions as 11 frames missed their display deadlines for SU\(_2\). While the discontinuities for SU\(_3\) were mainly caused by missing 17 frames for their display instants. The overall average PSNR qualities over the total number of frames (i.e. \(\bar{Q}\)) received by each SU in this case are: 38.5 dB, 35.6 dB and 34.4 dB, for SU\(_1\), SU\(_2\) and SU\(_3\), respectively.

![Figure 4.9: PSNR of NBC News sequence as received by SU\(_1\), SU\(_2\) and SU\(_3\), respectively, using fixed FEC with \(R_c = 1/3\).](image1)

![Figure 4.10: Buffer occupancies for SU\(_1\), SU\(_2\) and SU\(_3\), respectively, using fixed FEC with \(R_c = 1/3\).](image2)

We test the performance of the fixed FEC again with a higher code rate \(R_c = 1/2\). And the corresponding PSNR received by each SU are plotted in Figure 4.11. SU\(_1\) received the video with no discontinuity in the playback, only slight degradation in the quality at certain moments in the first half of the display (from the maximum of about 40 dB to lower levels of about 33 dB on average), as well as the display end when the PSNR quality dropped from 38 dB to about 32 dB. For the videos received by SU\(_2\) and SU\(_3\), lower levels of quality performance were achieved by both users compared to that
obtained by SU$_1$. And moreover, in addition to the slight degradation in the PSNR quality, they also experienced moments of interruptions when 4 and 2 frames missed their deadlines for SU$_2$ and SU$_3$, respectively, causing their buffers to underflow at $t \approx 3$ second for both as shown in Figure 4.12. The average PSNR qualities over the total number of frames received by each SU in this case are: 38.96 dB, 38.12 dB and 38.02 dB for SU$_1$, SU$_2$ and SU$_3$, respectively.

Figure 4.11: PSNR of NBC News sequence as received by SU$_1$, SU$_2$ and SU$_3$, respectively, using fixed FEC with $R_c = 1/2$.

Figure 4.12: Buffer occupancies for SU$_1$, SU$_2$ and SU$_3$, respectively, using fixed FEC with $R_c = 1/2$.

Figure 4.13 shows that, further improvement in the PSNR quality of the “NBC News” video sequence can be achieved by implementing fixed FEC with higher rate $R_c = 2/3$. Clearly we can notice that, the SUs received higher quality videos with less fluctuations throughout the display compared with the previous two cases as demonstrated by the average PSNR quality achieved by each SU: 39.86 dB, 39.9 dB and 38.96 dB for SU$_1$, SU$_2$ and SU$_3$, respectively. SU$_1$ and SU$_2$ have no interruptions in
their playback processes, whilst, the video received by SU$_3$ have a moment of discontinuity as 2 frames did not arrive by their display deadline causing an event of buffer starvation at time $t \approx 4.5$ seconds as shown in Figure 4.14.

Figure 4.13: PSNR of NBC News sequence as received by SU$_1$, SU$_2$ and SU$_3$, respectively, using fixed FEC with $R_c=2/3$.

Figure 4.14: Buffer occupancies for SU$_1$, SU$_2$ and SU$_3$, respectively, using fixed FEC with $R_c=2/3$.

A better PSNR quality performance was achieved when the code rate of the fixed FEC was increased further to $R_c = 3/4$ as shown in Figure 4.15. All SUs attained approximately the same levels of quality throughout the show. Their PSNR quality fluctuated between the maximum of about 42 dB and low values of about 32 dB on average with no interruptions in their display. Therefore, higher average PSNR qualities were achieved by each SU accordingly: 39.85 dB, 39.6 dB and 39.52 dB, for SU$_1$, SU$_2$ and SU$_3$, respectively.
Lastly, adaptive FEC was implemented in our proposed streaming system jointly with adaptive modulation, and the corresponding PSNR quality performance was compared against the cases of implemented fixed FEC with different code rates previously presented. Figure 4.17 shows the PSNR quality received by SU₁, SU₂ and SU₃, respectively. The SUs were able to achieve the maximum PSNR quality most of the display time with slight degradation in quality for short periods of time, when some ELs were discarded from the corresponding frames under show. No discontinuities were encountered during the display of the videos received by SU₁, SU₂ or SU₃. As a result, the SUs were able to obtain the highest averages of PSNR quality of about 40.58 dB, 39.9 dB and 40.47 dB, for SU₁, SU₂ and SU₃, respectively.
Figure 4.17: PSNR of NBC News sequence as received by SU₁, SU₂ and SU₃, respectively, using adaptive FEC.

Figure 4.18: Buffer occupancies for SU₁, SU₂ and SU₃, respectively, using adaptive FEC.

Figure 4.19 shows a comparison between the employed adaptive and fixed FEC with different code rates in terms of the average PSNR received by each SU. Evidently we can notice that, adaptive FEC outperforms the fixed FEC under different rates for all SUs, however, fixed FEC achieved a comparable performance when it was implemented with high code rate (i.e. $R_c = 3/4$). Other fixed FEC schemes have lower averages in comparison. Furthermore, we can observe that the average PSNR performance of the fixed FEC increases corresponding to increasing the code rates, and stabilizes for both $R_c = 2/3$ and $3/4$ (close performances). And that is mainly due to the overhead of increasing the payload data bits in the transmitted packets, while reducing the added FEC bits accordingly, which helps in delivering the entire frame with higher effective rate.
Figure 4.19: Average PSNR of NBC News sequence as received by SU₁, SU₂ and SU₃, respectively, for adaptive and fixed FEC with different code rates.

