Video over Cognitive Radio Networks: A Cross-Layer Optimization Approach

by

Donglin Hu

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 7, 2012

Keywords: Cognitive Radio, Femtocell, Cooperative Relay, Scalable Video Coding, Successive Interference Cancellation, Interference Alignment

Approved by

Shiwen Mao, Chair, Associate Professor of Electrical and Computer Engineering
Prathima Agrawal, Ginn Distinguished Professor of Electrical and Computer Engineering
Chwan-Hwa Wu, Professor of Electrical and Computer Engineering
Ming Liao, Professor of College of Sciences and Mathematics
Tin-Yau Tam, Professor of College of Sciences and Mathematics

Abstract

Cognitive radios (CR) are intelligent radio devices that can sense the radio environment and adapt to changes in the radio environment. CR represents a new paradigm of wireless communications and networking by efficiently sharing spectrum between licensed users and secondary users. To harvest the high potential of CRs, the mainstream CR research has focused on developing effective spectrum sensing and access techniques. Although considerable advances have been achieved, the important problem of guaranteeing application performance has not been well studied.

The first part of this dissertation develops effective algorithms and protocols for spectrum sensing and access. First, we present a spectrum sensing error aware MAC protocol for a CR network collocated with multiple primary networks. Second, we consider the problem of interference mitigation via channel assignment and power allocation for CR users.

The second part of this dissertation focuses on the problem of optimized video streaming over CR networks. First, we tackle the problem of scalable video multicast in emerging infrastructure-based CR networks. Second, we investigate the more challenging problem of streaming multiple videos over multi-hop CR networks.

Cooperative CR networks are discussed in the third part of this dissertation. First, we investigate the problem of cooperative relay in CR networks for further enhanced network performance. Then, we study the problem of cooperative relay in CR networks for video streaming incorporating interference alignment techniques.

In the fourth part of this dissertation, we consider femtocell CR networks, where femto base stations (FBS) are deployed to greatly improve network coverage and capacity. First, we investigate the problem of generic data multicast in femtocell networks. Second, we tackle the problem of streaming scalable videos in femtocell CR networks.

This dissertation research provides a new perspective on how robust multi-user video streaming can be achieved in highly dynamic CR networks. It is among the first efforts to address the important area of video over CR networks, and offers systematic and comprehensive results and solutions. The findings may shed new light on the feasibility of CR networks in transporting real-time video and be useful for developing practical CR video systems.

Acknowledgments

First of all, I would like to express the deepest appreciation to my advisor and committee chair Prof. Shiwen Mao, who led me into this exciting area of wireless networking research and provided me with valuable support and assistance throughout my doctoral program. Without his enlightening guidance and persistent support this dissertation would not have been possible.

I also would like to thank the other members of my dissertation committee, Prof. Prathima Agrawal, Prof. Chwan-Hwa Wu, and Prof. Ming Liao, for their valuable comments and suggestions regarding my research work. I am also benefited very much from the courses taught by Prof. Jitendra K. Tugnait. I am very grateful to my Master's advisor Prof. Nedret Billor for her guidance in Mathematics and Statistics. Special thanks should be given to Prof. Tin-Yau Tam for his help in my work.

I highly appreciate the friendship from my friends at Auburn University, including Dr. Tong Liu, Dr. Fangyang Shen, Dr. In Keun Son, Yingsong Huang, Xi Yu, Jing Ning, Yu Wang, and many others. I thank them for the discussion, cooperation, and assistance during these years. In addition, I am very grateful to Prof. Mao's wife, Prof. Yihan Li for elaborate Thanksgiving food. I really enjoy the Thanksgiving parties at their house.

Finally, I show my all appreciation to my dear wife Jiahui Shen, my two adorable sons Lucas and Lewis, my parents, my parents-in-law, my uncle Yukang Hu and my cousin Hao Tan. Without their caring and support, the achieving of my dissertation is impossible.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Due to significant advances in wireless access technologies and the proliferation of wireless devices and applications, there is a fundamental change in wireless network traffic. As predicted by a Cisco study, wireless data is expected to grow to 6.3 Exabytes per month by 2015, a 26-fold increase over 2010, and 66% of the increase in future wireless data traffic will be video related [2]. Such dramatic increase in wireless video traffic is driven by the proliferation of mobile PCs, smartphones, tablets, etc., with 300 Million to 400 Million new mobile phone users adopting mobile services around the world and 120,000 new base stations (BS) added every year to meet the compelling need for ubiquitous access of mobile multimedia data.

Such fundamental changes in wireless data volume and composition bring about great challenges for the design and operation of wireless access networks. The capacity of existing and future wireless networks will be greatly stressed. Although allocating more spectrum may help, we are facing the problem of spectrum depletion since it is not a regenerable resource. Improving spectrum efficiency thus becomes ultimate important. In addition, Quality of Service (QoS)/Quality of Experience (QoE) provisioning in wireless networks also becomes a very important problem in order to enable high quality video services in legacy and emerging wireless networks. Since most of the increase in wireless video will be concentrated in the hot-spot areas, interference becomes the major limiting factor of network capacity and QoS provisioning. Effective interference exploitation and mitigation technologies are needed to achieve more efficient use of the spectrum and power resources.

Among various potential techniques, we consider *Cognitive radios* (CR) as an effective solution to meeting the critical demand in wireless network capacity and video provisioning. We

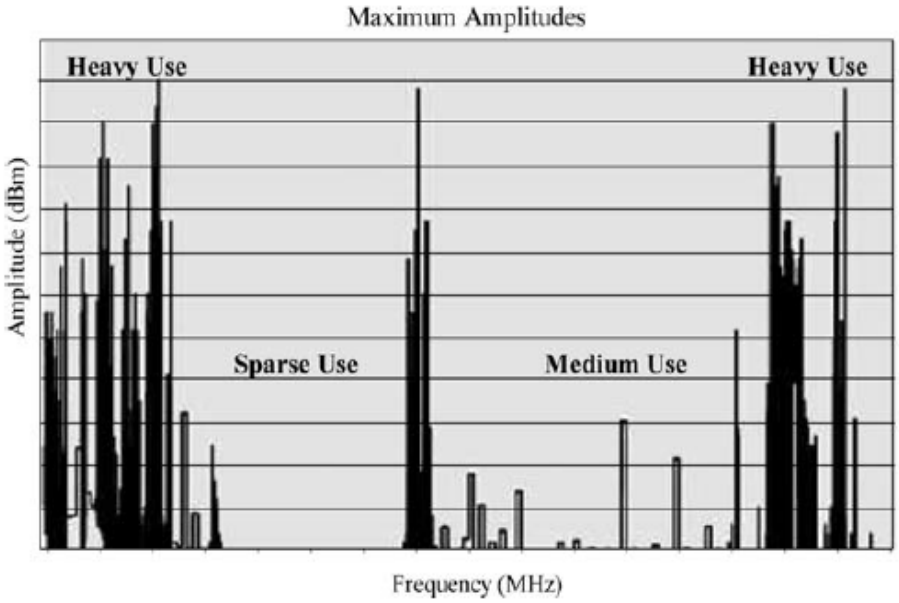


Figure 1.1: Spectrum measurement results [1].

also incorporate advanced wireless communications and networking technologies in the context of CR networks, such as cooperative communications, interference mitigation and alignment, and femtocells, for optimizing the quality of multiuser video communications.

A CR is an advanced device with interface to sense the radio environment, and dynamically access idle frequency bands so that CR networks are able to be deployed with existing primary network and improve network capacity. CR was motivated by a measurement study of FCC on the spectrum utilization, where many allocated spectrum are found to be seriously underutilized even in metropolitan areas (see Fig. 1.1). The CR concept represents a significant paradigm change in spectrum regulation and utilization, from exclusive use by licensed users (or, primary users) to sharing spectrum among licensed and unlicensed users (or, secondary or CR users). Although the basic concept of CR is intuitive, it is challenging to design effective CR medium access control (MAC) protocols to fully capitalize CR's potential. In order to exploit transmission opportunities in licensed channels, the tension between primary user protection and CR user spectrum access should be balanced. In addition, CR users access the licensed channels according to the sensing results. It is critical to take channel sensing errors into account in the design of CR MAC protocols.

Coupled with the depleting spectrum, interference will become the major capacity limiting factor. Thus, effective interference mitigation techniques are indispensable to realize the high potential of CRs. There are two types of interference that should be considered in such multi-channel environment. The first type, *co-channel interference* (CCI), is due to the coexisting transmitters occupying the same band as the victim receiver. The second type is *adjacent channel interference* (ACI), which is in the form of power leakage from adjacent channels. Both types of interference should be considered in the design of CR networking protocols.

Cooperative wireless relay represents another new paradigm for wireless communications and networking. It allows wireless relay nodes to assistant transmitters in data delivery. The objective of cooperative communication is to achieve cooperative diversity. Cooperation among wireless nodes enables opportunistic use of energy and bandwidth resources in wireless networks. Recently, researchers have been exploring the idea of combining these two techniques for enhanced network-wide performance [3, 4].

A femtocell is a small cellular base station (BS), typically used for serving approved users within a small coverage (e.g., a house). Femtocells usually have broadband wireline connections to the service provider network, which can be exploited to coordinate the transmissions of multiple femtocells for improved network-wide performance. Femtocells are shown effective in extending coverage, improving capacity, and reducing both power consumption and interference. Most of the benefits are achieved by the reduced distance of wireless transmissions, i.e., by bringing BS's closer to users [5].

Although considerable understandings have been gained on various aspects of CR. the problem of guaranteeing application performance has not been the focus of major CR research. To this end, we find spectrum-intensive and rate-adaptive video as a reference application, makes excellent use of the enhanced spectrum efficiency in CR networks. Unlike data, where each bit should be delivered, video is loss tolerant and rate adaptive. They are highly suited for CR networks, where the available bandwidth heavily depends on primary user behavior. We adopt scalable video coding, such as fine grained scalability (FGS) and medium grain scalable (MGS), to encode video streams.

We tackle the problems of video over various CR networking paradigms, such as infrastructure based CR networks, multi-hop CR ad hoc networks, cooperative relay based CR networks, and CR femtocell networks. We formulate cross-layer optimization problems that incorporates various system parameters and control knobs, and develop effective solution algorithms with proved optimal performance of tight performance bounds.

1.2 Major Contributions

The focus of this dissertation research is realtime video streaming in wireless networks, in particular, cognitive radio (CR) networks. The major contributions are summarized as follows.

We first work on a sensing error-aware MAC protocol that coordinates dynamic spectrum access for CR users, which considers channel sensing errors in the protocol design [6]. We develop analytical models to evaluate the performance of the proposed protocols. The accuracy of the analysis is demonstrated via our simulation study. In addition, interference mitigation in CR networks is crucial not only for primary user protection, but also the quality of service of CR user themselves. Therefore, we next consider the problem of interference mitigation via channel assignment and power allocation for CR users [7]. We propose both an RLT-based centralized algorithm and a distributed greedy algorithm which only needs local channel gain information. The distributed algorithm is shown to outperform the centralized algorithm and a heuristic algorithm with considerable gains in our simulations.

We further start to tackle the challenging problem of optimized real-time video multicast in an infrastructure-based CR network. The base station of the CR network exploits the spectrum opportunities in multiple licensed channels to multicast videos to groups of CR users [8]. A novel formulation of the CR video multicast system is developed, which considers important cross-layer design factors such as scalable video coding, video rate control, spectrum sensing, dynamic spectrum access, modulation, scheduling, and primary user protection. The design objective is to optimize CR video quality while protecting primary users from harmful collisions. Although the problem can be solved using advanced optimization techniques, we propose a sequential fixing

algorithm and a greedy algorithm with low complexity and proven optimality gap. We also investigate the more challenging problem of video streaming over multi-hop CR networks [9] which is formulated as a mixed integer nonlinear programming (MINLP) problem. We develop a centralized sequential fixing algorithm to derive upper and lower bounds for the achievable video quality, and then apply dual decomposition to develop a distributed algorithm with proven optimality and convergence conditions.

Next, we investigate the problem of cooperative relay in CR networks for further enhanced network performance [10]. In particular, we focus on the comparison of two representative cooperative relay strategies, *decode-and-forward* (DF) and *amplify-and-forward* (AF). Cross-point with the AF and DF curves are found when some parameter is valid, which indicates that each of them performs better in a certain parameter range and there is no case of dominance for the two strategies. We further extend our work to video streaming in cooperative CR networks [11], which incorporates interference alignment, a recent information theoretic breakthrough that allows current transmission of multiple signals. For the initial stochastic programming formulation, we first develop a reformulation that significantly reduces computational complexity. We then develop distributed optimal algorithms for the cases of a single channel and multi-channel with channel bounding, with proven convergence and convergence rate. We also developed a greedy algorithm for the multi-channel without channel bounding case, with bounded performance. This work is among the first efforts to harvest the information theoretic advances on interference alignment in the broader network context and practical perspective.

Femtocell networks are another theme that my dissertation focus on. We first investigate the problem of data multicast in femtocell networks that incorporates superposition coding and successive interference cancellation [12]. The objective is to minimize the total base station power consumption, while guaranteeing successful decoding of the multicast data at each user. We formulate a MINLP problem and reformulate it into a simpler form. To address the problem, we develop optimal and near-optimal algorithms for three typical connection scenarios, and derive upper and lower performance bounds. We also study the problem of streaming real-time scalable

videos in femtocell cognitive radio networks [13], with a multistage stochastic programming problem formulation. The proposed algorithms produce optimal solution in the case of non-interfering FBS's and near-optimal solution with proven lower bound in the case of interfering FBS's.

This dissertation research provides a new perspective on how robust multi-user video streaming can be achieved in highly dynamic CR networks. It is among the first efforts to address the important area of video over CR networks, and offers systematic and comprehensive results and solutions. The findings may shed new light on the feasibility of CR networks in transporting real-time video and be useful for developing practical CR video systems.

1.3 Dissertation Outline

The reminder of this dissertation is organized as follows.

In Chapter 2, we present our work on effective spectrum sensing and access protocols for CR networks. We introduce the general architecture of CR networks and present three protocol designs and analysis for different CR networking paradigms, including a sensing error aware CR MAC protocol and resource allocation for co-channel and adjacent channel interference mitigation in CR networks.

In Chapter 3, we investigate the problem of video streaming over three different CR network paradigms, including multiuser video multicast in the downlink of an infrastructure-based CR network and multiuser video unicast in a multi-hop CR network without any fixed network infrastructure.

In Chapter 4, we focus on cooperative CR networks. We compare and analyze two typical cooperative relay strategies in CR networks and investigate the problem of multiuser video streaming over a cooperative relay CR network by adopting interference alignment.

In Chapter 5, we consider femtocell networks. In particular, we tackle the problem of generic data mulicast in the downlink of femtocell networks that employs Superposition Codign (SC) and Successive Interference Cancellation (SIC), as well as resource allocation and quality of service (QoS) provisioning for multiuser video streaming in CR femtocell networks.

We conclude the dissertation and discuss our future work in Chapter 6.

Chapter 2

Cognitive Radio Networking

2.1 Introduction

According to Cisco's recent study, wireless data traffic is expected to increase by a factor of 66 times by 2013. Much of this future wireless data traffic will be video based services driven by the need for ubiquitous access to wireless multimedia content. Such drastic increase in traffic demand will significantly stress the capacity of future wireless networks.

Cognitive radios (CR) provide an effective solution to meeting this critical demand by exploiting co-deployed networks and aggregating underutilized spectrum for future wireless networks [1, 14, 15]. CR was motivated by the spectrum measurements by the FCC, where a significant amount of the assigned spectrum is found to remain underutilized. CR represents a paradigm change in spectrum regulation and access, from exclusive use by primary users to shared spectrum for secondary users, which can enhance spectrum utilization and achieve high throughput capacity.

Although the basic concept of CR is intuitive, it is challenging to design efficient cognitive network protocols to fully capitalize CR's potential. In order to exploit transmission opportunities in licensed bands, the tension between primary user protection and secondary user spectrum access should be judiciously balanced. Spectrum sensing and spectrum access are the two key CR functions. Important design factors include (i) how to identify transmission opportunities, (ii) how secondary users determine, among the licensed channels, which channel(s) and when to access for data transmission, and (iii) how to avoid harmful interference to primary users under the omnipresent of spectrum (or, channel) sensing errors. These are the problems that should be addressed in the medium access control (MAC) protocol design for CR networks. Although very good understandings on the availability process of licensed channels have been gained recently [16, 17], there

is still a critical need to develop analytical models that take channel sensing errors into account for guiding the design of CR MAC protocols.

To support many bandwidth-intensive applications in CR networks, it is desirable to achieve high network throughput under the constraint of limited interference to primary users. Due to the use of open space as transmission medium, wireless network capacity is usually constrained by interference. A CR user's transmission will generate interference not only to the neighboring primary users, but also to other CR users sharing the same or adjacent channels. Therefore, interference mitigation is crucial not only for primary user protection, but also for the quality of service of CR user themselves. Effective interference mitigation techniques are indispensable to realize the high potential of CRs.

The remainder of this chapter is organized as follows. The related work is discussed in Section 2.2. The system model and preliminary results are illustrated in Section 2.3. We present a sensing error aware MAC protocol in Section 2.4. Effective channel interference mitigation is discussed in Section 2.5. Section 2.6 concludes the chapter.

2.2 Background and Related Work

CR has been considered as a "spectrum agile radio" that enables dynamic spectrum access to exploit transmission opportunities in licensed spectrum bands [14, 15]. Several CR MAC protocols have been proposed in the literature. In [18], Le and Hossain propose a MAC protocol for opportunistic spectrum access in CR networks. A decentralized cognitive MAC protocol is developed in [19] that allows secondary users to explore spectrum opportunities without a central coordinator or a dedicated control channel. In a piece of recent work [20], Su and Zhang propose a negotiation-based sensing policy (NSP), in which a secondary user knows which channels are already sensed and will choose a different channel to sense. In [21], the authors consider two types of hardware constraints: sensing constraint and transmission constraint. In [22], based on the information obtained by a delegate secondary user, each secondary user group selects and switches to the best data channel for data communication during the next period. In [23], the authors describe a policy

such that a secondary user selects the channel that has the highest successful transmission probability to access. Many prior works [16, 18, 20, 21] assume perfect channel sensing, within which secondary users can always sense the channel correctly. Sensing errors are not considered. The joint design of opportunistic spectrum access and sensing policies is studied in a recent work [24] in the presence of sensing errors. The authors develop a separation principle that decouples the designs of sensing and access policy. This interesting study is based on a constrained partially observable Markov decision process (POMDP) formulation and thus has an exponentially growing computational complexity [24]. In [25, 26], the authors propose MAC protocols for multi-hop CR networks without the support of common control channels, but each CR user is requested to keep a list of available channels updated. The speed and accuracy of sensing process are high demanding.

Co-Channel Interefence (CCI) and Adjacent Channel Interference (ACI) are the two major factors limiting wireless network capacity. The impact of CCI on network performance is well-known and comprehensively investigated in [27]. Recently, the impact of ACI has attracted considerable interest in the wireless community. In [28, 29], the need was demonstrated for careful channel selection to mitigate ACI in IEEE 802.11 based systems. The impact of both CCI and ACI on network throughput and performance was evaluated in [30–32]. The interference models have been developed to analyze the channel interference in a few papers. In [33], the problem of statistical-physical modeling of CCI was investigated to analyze the outage probabilities in wireless networks and to design interference-aware transceivers. ACI was described by a simple quantification model that was verified by testbed experiments in [29]. In [34], a model for the aggregate ACI in TV white space was developed to demonstrate that the weighted sum of the total ACI power should be kept below certain threshold as well as ACI in each adjacent channel. A commonly used approach to reduce CCI is to assign different channels to neighboring transmitters [13, 35]. In [36, 37], frequency domain iterative multi-user detectors were adopted for CCI cancellation. A low-cost CCI avoidance MAC scheme was presented in [38]. In [39], ACI was minimized by optimizing reception of television receivers. In [40], Gidony and Kalet addressed the ACI mitigation problem by exploiting antenna diversity. In [41], statistical modeling of the aggregate interference

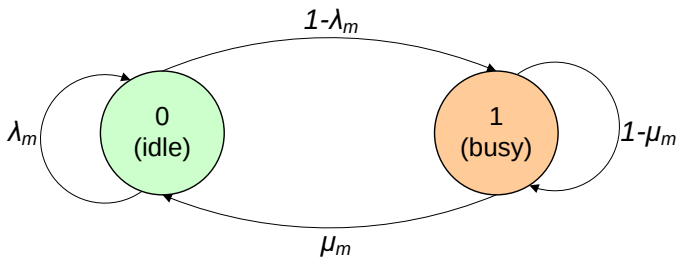


Figure 2.1: The discrete-time two-state Markov model for the state of channel m, S_m, for $m = 1, 2, \ldots, M$.

in a spectrum underlay CR network is proposed, in which CR networks coexist with primary networks and power constraints are imposed on the CR users to keep their power below the noise floor of primary receivers.

2.3 System Model and Preliminaries

2.3.1 Primary Network

We assume the primary users access the channels following a synchronous slot structure as in prior work [14, 20, 42]. The channel states are independent to each other and each of the M channels evolves over time following a discrete-time two-state Markov process, as shown in Fig. 2.1. Such channel model has been validated by recent measurement studies [14, 16, 20]. We define the *network state vector* in slot t as $\vec{S}(t) = [S_1(t), S_2(t), \ldots, S_M(t)]$, where $S_m(t)$ denotes the state of channel m, for $m = 1, 2, \cdots, M$. When channel m is idle, we have $S_m(t) = 0$; when channel m is busy, we have $S_m(t) = 1$.

Let λ_m and μ_m be the transition probability of remaining in state 0 and the transition probability from state 1 to 0 for channel m, respectively. Let $\eta_m = \Pr(S_m = 1)$ denote the *utilization* of channel m with respect to primary user transmissions. Let $\zeta_m = \Pr(S_m = 0)$ be the probability that channel m is idle (i.e., not being used by primary users). We then have

$$\eta_m = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} S_m(t) = \frac{1-\lambda_m}{1-\lambda_m+\mu_m} \tag{2.1}$$

$$\zeta_m = 1 - \Pr(S_m = 1) = \frac{\mu_m}{1-\lambda_m+\mu_m}. \tag{2.2}$$

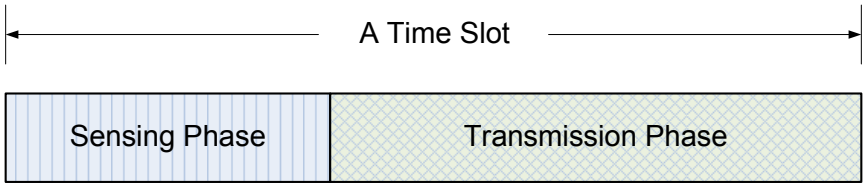


Figure 2.2: Time slot structure: a time slot consists of a sensing phase and a transmission phase.

2.3.2 CR Network

As presented in prior work [16, 20, 43], we assume that each secondary user is equipped with two transceivers: a *control transceiver* that operates over a dedicated control channel, which we assume is always available (e.g., a channel in the industrial, scientific and medical (ISM) band), and a *data transceiver* that is used for data communications through the M licensed channels. The data transceiver consists of an SDR that can be tuned to any of the M licensed channels to transmit and receive data. Secondary users also use their transceivers for spectrum sensing and exchanging sensing results.

We assume CR nodes access the licensed channels following the same time slot structure [14]. Each time slot is divided into two phases, the *sensing phase* and the *transmission phase*, as shown in Fig. 2.2. In the sensing phase, a CR node chooses one of the M channels to sense using one of its transceivers, and then exchanges sensed channel information with other CR nodes using the other transceiver over the control channel. During the transmission phase, the CR transmitter and/or relay transmit data frames on licensed channels that are believed to be idle based on sensing results, using one or both of the transceivers.

2.3.3 Spectrum Sensing

We explicitly consider channel sensing errors in the design. During the sensing process, two kinds of detection errors may occur. A *false alarm* refers to the case when an idle channel is considered busy. Consequently, the CR nodes will not attempt to access that channel and a spectrum opportunity will be wasted. A *miss detection* refers to the case when a busy channel is

considered idle. Since CR nodes will attempt to access this channel in the transmission phase, collisions with primary user transmissions will occur subsequently.

We adopt hypothesis test to detect the availability of channel m. The null hypothesis H_0^m is "channel m is idle." The alternative hypothesis H_1^m is "channel m is busy." Let ϵ_i^m and δ_i^m be the probabilities of false alarm and miss detection, respectively, when CR node i senses channel m. We have

$$\epsilon_i^m = \Pr\{\Theta_i^m = 1 | H_0^m\} \text{ and} \tag{2.3}$$

$$\delta_i^m = \Pr\{\Theta_i^m = 0 | H_1^m\}, \tag{2.4}$$

where $\Theta_i^m \in \{0, 1\}$ is the channel m sensing result of channel m at node i.