Fixing the code rate $R_c$ at low levels may guarantee the correct reception of the delivered coded frames regardless of any errors that may occur during their transmission over the wireless channels. However, at the same time it keeps the effective channel bit rate (at which we transmit) very low even if the transmission channels are in good condition, and that may cause discontinuities at the end SUs for their display, as the requested frames may arrive correctly, but after their display deadlines. On the other hand, coding the scheduled frames with higher rates may improve the effective bit rate but at the cost of receiving more frames with errors that may not be correctable with insufficient FEC bits attached to each frame. Subsequently, the corrupted frames become useless as they cannot be decoded and played-back. And hence resulted in degrading the received quality and interrupting the playback. Adaptive FEC compromises between high and low rated channel coding and improve the effective bit rate according to the dynamic channel conditions, and that consequently lead to an efficient exploitation of the transmission opportunities over the CR spectrum.
Chapter 5. Conclusion and Future Work

In this thesis, different systems for video streaming over cognitive radio networks were studied and implemented using SimEvents discrete-event simulator of MATLAB. It was assumed that a base station exploits the available opportunities in the primary users’ spectrum to deliver the video sessions to multiple secondary users. The proposed streaming schemes were tested under diverse primary users’ traffic patterns using a two-state discrete-time Markov chain model.

The proposed streaming schemes aim to ensure the continuity of the video playback at the secondary users end with acceptable perceptual quality level. To achieve such a goal, a scalable video coding technique is considered to adapt to the dynamic nature of the primary users’ channels. Scalable video coding allows for immediate codec reconfigurations and improves the quality of the received video by transmitting the enhancement layers of the frames, if certain constraints specified by the proposed streaming algorithms are satisfied. The proposed algorithms are characterised by their reduced complexity compared to other source rate control approaches, as they do not require any reconfiguration for the codec parameters and hence they can be used for real-time streaming.

The first streaming architecture uses adaptive modulation to transmit the video information to meet a fixed target bit error rate for video on-demand applications. One-GoP based streaming algorithm was used to schedule the frames from different coded layers on the channels that are allocated to secondary users based on their feedback channel state information. While the second implemented architecture employs adaptive modulation based on the sensitivity of the transmitted video information. A streaming algorithm that considers different GoP partitions was used to schedule the frames on the available channels that are assigned to secondary users based jointly on their buffer occupancies and channel state information. The results show that, scalable video sequences outperforms their single-layer counterparts in terms of the achieved PSNR video quality as validated for different target bit error rates. The results also reveal that, the channel allocation that jointly considers the buffer occupancies of the secondary users as well as the quality of their assigned channels outperforms other allocation algorithms that are based only on the buffer occupancies or the channel state information. It has also been shown that, the 1/2-GoP, 3/4-GoP and one-GoP based
streaming mechanisms achieve better PSNR quality with minimum interruptions in the playback compared with the $1/4$-GoP and frame based streaming mechanisms.

The third streaming architecture employs the joint buffer-and-CSI based channel allocation and layer-based adaptive modulation algorithms to provide a real-time video streaming service to the secondary users. The video frames were scheduled for transmission under strict constraints using two algorithms, half-GoP based and frame-based. The results show that, the half-GoP based streaming technique achieves better PSNR quality with less interruption in the playback compared with the classical frame based streaming mechanism that has longer periods of discontinuity with fluctuations in the received PSNR quality.

The last implemented architecture adapts the modulation level, the amount of channel coding and the number of retransmissions while benefiting from the scalability features of the video encoder to enhance the perceptual quality of the received videos at the end users while maintaining continuous video playback. A probabilistic-based streaming algorithm was used to control the source rate of the transmitted video information while satisfying a lower bound on the probability of correct reception. The results show that the proposed scheme ensures the improvement in the perceptual quality received by the secondary users with an uninterrupted playback.

As a future work, error concealment approaches could be integrated in our steaming systems to limit the error propagation through the frames, and hence facilitate the continuity of video playback at the secondary users end. Adaptive playback could also be employed on the video displays at the receiver side, to minimize the interruption periods when happened. On the other hand, the traffic patterns of the primary users on the cognitive radio spectrum can be assumed to be continuous according to an exponentially distributed activity time for the primary users, instead of being discrete. Additionally, we can consider primary users’ channels with fast varying SNR, on which the channel condition could change during one transmission slot or within the delivery time of the scheduled frame.
REFERENCES


VITA

Ala Eldin Omer was born in 1991, in Omdurman, Sudan. He received his primary and secondary education in Khartoum, Sudan. He was ranked as the third top student over whole Sudan ranking in the Sudanese High School Certificate (SHSC) with an average of (96.4%). He received his B.Sc. degree in Electrical and Electronic Engineering from the University of Khartoum, Sudan, in 2013 and graduated with First class as the second top student over whole the Engineering disciplines. From 2013 to 2014, he worked as a Research and Development Communication Engineer in the Telecommunication Research Center (TRC) in Khartoum, Sudan.

In September 2014, he joined the Electrical Engineering Master's program in the American University of Sharjah where he was awarded a two-year Graduate Assistantship from the Department of Electrical Engineering. During his Master's study, he co-authored 3 papers which were presented in international conferences. His research interests are in multimedia streaming, wireless communications, digital signal processing and mobile networks.