2.4 CR MAC Protocol

In this section, we present a channel sensing error aware MAC protocol for a CR network collocated with multiple primary networks. We assume primary users access the licensed channels following a synchronous time slot structure [14, 20]. The channel states are independent to each other and each evolves over time following a discrete-time Markov process [14, 16]. Secondary users use their software-defined radio (SDR)-based transceivers to tune to any of the licensed channels, to sense and estimate channel status and to access the channels when they are found (or, believed) to be available.

In particular, we develop two channel sensing polices, with which secondary users collaboratively sense the licensed channels and predict channel states. With the *memoryless sensing* policy, each secondary user chooses one of the M licensed channels to sense with equal probability. During the sensing phase, secondary users also exchange sensing results through a separate control channel. This sensing policy is further improved with a mechanism to spread out secondary users to sense different channels, therefore reducing the chance that a channel is not sensed by any of the users. When spreading out secondary users to the channels, the mechanism also considers

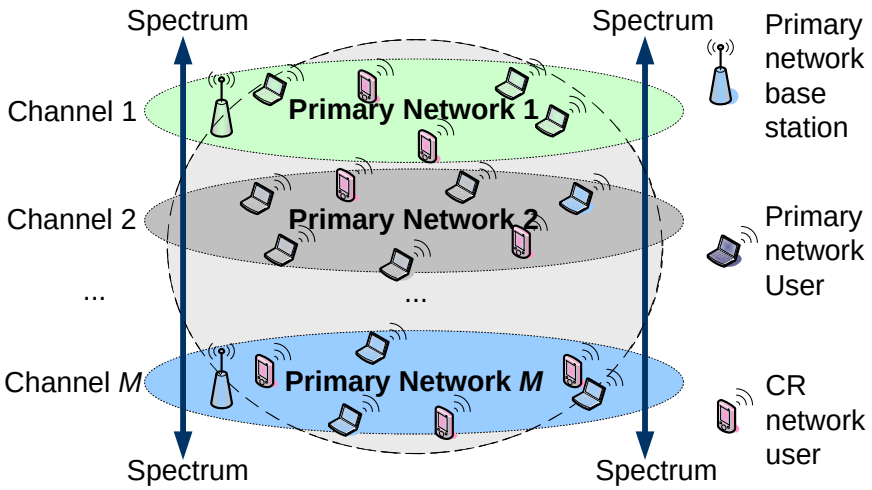


Figure 2.3: The CR secondary network is collocated with M primary networks.

the autocorrelation of channel processes to obtain more accurate sensing results. This is termed *improved sensing* policy.

These two sensing polices are then incorporated into the p-Persistent Carrier Sense Multiple Access (CSMA) mechanism to make sensing error aware CR MAC protocols. We analyze the proposed CR MAC protocols with respect to the interference and throughput performance and derive closed-form expressions. Primary user protection is achieved via tunning the channel access probability p of p-Persistent CSMA according to the interference analysis. The CR MACs also aims to maximize the CR network throughput while satisfying the primary user protection constraints. Through simulations, we find that the analysis is highly accurate as compared to simulation results. In addition, the proposed sensing error aware CR MAC protocols outperform two existing schemes with considerable gain margins, which justify the importance of considering channel sensing errors in CR MAC design.

2.4.1 Network Model and Assumptions

The network model considered in this section is illustrated in Fig. 2.3. Consider M primary networks, each allocated with a licensed channel. We assume the primary users access the channels following a synchronous slot structure as in prior work [14,20,42]. The channel state model evolves independently following a discrete Markov process (see Section 2.3.1).

We assume a secondary network collocated with the M primary networks, within which N secondary users take advantage of the spectrum white spaces in M licensed channels for data transmissions. For protection of primary users, the probability of collision caused by secondary user transmissions to primary users should be upper bounded by a prescribed threshold γ_m, for $m = 1, 2, \cdots, M$.

As illustrated in Section 2.3.2, we assume that each secondary user is equipped with two transceivers: a *control transceiver* that operates over a dedicated control channel, which we assume is always available, and a *data transceiver* that is used for data communications through the M licensed channels. The data transceiver consists of an SDR that can be tuned to any of the M licensed channels to transmit and receive data. Secondary users also use their transceivers for spectrum sensing and exchanging sensing results.

2.4.2 Sensing Error Aware CR MAC Protocol

For the CR network described in Section 2.4.1, we develop sensing aware MAC protocols for opportunistic spectrum access. The time slot structure of the proposed MAC protocols is shown in Fig. 2.4, which consists of a *sensing phase* and a *transmission phase*. The sensing phase is further divided into $\bar{K}$ mini-slots, within which each secondary user senses one of the licensed channels. CR users access the channels for data transmission during the transmission phase. Let T_s, T_{ms}, and T_{data} denote the duration of a time slot, a mini-slot, and the transmission phase, respectively (see Fig. 2.4), we have

$$T_s = \bar{K} \times T_{ms} + T_{data}. \tag{2.5}$$

We first discuss the two key components of the proposed protocols, i.e., channel sensing and channel access, and then analyze their performance with respect to primary user protection and the expected throughput.

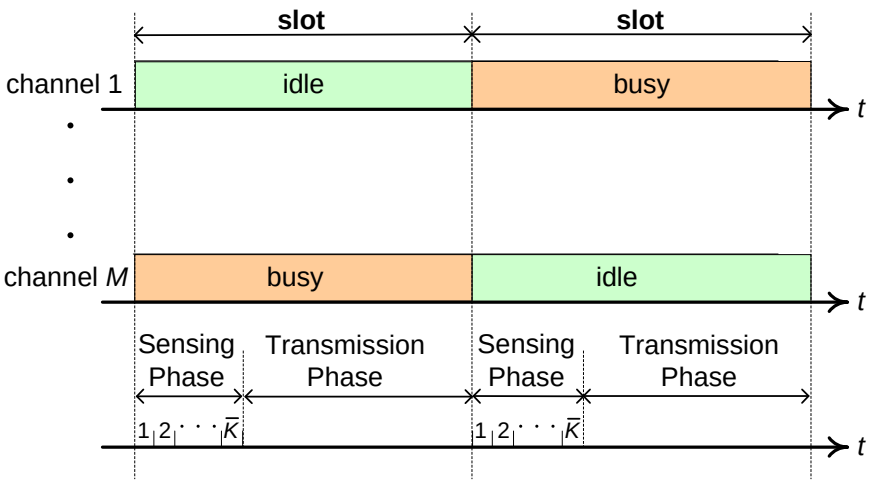


Figure 2.4: The time slot structure of the proposed sensing error aware CR MAC protocol.

Sensing Phase

The first key element of the proposed MAC protocols is spectrum, or channel sensing. Although precise and timely channel state information is highly desirable for opportunistic spectrum access and primary user protection, contiguous full-spectrum sensing is both energy inefficient and hardware demanding. Since we assume a secondary user is equipped with one transceiver for spectrum sensing, i.e., the data transceiver with SDR capability, only one of the licensed channels can be sensed by the secondary user at a time.

During the sensing phase (see Fig. 2.4), a secondary user picks a licensed channel and keeps on sensing it for one or multiple mini-slots. As discussed in Section 2.3.3, two kinds of detection errors may occur: false alarm and miss detection. We assume all secondary users have the same probability of detection errors when sensing channel m, $m = 1, 2, \cdots, M$. Let ϵ_m and δ_m denote the probabilities of false alarm and miss detection on channel m, respectively. The spectrum sensing performance can be represented by the Receiver Operation Characteristic (ROC) curve, where $(1 - \delta_m)$ is plotted as a function of ϵ_m [14]. For a specific channel m in a certain time slot

t, the sensing error probabilities can be written as:

$$\Pr(W_{m,i} = 1 \mid S_m = 0) = \epsilon_m, \text{for all } i = 1, 2, \cdots \tag{2.6}$$

$$\Pr(W_{m,i} = 0 \mid S_m = 1) = \delta_m, \text{for all } i = 1, 2, \cdots, \tag{2.7}$$

where $W_{m,i}$ is the ith sensing result of channel m and S_m is state of channel m.

We assume that the sensing results from different users are independent and the sensing results in different mini-slots are also independent to each other. Suppose a secondary user continues to sense channel m for k mini-slots and obtains k sensing results. The conditional probability that channel m is available after the kth sensing mini-slot, denoted by $a_{m,k}$, can be derived as

$$\begin{aligned}
a_{m,k} &= \Pr(S_m = 0 \mid W_{m,1} = \theta_{m,1}, \cdots, W_{m,k} = \theta_{m,k}) \\
&= \frac{\Pr(W_{m,i} = \theta_{m,i}, i = 1, \cdots, k | S_m = 0) \Pr(S_m = 0)}{\sum_{j=0}^{1} \Pr(W_{m,i} = \theta_{m,i}, i = 1, \cdots, k | S_m = j) \Pr(S_m = j)} \\
&= \frac{\Pr(S_m = 0) \prod_{i=1}^{k} \Pr(W_{m,i} = \theta_{m,i} | S_m = 0)}{\sum_{j=0}^{1} \Pr(S_m = j) \prod_{i=1}^{k} \Pr(W_{m,i} = \theta_{m,i} | S_m = j)} \\
&= \left[1 + \frac{\Pr(S_m = 1)}{\Pr(S_m = 0)} \prod_{i=1}^{k} \frac{\Pr(W_{m,i} = \theta_{m,i} | S_m = 1)}{\Pr(W_{m,i} = \theta_{m,i} | S_m = 0)}\right]^{-1} \\
&= \left[1 + \alpha_m^{d_m} \beta_m^{k-d_m} \frac{\Pr(S_m = 1)}{\Pr(S_m = 0)}\right]^{-1} \\
&= \left(1 + \alpha_m^{d_m} \beta_m^{k-d_m} \frac{\eta_m}{\zeta_m}\right)^{-1},
\end{aligned} \tag{2.8}$$

where d_m is the number of observations whose sensing result is 0 on channel m, and α_m and β_m are defined as follows.

$$\alpha_m = \frac{\Pr(W_{m,i} = 0 | S_m = 1)}{\Pr(W_{m,i} = 0 | S_m = 0)} = \frac{\delta_m}{1 - \epsilon_m}, \text{ for } \theta_{m,i} = 0 \tag{2.9}$$

$$\beta_m = \frac{\Pr(W_{m,i} = 1 | S_m = 1)}{\Pr(W_{m,i} = 1 | S_m = 0)} = \frac{1 - \delta_m}{\epsilon_m}, \text{ for } \theta_{m,i} = 1. \tag{2.10}$$

For the secondary user, it is also possible that it obtains some of the k sensing results by local measurements, and receives the remaining sensing results from the control channel in the case that some other secondary users are sensing the same channel m. By abuse of notation, we also use $a_{m,k}$ to denote the conditional channel availability probability in this case, due to independence of the sensing results. We plot $a_{m,k}$ as a function of k for the channel idle and busy cases in Fig. 2.5, using the same parameters as one of the simulations (see Section 2.4.4). We have the following proposition for $a_{m,k}$.

Definition 2.1. *A random variable X is said to be dominated by Y in the sense of stochastic ordering if $\Pr(X \geq x) \leq \Pr(Y \geq x)$ for all x*

Proposition 2.1. *When channel m is idle, $a_{m,k}$ is a monotone increasing function of k; when channel m is busy, $a_{m,k}$ is a monotone decreasing function of k in the sense of stochastic ordering.*

Proof. From the defintion of $a_{m,k}$ in (2.8), it follows that

$$\begin{aligned}
& \Pr(a_{m,k} \geq \theta_1) \\
= \; & \Pr\left(\left[1 + \frac{\eta_m}{\zeta_m}\left(\frac{\delta_m}{1-\epsilon_m}\right)^{\sum_{i=1}^{k}\bar{W}_{m,i}}\left(\frac{1-\delta_m}{\epsilon_m}\right)^{\sum_{i=1}^{k}W_{m,i}}\right]^{-1} \geq \theta_1\right) \\
= \; & \Pr\left(\left(\frac{\delta_m}{1-\epsilon_m}\right)^{\sum_{i=1}^{k}\bar{W}_{m,i}}\left(\frac{1-\delta_m}{\epsilon_m}\right)^{\sum_{i=1}^{k}W_{m,i}} \leq \left(\frac{1}{\theta_1} - 1\right)\frac{\zeta_m}{\eta_m}\right) \\
= \; & \Pr\left(\sum_{i=1}^{k}\left[W_{m,i}\log\left(\frac{1-\delta_m}{\epsilon_m}\right) - \bar{W}_{m,i}\log\left(\frac{1-\epsilon_m}{\delta_m}\right)\right] \leq \chi_m\right) \qquad (2.11)
\end{aligned}$$

where $\bar{W}_{m,i} = 1 - W_{m,i}$.

Since $\epsilon_m < 0.5$ and $\delta_m < 0.5$ for practical sensors, both $\log\left(\frac{1-\delta_m}{\epsilon_m}\right)$ and $\log\left(\frac{1-\epsilon_m}{\delta_m}\right)$ are positive. If $S_m(t) = 0$, we have that

$$\Pr(W_{m,i} = 1) < \Pr(W_{m,i} = 0) = \Pr(\bar{W}_{m,i} = 1).$$

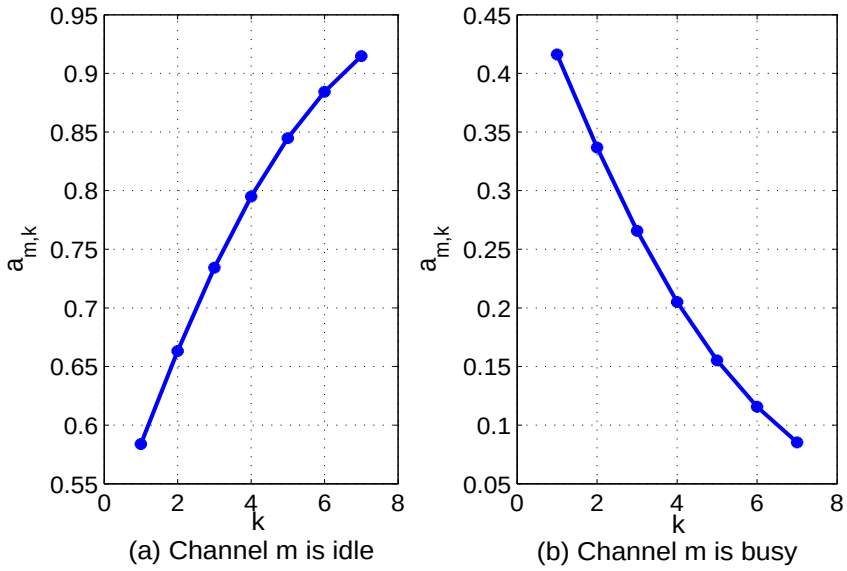


Figure 2.5: Illustration of $a_{m,k}$ as a monotone function of k, when $\epsilon_m = 0.3$, $\delta_m = 0.3$, and $\bar{K} = 7$.

It follows that $\Pr(a_{m,k} \geq \theta_1) < \Pr(a_{m,k+1} \geq \theta_1)$. That is, $a_{m,k}$ is a monotone increasing function of k in the sense of stochastic ordering. Similarly, we can show that

$$\Pr(a_{m,k} \leq \theta_0) < \Pr(a_{m,k+1} \leq \theta_0)$$

when $S_m(t) = 1$. That is, $a_{m,k}$ is a monotone decreasing function of k in the sense of stochastic ordering when the channel is busy. □

During the sensing phase, each secondary user chooses one channel to sense with equal probability at the beginning of the time slot. Secondary users also report their sensing results over the control channel, and share the corresponding channel sensing results during the mini-slots. Two threshold probabilities $\Theta_0 < \Theta_1$ are used for decision making.

- If the availability of channel m, i.e., $a_{m,k}$, is below Θ_0, the channel is believed to be busy and the secondary users will wait till the next time slot to start sensing again.
- If the availability of channel m is between Θ_0 and Θ_1, secondary users will keep on sensing the same channel to obtain more sensing results for more accurate estimation of the channel state, until the maximum number of mini-slots, $\bar{K}$, is reached.

- If the availability of channel m exceeds Θ_1, the channel is believed to be idle and the secondary users stop sensing and prepare to access the channel (see Section 2.4.2).

The *stop time* K_m when secondary users stop sensing channel m, is a random variable that takes value between 1 and $\bar{K}$, the maximum number of mini-slots that can be used for sensing (see Fig. 2.4). If we have $\Theta_0 < a_{m,k} < \Theta_1$ by the end of the sensing phase, then channel m state is not identified due to lack of time (or sensing results) and the channel will not be accessed.

When there are k sensing results available (e.g., one user senses channel m for k mini-slots, or it senses channel m for less than k mini-slots and receives some channel m sensing results from other secondary users), we define three sets of estimates for the state of channel m, as:

$$\Psi^0_{m,k} = \{d_m \mid a_{m,k} \leq \Theta_0, \forall\ 0 \leq d_m \leq k\} \tag{2.12}$$

$$\Psi^1_{m,k} = \{d_m \mid a_{m,k} \geq \Theta_1, \forall\ 0 \leq d_m \leq k\} \tag{2.13}$$

$$\begin{aligned} \Psi^2_{m,k} &= \{d_m \mid \Theta_0 < a_{m,k} < \Theta_1, \forall\ 0 \leq d_m \leq k\} \\ &= \overline{(\Psi^0_{m,k} \cup \Psi^1_{m,k})}, \end{aligned} \tag{2.14}$$

where d_m is the number of observations whose sensing result is 0 on channel m. We then present two channel sensing policies based on this classification in the following.

Memoryless Sensing Policy In this section, we first present a memoryless sensing policy, with which secondary users cooperatively sense the licensed channels. We call the policy "memoryless" since it does not consider the channel sensing and access results in the previous time slot for simplicity. With this memoryless policy, each secondary user chooses one of the M licensed channels to sense with equal probability, i.e., $1/M$. Furthermore, channel selections of the N secondary users are independent and identically distributed (i.i.d.).

Let U_m be the random variable representing the number of secondary users that select channel m to sense. The probability that u_m secondary users choose channel m to sense is

$$\Pr(U_m = u_m) = \binom{N}{u_m} \left(\frac{1}{M}\right)^{u_m} \left(\frac{M-1}{M}\right)^{N-u_m}. \tag{2.15}$$

The joint distribution that there are u_1 secondary users sensing channel 1, u_2 secondary users sensing channel 2, $\cdots$, and u_M secondary users sensing channel M, is

$$\Pr(u_1, u_2, \cdots, u_M) = \begin{cases} \frac{N!}{\prod_{m=1}^{M} u_m!} \times \frac{1}{M^N}, & \text{if } \sum_{m=1}^{M} u_m = N \\ 0, & \text{otherwise.} \end{cases} \tag{2.16}$$

We next derive the conditional probability that secondary users compete for the channel after the sensing phase stops at the end of mini-slot $K_m < \bar{K}$. The stop time $K_m < \bar{K}$ has two implications. First, it means that secondary users stop sensing channel m after mini-slot K_m. Second, it indicates that the estimated availability of channel m, $a_{m,k}$, has already exceeded the threshold Θ_1. Thus these secondary users think channel m is idle and are ready to access the channel for data transmission. Note that a secondary user also stops sensing a channel m when $a_{m,k} < \Theta_0$ (when it is sure that the channel is busy). We are not interested in this case, since the secondary user will back off until the next time slot. Thus K_m is defined with regard to the event $a_{m,k} > \Theta_1$.

There are U_m users sensing channel m and $U_m K_m$ observations are available after mini-slot K_m, which is also a random variable. We first derive the conditional probability for event $K_m = 1$, as

$$\begin{aligned} &\Pr(K_m = 1 \mid U_m = u\ ,\ S_m = 0) = \Pr(a_{m,u} \geq \Theta_1) \\ = &\sum_{d_m^1 \in \Psi_{m,u}^1} \binom{u}{d_m^1} \left[(\epsilon_m)^{u - d_m^1} (1-\epsilon_m)^{d_m^1} \right], \end{aligned} \tag{2.17}$$

where d_m^1 is the number of observations whose sensing result is 0 in the first mini-slot.

Following similar reasoning as in (2.17), we can obtain the conditional probability for the event that the stop time $K_m = 2$ as

$$\begin{aligned}
&\Pr(K_m = 2 \mid U_m = u\ ,\ S_m = 0) = \Pr\left[(\Theta_0 < a_{m,u} < \Theta_1) \cap (a_{m,2u} \geq \Theta_1)\right] \\
= &\sum_{D_m^2 \in \Psi_{m,2u}^1} \sum_{d_m^1 \in \Psi_{m,u}^2} \binom{u}{d_m^1} \binom{u}{d_m^2} \left[(\epsilon_m)^{2u - D_m^2} (1-\epsilon_m)^{D_m^2}\right], \qquad (2.18)
\end{aligned}$$

where $\Psi_{m,k}^2$ is defined in (2.14) and $D_m^2 = d_m^1 + d_m^2$. In the general case, we can derive the conditional probability for the event that the stop time is $K_m = k$ as:

$$\begin{aligned}
&\Pr(K_m = k \mid U_m = u\ ,\ S_m = 0) \\
&= \Pr\left[(\Theta_0 < a_{m,u} < \Theta_1) \cap (\Theta_0 < a_{m,2u} < \Theta_1) \cap \cdots \cap (\Theta_0 < a_{m,(k-1)u} < \Theta_1) \cap (a_{m,ku} \geq \Theta_1)\right] \\
&= \sum_{D_m^k \in \Psi_{m,ku}^1} \sum_{D_m^{k-1} \in \Psi_{m,(k-1)u}^2} \cdots \sum_{d_m^1 \in \Psi_{m,u}^2} \binom{u}{d_m^1} \binom{u}{d_m^2} \cdots \binom{u}{d_m^k} \left[(\epsilon_m)^{ku - D_m^k} (1-\epsilon_m)^{D_m^k}\right], \quad (2.19)
\end{aligned}$$

where $k = 1, \cdots, \bar{K}$ and $D_m^k = \sum_{i=1}^k d_m^i$. We will apply these results in Section 2.4.3 to derive the throughput of the CR network by the *law of total probability*.

Improved Sensing Policy Under the memoryless sensing policy, some channels may not be sensed by any of the secondary users. Such an event occurs with probability $\Pr(U_m = 0) = \left(\frac{M-1}{M}\right)^N$, which is sufficiently large when M is large and/or the number of secondary users is close to the number of channels. Secondary users will not be able to estimate the state of a channel that nobody senses, and will neither access it in the transmission phase. Therefore, the spectrum opportunities in that channel will be wasted when such events occur.

Motivated by this observation, we develop an *improved sensing* policy that attempts to reduce the chance that a channel is not sensed by any of the secondary users. The improved sensing policy

incorporates a mechanism to spread secondary users to the channels. It also exploits channel state autocorrelation by considering sensing results and channel states in the previous time slot.

By the end of the sensing phase in a time slot t, the secondary users compute the channel availability $a_{m,k}$ for each channel m. During the following transmission phase, if a secondary user transmits on channel m, it can obtain more accurate channel state information: if its transmission is successful, then channel m is idle in time slot t; otherwise, channel m is busy in the time slot. Such channel information can be exchanged at the beginning of the sensing phase in the next time slot. Then, we can classify the M channels into three sets according to the channel states in time slot t, including

- The set of channels that are detected or believed to be idle, denoted by $\mathcal{B}_0(t)$.
- The set of channels that are detected or believed to be busy, denoted by $\mathcal{B}_1(t)$.
- The set of channels whose states are not identified due to lack of time or not sensed by any of the secondary users, denoted by $\mathcal{B}_2(t)$.

Let $|\mathcal{B}_0(t)|$, $|\mathcal{B}_1(t)|$ and $|\mathcal{B}_2(t)|$ be the cardinalities of $\mathcal{B}_0(t)$, $\mathcal{B}_1(t)$, and $\mathcal{B}_2(t)$, respectively.

If channel m is in set $\mathcal{B}_0(t)$ and the stop time on channel m is less than the maximum stop time $\bar{K}$, one user among those u_m users that are sensing this channel will be randomly chosen to switch to sense another channel in the set $m \cup \mathcal{B}_1(t) \cup \mathcal{B}_2(t)$ in time slot $(t+1)$. If channel m is in set $\mathcal{B}_1(t)$ and the stop time on channel m is less than the maximum stop time $\bar{K}$, the secondary users that are sensing this channel will randomly choose a channel in $m \cup \mathcal{B}_2(t)$ to sense in time slot $(t+1)$. With the above mechanism that reassigns secondary users to channels based on the sensing results in the previous time slot, we can reduce the chance that a licensed channel is not sensed by any of the users. This approach achieves the *load balancing* effect since it attempts to spread out secondary users to the channels.

Transmission Phase

We adopt the p-persistent CSMA protocol for data channel access for secondary users during the data transmission phase. Under this protocol, a secondary user delays its transmission when the channels are busy. Once one or more channels are detected idle, the secondary user will attempt to access the idle channel(s) for data transmission with probability p. We consider the heavy load domain, where each secondary user always has data to send to every other secondary user. The following two cases are investigated for opportunistic spectrum access for secondary users.

Case 1 Once the estimate of channel m, i.e., $a_{m,k}$, exceeds threshold Θ_1, each of the secondary users sensing channel m will send an RTS packet on channel m with probability p, to contend for the transmission opportunity on this channel. If there is only one secondary user that sends RTS, then it wins the channel; if there is no secondary user that sends RTS, then the channel will not be accessed and will be wasted; if there are more than one RTS packets sent on channel m, there is collision and none of the secondary users can use the channel.

We define P_m^{idle}, P_m^{succ} and P_m^{coll} as the probability that there is no RTS transmission on channel m, the probability that exactly one secondary user successfully transmits an RTS on channel m, and the probability that there is collision on channel m when multiple RTS packets are transmitted, respectively. Recall that U_m is the number of secondary users that choose channel m to sense. This set of secondary users also attempt to access channel m if it is found idle. With p-persistent CSMA, it follows that

$$P_m^{idle}(U_m) = (1-p)^{U_m} \tag{2.20}$$

$$P_m^{succ}(U_m) = U_m \times p \times (1-p)^{U_m-1} \tag{2.21}$$

$$\begin{aligned} P_m^{coll}(U_m) &= 1 - P_m^{idle}(U_m) - P_m^{succ}(U_m) \\ &= 1 - (1-p)^{U_m} - U_m \times p \times (1-p)^{U_m-1}. \end{aligned} \tag{2.22}$$

Case 2 We assume that the CR users can transmit data over more than one channels using the channel bonding/aggregation techniques [20, 44]. In this case, every secondary user keeps on sensing the channel until the channel state is identified or until the end of the sensing phase. At the beginning of the transmission phase, the set of idle channels are identified and are know to all the secondary users. Then every secondary user will transmit an RTS packet with probability p on the control channel, to contend for the entire set of idle channels. If there is only one secondary user that sends RTS on the control channel, it wins the entire set of idle channels. Otherwise, the idle channels will be wasted (i.e., when no RTS is sent, or more than one RTS are sent on the control channel).

We define P^{idle}, P^{succ} and P^{coll} as the probability of no RTS transmission on the control channel, the probability that exactly one RTS sent on the control channel, and the probability of collision on the control channel, respectively. For p-Persistent CSMA, we have

$$P^{idle}(N) = (1-p)^N \tag{2.23}$$

$$P^{succ}(N) = N \times p \times (1-p)^{N-1} \tag{2.24}$$

$$\begin{aligned} P^{coll}(N) &= 1 - P^{idle}(N) - P^{succ}(N) \\ &= 1 - (1-p)^N - N \times p \times (1-p)^{N-1}. \end{aligned} \tag{2.25}$$

2.4.3 Performance Analysis

Interference Analysis

One of the main challenges in designing a CR network MAC protocol is how to balance the tension between maximizing the capacity of secondary users and protecting primary users from harmful collisions. Let $\gamma_m \in [0,1]$ be the maximum tolerable collision probability to primary users on channel m: $\gamma_m = 0$ means that no secondary transmission is allowed, while $\gamma_m = 1$ means that secondary users have the same privilege as primary users when accessing the channels. The probability of collision caused by secondary users to primary users should be kept below γ_m.

We first derive the conditional probability that channel m is miss detected to be idle by u secondary users after mini-slot k, as follows.

$$\begin{aligned}&\Pr(K_m = k \mid U_m = u, S_m = 1)\\ &= \sum_{D_m^k \in \Psi^1_{m,ku}} \sum_{D_m^{k-1} \in \Psi^2_{m,(k-1)u}} \cdots \sum_{d_m^1 \in \Psi^2_{m,u}} \binom{u}{d_m^1}\binom{u}{d_m^2}\cdots\binom{u}{d_m^k} (\delta_m)^{D_m^k}(1-\delta_m)^{ku-D_m^k}. \quad (2.26)\end{aligned}$$

In Case 1, the idle channels are accessed by different secondary users. The probability that secondary users collide with primary users on channel m is

$$P_{m,1}^{intf} = \sum_{k=1}^{\bar{K}}\sum_{u=0}^{N} \Pr(K_m = k \mid U_m = u, S_m = 1) \times \Pr(U_m = u) \times \left[P_m^{succ}(u) + P_m^{coll}(u)\right]. \quad (2.27)$$

In Case 2, a winning secondary user takes all the idle channels using the technique of channel bonding/aggregation. The probability that secondary users collide with primary users on channel m is

$$P_{m,2}^{intf} = \sum_{k=1}^{\bar{K}}\sum_{u=0}^{N} \Pr(K_m = k \mid U_m = u, S_m = 1) \times \Pr(U_m = u) \times P^{succ}(N). \quad (2.28)$$

For primary user protection, the probability of secondary users causing collision with primary users on channel m should be kept lower than or equal to γ_m, i.e.,

$$P_{m,i}^{intf} \leq \gamma_m, \text{ for } i = 1, 2. \quad (2.29)$$

This constraint is used to set the channel access probability p for the p-persistent CSMA protocol.

Throughput Analysis

Based on previous analysis, the expected throughput of the proposed CR MAC protocols adopting the two sensing policies, can be derived after the system attains steady state. Without loss of generality, we ignore the time spent on RTS/CTS exchanges, which can be approximated by a fixed amount of overhead.

In Case 1, the expected throughput of channel m that is sensed by u users, denoted by $\Lambda_m^1(u)$, can be derived as

$$\Lambda_m^1(u) \;=\; \sum_{k=1}^{\bar{K}} \Pr(K_m = k \mid U_m = u, S_m = 0) \times R_m \times \frac{1}{T_s} \times \left[(\bar{K}-k)T_{ms} + T_{data}\right], \tag{2.30}$$

where R_m is the data rate of channel m, and T_s is the time slot duration given in (2.5).

Let $\vec{U} = [U_1, U_2, \cdots, U_M]$ denote the secondary user *sensing state vector*, where each element U_m represents the number of secondary users that choose channel m to sense and access. The aggregate throughput for the CR network, denoted by Ω_1, is

$$\Omega_1 = \sum_{\vec{U}} \Pr(\vec{U}) \sum_{\vec{S}} \Pr(\vec{S}) \sum_{m=1}^{M} \left[I_{[S_m=0]} \Lambda_m^1(U_m = u) P_m^{succ}(U_m = u) \right], \tag{2.31}$$

where $\vec{S}$ is the channel state vector defined in Section 2.4.1, $P_m^{succ}(u)$ is given in (2.21) and $I_{[S_m=0]}$ is an indicator that channel m is idle, i.e.,

$$I_{[S_m=0]} = \begin{cases} 1, & \text{if } S_m = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{2.32}$$

In Case 1, the sensing process on channel m can stop early if the estimate of channel availability $a_{m,k}$ exceeds threshold Θ_1 or drops below the threshold Θ_0. In the former case, the remaining mini-slots can be used to transmit data. In Case 2, all CR users wait till the beginning of the transmission phase, and then contend for the idle channels by sending RTS packets on the control channel. The winning secondary user's data transmissions start at the beginning of the transmission

phase (i.e., after $\bar{K}$ mini-slots). We can derive the throughput for channel m as follows.

$$\Lambda_m^2(u) \;=\; \sum_{k=1}^{\bar{K}} \Pr(K_m = k \mid U_m = u, S_m = 0) \times R_m \times \frac{T_{data}}{T_s}, \tag{2.33}$$

The aggregate throughput for the CR network, denoted by Ω_2, is

$$\Omega_2 \;=\; \sum_{\vec{U}} \Pr(\vec{U}) \sum_{\vec{S}} \Pr(\vec{S}) \times \sum_{m=1}^{M} \left[I_{[S_m=0]} \Lambda_m^2(u) P^{succ}(N) \right]. \tag{2.34}$$

2.4.4 Simulation Study

Simulation Settings

We evaluate the performance of the proposed CR MAC protocol using a customized simulator developed with MATLAB. We compare the following four schemes in the simulations:

- A simple random sensing scheme that each user chooses one channel to sense with equal probability, termed *Random* in the plots.
- The negotiate sensing scheme presented in [20], termed *Negotiate* in the plots.
- The memoryless sensing scheme as described in Section 2.4.2. In the figures, *Memoryless*1 refers to transmission scheme Case 1 (i.e., idle channels are accessed by different secondary users, see Section 2.4.2), and *Memoryless*2 refers to transmission scheme Case 2 (i.e., idle channels are accessed by a winning secondary user using channel bonding/aggregation techniques [44]).
- The improved sensing scheme presented in Section 2.4.2. In the figures, *Improved*1 refers to transmission scheme Case 1, and *Improved*2 refers to transmission scheme Case 2.

We choose the negotiate sensing scheme since it adopts a similar network model and assumptions. With this scheme, different secondary users attempt to select distinct channels to sense by overhearing the control packets on the control channel [20]. One of the major differences between

Table 2.1: Simulation Parameters

Symbol	*Value*	*Definition*
T_{ms}	9 μs	mini-slot interval
T_s	1.89 ms	time slot interval
M	5	number of licensed channels
N	8	number of secondary users
η	0.3	utilization of the licensed channels
ϵ	0.3	probability of false alarm
δ	0.3	probability of miss detection
R	1 Mb/s	data rate of each licensed channel
Θ_1	0.8	upper threshold for channel decision
Θ_0	0.2	lower threshold for channel decision
$\bar{K}$	5	maximum stop time for channel sensing

negotiate sensing and the proposed schemes in this paper, is that negotiate sensing does not consider spectrum sensing errors in the MAC protocol design.

The simulation parameters are summarized in Table 2.1, which follow the typical values used in [20]. We run each simulation scenario for 10 times with different random seeds. Each point in the plots shown in this section is the average of 10 simulation runs. We plot 95% confidence intervals as error bars on the simulation curves, which are negligible in all the figures.

Simulation Results

We first verify our throughput analysis presented in Section 2.4.2. In Figs 2.6 and 2.7, we plot the throughputs for the CR MACs incorporating the memoryless sensing policy and the improved sensing policy, with both simulation and analysis curves (dashed curves). We observe that the simulation and analysis curves for the memoryless sensing CR MACs overlap completely with each other, indicating that our analysis is exact. Furthermore, although there is a gap between the simulation and analysis curves for the CR MACs with the improved sensing policy, the gap is generally very small. The gap is actually due to an approximation we used for the secondary user sensing state vector $\vec{U}$, for which deriving the exact form is non-trivial. In the analysis, we assume that the probability is 0 that a channel is not sensed by any secondary user. We find the analysis can

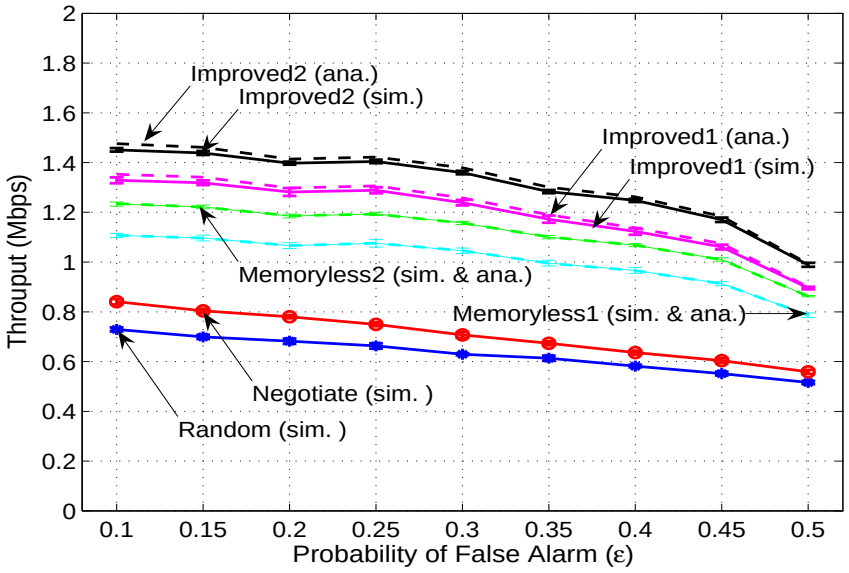


Figure 2.6: Throughput versus false alarm probability (with 95% confidence intervals for the simulation results).

serve as a tight upper bound for the CR MAC throughput performance when the improved sensing policy is incorporated.

We next investigate the impact of sensing errors on the CR MAC performance. We assume identical false alarm probabilities $\epsilon_m = \epsilon$, and identical miss detection probabilities $\delta_m = \delta$ for all the licensed channels. In Fig. 2.6, we plot the throughputs obtained by the four schemes versus the false alarm probability ϵ. Specifically, we fix δ at 0.3 and increase ϵ from 0.1 to 0.5. Intuitively, a higher false alarm probability results in lower probability for secondary users to exploit the transmission opportunities in the licensed channels. This is illustrated in the figure, as all the four throughput curves decrease as ϵ is increased. The improved sensing MAC achieves the best performance, with about 10% gain over the memoryless sensing MAC and about 200% gain over the two existing approaches. The advantage of channel bonding/aggregation is also demonstrated in the figure, where Case 2 transmission scheme always achieves higher throughput than Case 1 scheme.

In Fig. 2.7, we examine the impact of miss detection probability δ on the CR network throughput. In these simulations, we fix ϵ at 0.3 and increase δ from 0.1 to 0.5. We find that the miss detection error has small impact on the throughputs of the random sensing and negotiate sensing

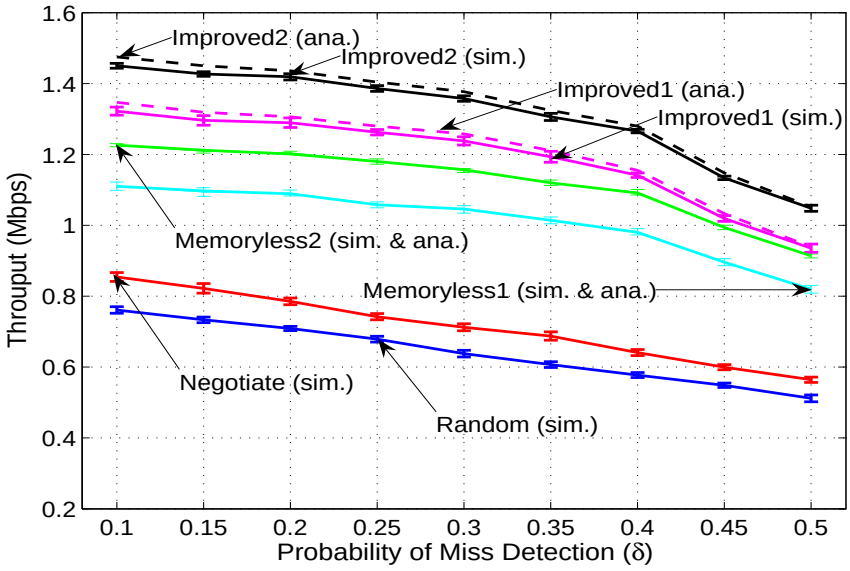


Figure 2.7: Throughput versus miss detection probability (with 95% confidence intervals for the simulation results).

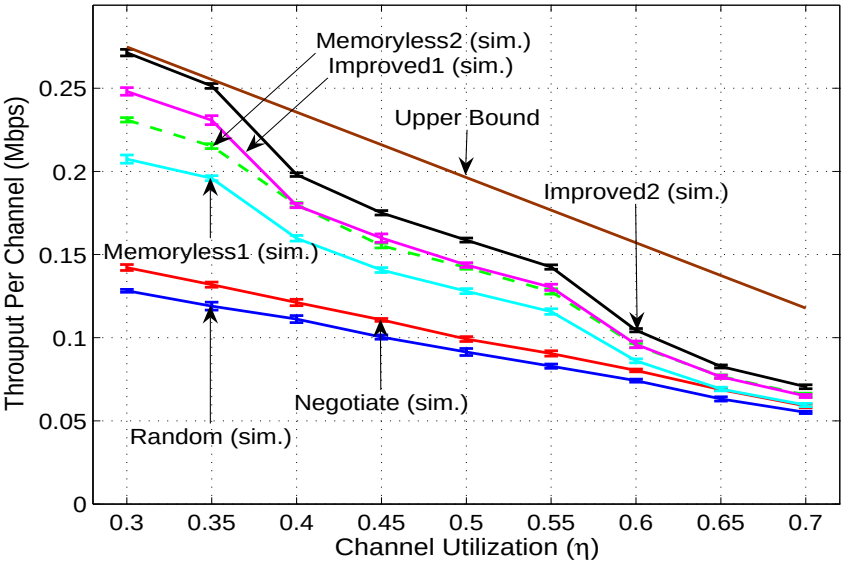


Figure 2.8: Throughput versus channel utilization (with 95% confidence intervals for the simulation results).

protocols, since miss detection errors are not considered in the design of these protocols. However, both our proposed CR MAC schemes achieve considerable throughput gains over the random sensing and negotiate sensing schemes.

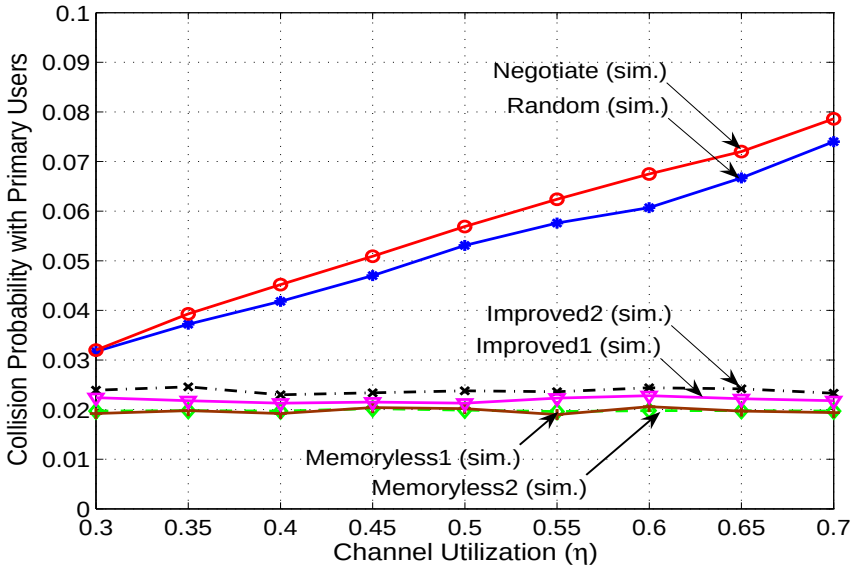


Figure 2.9: Collision probability with primary users when the maximum tolerable collision probability is $\gamma = 3.5\%$.

In Fig. 2.8, we plot the throughput of the four schemes under different channel utilization values ranging from 0.3 to 0.7. As the utilization of the licensed channels is increased, the transmission opportunities for secondary users are clearly reduced. Therefore the four curves are all decreasing function of η. The improved policy with transmission scheme Case 2 achieves the best performance among the four schemes, while random sensing has the poorest performance. When the channel utilization is $\eta = 0.3$, the improved policy achieves a 10% gain in throughput over the memoryless sensing policy. We also plot the upper bound on the CR network throughput, as given by the channel idle probability in (2.2). When the channel utilization is low, the improved policy with transmission scheme Case 2 can achieve a throughput very close to the upper bound. The gap between the upper bound and the achievable throughput increase when the primary users get more busy.

In Fig. 2.9, we plot the collision probability caused by secondary transmissions to primary users, when the maximum allowable collision probability is set as $\gamma = 3.5\%$. We plot the measured collision probabilities in the simulations when the channel utilization is increased from 30% to 70%. It can be seen that the collision probabilities of random and negotiate sensing schemes

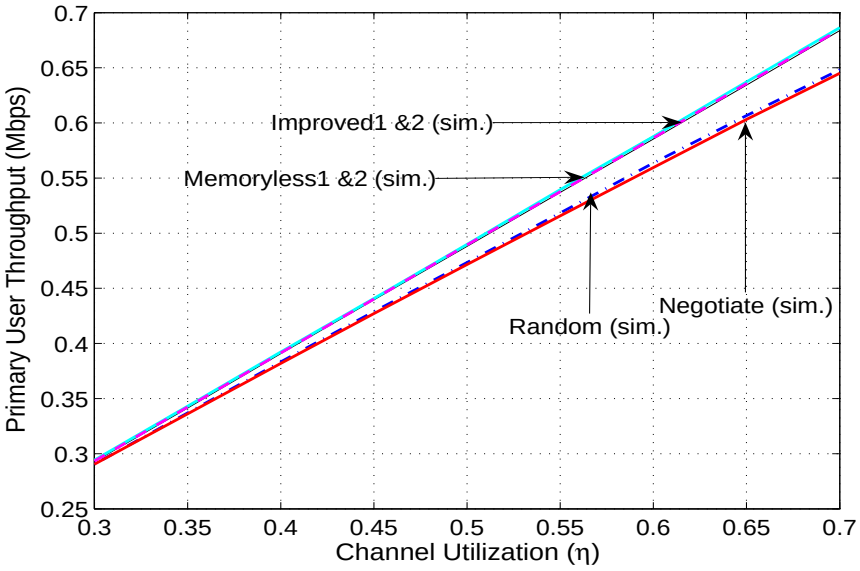


Figure 2.10: Total throughput of primary users when they become more active.

increases along with η and soon exceed the 3.5% threshold. On the other hand, the collision probabilities of the proposed schemes are kept around 2.5% for the entire range of η examined.

Finally, we plot the throughput of the primary users in Fig. 2.10. The primary user throughput curves for all the four schemes increase when the channel utilization η is increased. The gap between the curves of the proposed schemes and those of random and negotiate sensing schemes, is due to the different collision rates secondary users introduce to primary users under these schemes (see Fig. 2.9). As η is increased, the proposed schemes introduces relatively constant collision rates to primary users (i.e., around 2.5%), while the random and negotiate sensing schemes introduce increasingly higher collision rates to primary users, which degrade the throughput of primary users.

2.5 Co-channel and Adjacent Channel Interference Mitigation

In this section, we consider a CR network consisting of multiple CR transmitter receiver pairs. The primary network comprises a base station sending data to primary users using a set of licensed channels. The CR nodes collaboratively sense the licensed channels and exploit spectrum opportunities for data transmission. We investigate the problem of maximizing the CR network throughput while bounding the interference to primary users. We incorporate several important components

such as cooperative spectrum sensing, spectrum sensing errors, and opportunistic spectrum access into the cross-layer optimization framework. In the problem formulation, we specifically consider mitigating CCI among CR users and ACI for both CR and primary users, through optimized channel assignment and transmit power control for CR users.

The formulated problem is a Mixed Integer Nonlinear Programming (MINLP) problem, due to the use of index variables for channel assignment and logarithmic relationship between link capacity and signal to noise ratio (SNR). Such problems are NP-hard in general. We first propose a reformulation-linearization technique (RLT)-based centralized algorithm that computes near-optimal solutions in polynomial time [45]. We then develop a distributed greedy algorithm that uses only local information and computes near-optimal solutions. Through simulation studies, we find the distributed greedy algorithm outperforms both the RLT-based centralized algorithm and a heuristic channel assignment algorithm that exploits multiuser diversity with considerable gains.

2.5.1 Network Model and Assumptions

As shown in Fig. 2.11, we consider a primary network where a base station transmits data to primary users using M licensed channels with non-overlapping spectrum. Without loss of generality, we assume the channels have identical bandwidth. We assume that each primary user is equipped with one transceiver and can communicate with the primary base station via one of the licensed channels. Let $\mathcal{P}_m$ be the subset of primary users that are tuned to channel m. All the $\mathcal{P}_m$'s are generally assumed non-empty.

As illustrated in Section 2.3.1, we assume that the primary network uses a synchronous time slot structure. The occupancy of each channel is modeled as a discrete-time Markov process. As discussed before, the utilization of channel m with respect to primary user transmission, denoted by η_m, can be computed as (2.1)

Within the coverage of the primary network, there are K pairs of CR transmitters and receivers that explore the spectrum opportunities in the M licensed channels for data communications. Each CR node is equipped with two transceivers: a *control transceiver* that operates on a dedicated

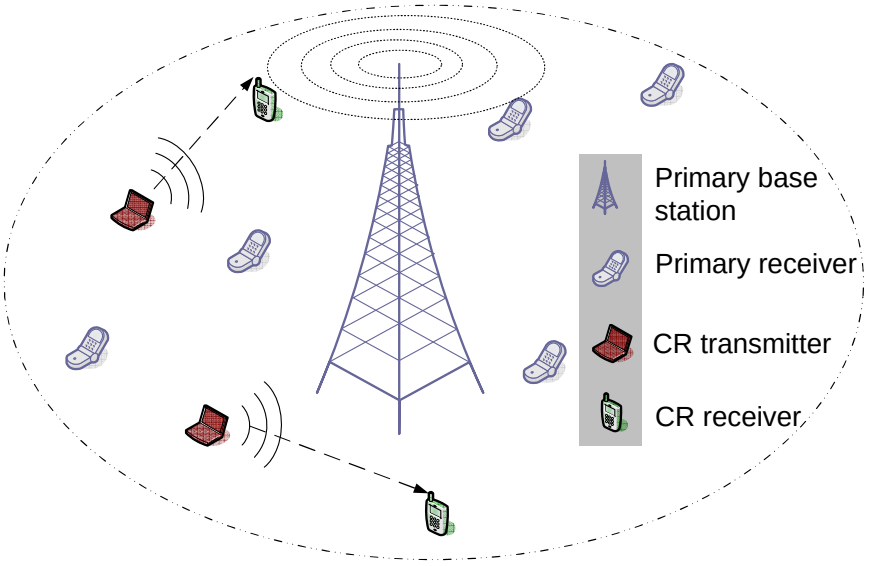


Figure 2.11: The primary and CR network model.

control channel (which we assume is always available), and a *data transceiver* incorporating a software defined radio (SDR) that is able to tune to any of the M licensed channels.

CR nodes access the licensed channels following the same time slot structure as in the primary network. For CR nodes, each time slot consists of a sensing phase and a transmission phase (see Section 2.3.2). In the sensing phase, a CR node chooses one of the M channels to sense using its data transceiver, and then exchanges the sensed channel information with other CR nodes using its control transceiver over the control channel. During the transmission phase, the CR node turns its data transceiver to one of the M channels to transmit or receive data based on sensing results.

Cooperative Spectrum Sensing

As illustrated in Section 2.3.3, we consider two types of sensing errors (false alarm ϵ_n^m and miss detection δ_n^m) in the sensing phase. During the sensing phase, the CR nodes exchange their sensing results through the control channel. As presented in (4.1), the conditional probability that channel m is available for given N sensing results on channel m, denoted by $P_m^A(\Theta_1^m, \cdots, \Theta_N^m)$,

can be computed as [46]:

$$\begin{aligned} & P_m^A(\Theta_1^m, \cdots, \Theta_N^m) \\ = & \frac{\Pr\{\Theta_1^m, \Theta_2^m, \cdots, \Theta_N^m | H_0^m)\} \Pr\{H_0^m\}}{\Pr\{\Theta_1^m, \Theta_2^m, \cdots, \Theta_N^m | H_0^m\} \Pr\{H_0^m\} + \Pr\{\Theta_1^m, \Theta_2^m, \cdots, \Theta_N^m | H_1^m\} \Pr\{H_1^m\}} \\ = & \frac{\prod_{n=1}^N \Pr\{\Theta_n^m | H_0^m\} \Pr\{H_0^m\}}{\prod_{n=1}^N \Pr\{\Theta_n^m | H_0^m\} \Pr\{H_0^m\} + \prod_{n=1}^N \Pr\{\Theta_n^m | H_1^m\} \Pr\{H_1^m\}} \\ = & \left[1 + \frac{\Pr\{H_1^m\}}{\Pr\{H_0^m\}} \prod_{n=1}^N \frac{\Pr\{\Theta_n^m | H_1^m\}}{\Pr\{\Theta_n^m | H_0^m\}}\right]^{-1} \\ = & \left[1 + \frac{\eta_m}{1-\eta_m} \prod_{n=1}^N \frac{(\delta_n^m)^{1-\Theta_n^m}(1-\delta_n^m)^{\Theta_n^m}}{(\epsilon_n^m)^{\Theta_n^m}(1-\epsilon_n^m)^{1-\Theta_n^m}}\right]^{-1} \end{aligned} \tag{2.35}$$

where Θ_n^m is the n-th sensing result on channel m.

When one or more sensing results are received at a CR node, let the sensing result vector be $\vec{\Theta}_n^m = [\Theta_1^m, \Theta_2^m, \cdots, \Theta_n^m]$ for the n received sensing results on channel m. The conditional channel availability probability can be computed iteratively as follows.

$$P_m^A(\Theta_1^m) = \left[1 + \frac{\eta_m}{1-\eta_m} \times \frac{(\delta_1^m)^{1-\Theta_1^m}(1-\delta_1^m)^{\Theta_1^m}}{(\epsilon_1^m)^{\Theta_1^m}(1-\epsilon_1^m)^{1-\Theta_1^m}}\right]^{-1} \tag{2.36}$$

$$\begin{aligned} P_m^A(\vec{\Theta}_n^m) &= P_m^A(\Theta_1^m, \Theta_2^m, \cdots, \Theta_n^m) \\ &= \left\{1 + \left[\frac{1}{P_m^A(\Theta_1^m, \Theta_2^m, \cdots, \Theta_{n-1}^m)} - 1\right] \times \right. \\ &\quad \left. \frac{(\delta_n^m)^{1-\Theta_n^m}(1-\delta_n^m)^{\Theta_n^m}}{(\epsilon_n^m)^{\Theta_n^m}(1-\epsilon_n^m)^{1-\Theta_n^m}}\right\}^{-1}, n = 2, \cdots, N. \end{aligned} \tag{2.37}$$

Opportunistic Channel Access

Let $D_m(t)$ be a decision variable indicating whether channel m will be accessed in time slot t. It is defined as

$$D_m(t) = \begin{cases} 0, & \text{if channel } m \text{ is considered idle} \\ 1, & \text{otherwise} \end{cases} \quad \text{for } m = \{1, 2, \cdots, M\}. \tag{2.38}$$

Based on the spectrum sensing result $P_m^A(\vec{\Theta}_N^m)$, channel m will be accessed (i.e., when $D_m(t) = 0$) with probability $P_m^D(\vec{\Theta}_N^m)$, and it will not be accessed (i.e., when $D_m(t) = 1$) with probability $1 - P_m^D(\vec{\Theta}_N^m)$. We show how to compute $P_m^D(\vec{\Theta}_N^m)$ in the following.

For primary user protection, the probability that a CR transmission collides with primary user transmissions should be smaller with a threshold prescribed by the primary network, denoted by γ_m for channel m. The primary user protection condition can be written as

$$\left[1 - P_m^A(\vec{\Theta}_N^m)\right] P_m^D(\vec{\Theta}_N^m) \leq \gamma_m. \tag{2.39}$$

To maximize the CR network throughput, $P_m^D(\vec{\Theta}_N^m)$ should be set to a probability as large as possible, as allowed by the maximum collision rate constraint. We have from (2.39)

$$P_m^D(\vec{\Theta}_N^m) = \min\left\{\frac{\gamma_m}{1 - P_m^A(\vec{\Theta}_N^m)}, 1\right\}. \tag{2.40}$$

Let $\mathcal{A}(t) := \{m | D_m(t) = 0\}$ be the subset of channels that are identified to be idle in time slot t. Then its complement set $\mathcal{B}(t) = \overline{\mathcal{A}}(t)$ is the subset of channels that are believed to be busy in time slot t (i.e., being used by primary users). We next investigate how to effectively assign the channels in $\mathcal{A}(t)$ to the CR transmitters and how to choose transmit power for the CR transmitters, such that the CR network throughput is maximized under interference and collision rate constraints.

Channel Interference Model

We consider CCI among CR users sharing the same licensed channel and ACI for both CR and primary users in this paper. The CCI and ACI models are presented in the following.

Co-channel Interference CCI is caused by the CR user transmissions sharing the same channel as the victim receiver. Before introducing the interference model, we define index variables x_k^m

indicating the channel assignment for the CR links as follows:

$$x_k^m = \begin{cases} 1, & \text{if channel } m \text{ is used by CR link } k \\ 0, & \text{otherwise,} \end{cases} \quad m = 1, 2, \cdots, M,\ k = 1, 2, \cdots, K. \tag{2.41}$$

Let $\mathcal{T}_k$ and $\mathcal{R}_k$ be the transmitter and receiver of CR link k, respectively. The CCI at CR receiver k on channel m, denoted by C_k^m, is

$$\begin{aligned} C_k^m &= \sum_{i \in \Phi, i \neq k} G_{\mathcal{T}_i, \mathcal{R}_k}^m P_i^m x_i^m \\ &= \sum_{i \in \Phi} G_{\mathcal{T}_i, \mathcal{R}_k}^m P_i^m x_i^m - G_{\mathcal{T}_k, \mathcal{T}_k}^m P_k^m x_k^m, \end{aligned} \tag{2.42}$$

where $G_{\mathcal{T}_i, \mathcal{R}_k}^m$ is the channel gain from CR transmitter i to CR receiver k on channel m, P_i^m is the transmit power of CR transmitter i on channel m, and $\Phi := \{1, 2, \cdots, K\}$ is the set of transmitter/receiver pairs in the CR network.

Adjacent Channel Interference In addition to CCI, a CR receiver may also be interfered by transmissions on an adjacent channel, when the channels are not strictly orthogonal. The interferer could be either a CR transmitter or a primary transmitter (e.g., the primary base station) on the adjacent channel. Such ACI is shown to be harmful with testbed experiments in a recent work [29].

Due to the imperfect design of band-pass filters, a portion of the power on the adjacent channel may leak to the channel being used by CR users. Such leakage is also considered as noise. For ease of explanation, we only consider the ACI from a direct neighboring channel in this paper. Let β_m^{+1} be the ratio of leakage power from channel $(m+1)$ to m, and β_m^{-1} the ratio of leakage power from channel $(m-1)$ to m. We term these leakage power ratios *ACI factor*, which depends on the spectral properties, such as inter-channel distance and channel width, and band-pass filter design.

For a channel m, if its adjacent channel $(m+1)$ is idle, then the ACI is due to the concurrent CR transmissions on channel $(m+1)$. We have

$$\begin{aligned} AC_{m,k}^{+1} &= [1 - D_{m+1}(t)]\beta_m^{+1} \sum_{i \in \Phi, i \neq k} G_{\mathcal{T}_i,\mathcal{R}_k}^{m+1} P_i^{m+1} x_i^{m+1} \\ &= [1 - D_{m+1}(t)]\beta_m^{+1} C_k^{m+1}. \end{aligned} \tag{2.43}$$

Alternatively, if the adjacent channel $(m+1)$ is busy, then the ACI is caused by a primary transmission on channel $(m+1)$. We have

$$AP_{m,k}^{+1} = D_{m+1}(t)\beta_m^{+1} G_{0,\mathcal{R}_k}^{m+1} Q_{m+1}, \tag{2.44}$$

where $G_{0,\mathcal{R}_k}^{m+1}$ is the channel gain from the primary transmitter to CR receiver k on channel $(m+1)$, and Q_{m+1} is the transmit power of the primary transmitter on channel $(m+1)$.

Similarly, ACI on channel m may also come from the adjacent channel on the other side, i.e., channel $(m-1)$. We define $AC_{m,k}^{-1}$ and $AP_{m,k}^{-1}$ as the interference due to CR transmission and primary transmission on channel $m-1$, respectively. These can be computed as

$$AC_{m,k}^{-1} = [1 - D_{m-1}(t)]\beta_m^{-1} C_k^{m-1} \tag{2.45}$$

$$AP_{m,k}^{-1} = D_{m-1}(t)\beta_m^{-1} G_{0,\mathcal{R}_k}^{m-1} Q_{m-1}. \tag{2.46}$$

The total ACI on channel m from its two adjacent channels can be written as

$$A_k^m = AC_{m,k}^{+1} + AP_{m,k}^{+1} + AC_{m,k}^{-1} + AP_{m,k}^{-1}. \tag{2.47}$$

Without loss of generality, we assume $AC_{1,k}^{-1}$, $AP_{1,k}^{-1}$, $AC_{M,k}^{+1}$ and $AP_{M,k}^{+1}$ are all zero for channels 1 and M. This is because the adjacent channels 0 and $(M+1)$ are used by neither primary nor CR users.

On the other hand, primary users may also be interfered by CR users transmitting on an adjacent channel. If channel m is used by primary user j and channel $(m+1)$ is available for CR user access, the ACI received by the primary user is

$$BC_{m,j}^{+1} = [1 - D_{m+1}(t)]\beta_m^{+1} \sum_{i \in \Phi} G_{\mathcal{T}_i,j}^{m+1} P_i^{m+1} x_i^{m+1}. \tag{2.48}$$

The ACI received by the primary user from CR transmissions on channel $(m-1)$ is

$$BC_{m,j}^{-1} = [1 - D_{m-1}(t)]\beta_m^{-1} \sum_{i \in \Phi} G_{\mathcal{T}_i,j}^{m-1} P_i^{m-1} x_i^{m-1}. \tag{2.49}$$

Considering ACI from both sides of channel m, the total ACI at a primary receiver can be written as:

$$B_j^m = BC_{m,j}^{+1} + BC_{m,j}^{-1}. \tag{2.50}$$

Again, we assume $BC_{1,j}^{-1}$ and $BC_{M,j}^{+1}$ are zero for the two edge channels 1 and M.

2.5.2 Channel Selection and Power Allocation

Problem Statement

At each CR receiver, both types of interference from co-channel and adjacent channels are treated as noise. Let υ_k^m be the SNR at CR receiver k on channel m. Then υ_k^m can be written as

$$\upsilon_k^m = \frac{G_{\mathcal{T}_k,\mathcal{R}_k}^m P_k^m x_k^m}{N_0 + C_k^m + A_k^m}, \tag{2.51}$$

where N_0 is the channel noise power. The objective is to maximize the capacity of the CR network as approximated by Shannon capacity. Without loss of generality, we assume that each channel

has unit bandwidth (e.g., 1 MHz). The objective function becomes

$$\max_{P_k^m, x_k^m} \sum_{k\in\Phi} \sum_{m\in\mathcal{A}(t)} \log_2(1+\upsilon_k^m). \tag{2.52}$$

Since each CR user is able to access one channel in each time slot, we have the following channel access constraint.

$$\sum_{m\in\mathcal{A}(t)} x_k^m \leq 1, \quad \text{for all } k \in \Phi. \tag{2.53}$$

Furthermore, each CR transmitter is limited by a peak power constraint. That is

$$\sum_{m\in\mathcal{A}(t)} P_k^m x_k^m \leq \Gamma, \quad \text{for all } k \in \Phi. \tag{2.54}$$

As discussed, the interference from CR transmissions to primary users should be bounded. Recall that $\mathcal{B}(t)$ is the set of busy channels and $\mathcal{P}_m$ is the set of primary users using channel m. Letting the ACI bound be Ω, we have

$$B_j^m \leq \Omega, \quad \text{for all } m \in \mathcal{B}(t), j \in \mathcal{P}_m. \tag{2.55}$$

Problem (2.52) with constraints (2.53), (2.54) and (2.55) maximizes the CR network capacity while bounding the total interference (i.e., both CCI and ACI) to primary users. Note that the maximum collision rate constraint caused by CR transmissions (2.39) is satisfied by choosing the channel m access probability $P_m^D(\vec{\Theta}_N^m)$ as in (2.40). Based on spectrum sensing results, we need to determine channel access (as given by the x_k^m's) as well transmit powers (as given by the P_k^m's) for CR users. This is an MINLP problem, which is NP-hard in general and cannot be solved exactly in polynomial time. For this problem, we first describe below how to derive upper and lower bounds with a centralized algorithm, and then present a distributed algorithm that decomposes

Problem (2.52) into a channel assignment subproblem and a power allocations subproblem in the next section.

Centralized Algorithm and Performance Bounds

In this section, we first obtain an upper bound by relaxing the problem with RLT [45]. The lower bound is then computed with a sequential fixing (SF) algorithm [8, 47]. To obtain a linear relaxation of the MINLP problem, we first allow the binary variables x_k^m to take real values in $[0, 1]$. Second, the product term $P_k^m x_k^m$ is replaced by a substitution variable $\phi_k^m = P_k^m x_k^m$. Since $0 \leq P_k^m \leq \Gamma$ and $0 \leq x_k^m \leq 1$, we derive the following *RLT bound-factor product constraints*.

$$\begin{cases} (P_k^m - 0)(x_k^m - 0) \geq 0 \\ (P_k^m - 0)(1 - x_k^m) \geq 0 \\ (\Gamma - P_k^m)(x_k^m - 0) \geq 0 \\ (\Gamma - P_k^m)(1 - x_k^m) \geq 0 \end{cases} \tag{2.56}$$

Rearranging the terms, we have

$$\begin{cases} \phi_k^m \geq 0 \\ P_k^m - \phi_k^m \geq 0 \\ \Gamma x_k^m - \phi_k^m \geq 0 \\ P_k^m + \Gamma x_k^m - \phi_k^m \leq \Gamma. \end{cases} \tag{2.57}$$

Finally, the logarithm term $\log_2(1 + \upsilon_k^m)$ in the objective function can be decomposed into the difference between two logarithm terms, denoted by y_k^m and z_k^m, respectively, as follows.

$$\begin{aligned} \log_2(1 + \upsilon_k^m) &= \log_2(N_0 + C_k^m + A_k^m + G_{\mathcal{T}_k,\mathcal{R}_k}^m P_k^m x_k^m) - \log_2(N_0 + C_k^m + A_k^m) \\ &= y_k^m - z_k^m, \end{aligned} \tag{2.58}$$

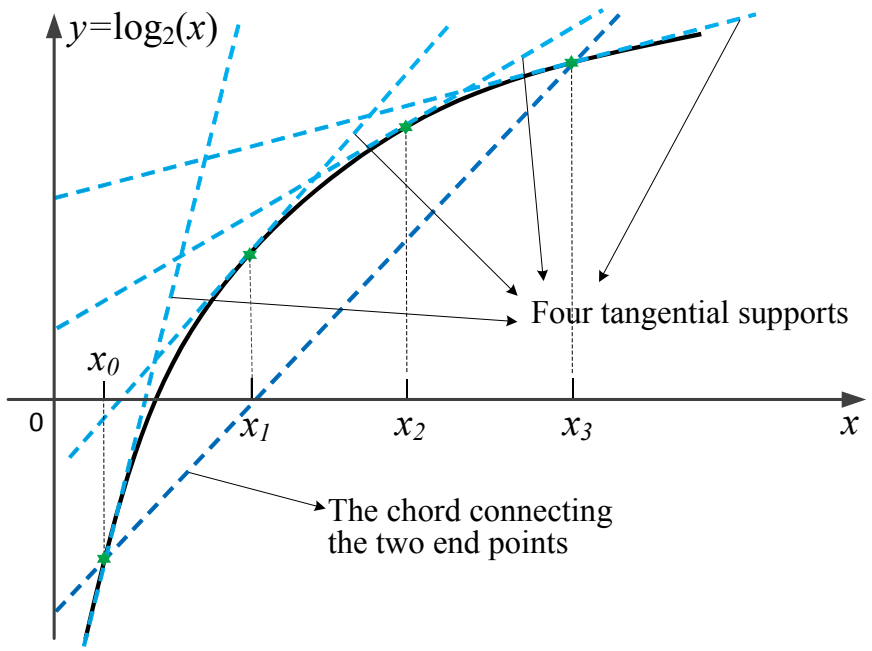


Figure 2.12: The polyhedral outer approximation of a logarithm function $y = \log_2(x)$ in $x_0 \leq x \leq x_3$.

where $y_k^m := \log_2(N_0 + C_k^m + A_k^m + G_{\mathcal{T}_k,\mathcal{R}_k}^m P_k^m x_k^m)$ and $z_k^m := \log_2(N_0 + C_k^m + A_k^m)$. For a general logarithm term $\log_2(x)$, we can linearize it over some tightly bounded regions with a *polyhedral outer approximation*. For example, if x is bounded by $x_0 \leq x \leq x_L$, we can determine L evenly spaced points as

$$x_l = x_0 + \frac{l}{L}(x_L - x_0), \text{ for } l = 0, 1, \cdots, L. \tag{2.59}$$

Then the logarithmic function $y = \log_2(x)$ can be substituted with the following linear constraints.

$$\begin{cases} y \geq \frac{\log_2(x_L) - \log_2(x_0)}{x_L - x_0}(x - x_0) + \log_2(x_0) \\ y \leq \frac{1}{\ln(2) x_l}(x - x_l) + \log_2(x_l), \text{ for } l = 0, 1, \cdots, L. \end{cases} \tag{2.60}$$

In this paper, we use a four-pointer (i.e., $L = 3$) tangential approximation. The upper and lower bound of y_k^m and z_k^m can be obtained by letting ϕ_k^m be 0 and Γ, respectively. The five linear constraints given in (2.60) that form the polyhedral outer approximation are plotted in Fig. 2.12.

Table 2.2: Sequential Fixing (SF) Algorithm

1:	Use RLT to linearize the original problem;
2:	Solve the LP relaxation;
3:	Find the minimum value among all x_k^m and $1 - x_k^m$;
4:	IF (the minimum value is in the form of $1 - x_{k'}^{m'}$)
5:	Fix $x_{k'}^{m'}$ to 1;
6:	Fix all $x_{k'}^{m}$ to 0 for all $m \neq m'$;
7:	ELSE
8:	Fix $x_{k'}^{m'}$ to 0;
9:	END IF
10:	IF (all x_k^m's are fixed)
11:	Go to Step 16;
12:	ELSE
13:	Reformulate and solve a new LP relaxation based on all the x_k^m's that have been fixed;
14:	Go to Step 3;
15:	END IF
16:	Formulate and solve a new LP for the P_k^m's;
17:	Substitute the $x_{k'}^{m'}$'s and P_k^m's into (2.52) to obtain a lower bound on the CR network capacity;

With the above three-step relaxations, we thus obtain a linear programming (LP) relaxation for Problem (2.52). Solving the LP relaxation with an LP solver, we can obtain a possibly infeasible solution due to the relaxations, which can serve as an *upper bound* for the original problem.

We next present an SF Algorithm in Table 2.2 for deriving a feasible near-optimal solution. In Steps $3 - 9$, the variable $x_{k'}^{m'}$ that is closest to 0 or 1 is chosen and rounded to the nearest binary integer. Once $x_{k'}^{m'}$ is fixed to 1, all the other variables $x_{k'}^{m}$ with the same subscript k' are fixed to 0, due to constraint (2.53). Then the problem can be reformulated with a reduced size, and solved again iteratively, until all the binary variables x_k^m's are fixed. In Step 16, the transmit powers P_k^m's are derived when the channel assignment is determined. Note that here we still need to formulate and solve an LP relaxation with respect to the logarithmic terms, since even when the binary variables are fixed, the problem is still non-convex. Finally, in Step 17, we substitute the near-optimal feasible solution into the original objective function (2.52) to obtain a *lower bound* for the global optimum.

Generally, solving such an LP relaxation may produce an infeasible solution to the original problem due to the relaxations. A local search algorithm is then needed to find a feasible solution in the neighborhood. However, local search is not necessary in our problem. First, the SF algorithm determines binary values for all the x_k^m's, such that constraint (2.41) is satisfied. Second, from the RLT bound-factor product constraints (2.57), we have that $\phi_k^m = P_k^m$ when $x_k^m = 1$; and $\phi_k^m = 0$ when $x_k^m = 0$. Once all the x_k^m's are set to binary values, by replacing ϕ_k^m with P_k^m or 0, the linear inequality constraints now only contain the P_k^m's. Finally, although we use the polyhedral outer approximation for the logarithmic terms, these terms are all in the objective function (2.52). It is easy to show that the feasible solutions of the relaxed LP problem is a subset of the feasible solution of the original problem, since the constraints of the original problem is a subset of those of the relaxed LP problem. Therefore, a feasible problem to the LP relaxation is also feasible to the original problem. We only need to substitute the feasible solution into the original objective function (2.52) to obtain the corresponding objective value, which is a lower bound for the original problem.

In our simulations, we find the upper bound quite loose, but the lower bound is reasonably tight. The average-case time complexity of the *simplex method*, a popular LP solving algorithm, is $O(n \log n)$ for a problem with size n [48]. Thus the computational complexity of one iteration in SF is $O(MK \log(MK))$. Since the number of iterations in SF is MK in the worst case, the overall average-case computational complexity of SF is $O(M^2K^2 \log(MK))$.

Distributed Algorithm

The Algorithm Although the SF algorithm in Table 2.2 can compute a near-optimal solution in polynomial time, it is a centralized algorithm that needs to know all the channel gains. In this section, we present a distributed greedy algorithm for solving Problem (2.52). With this algorithm, each CR transmitter estimates channel gains from itself to primary users and all other CR receivers, and each CR receiver estimates channel gains from the primary base station and all other CR transmitters.

Table 2.3: Channel Assignment Algorithm for CR Link k

1:	Initialize $x_k^m = 1$ for all $m \in \mathcal{A}(t)$ and $\mathcal{A}_k(t) = \mathcal{A}(t)$;
2:	WHILE ($\|\mathcal{A}_k(t)\| > 1$)
3:	Run the power allocation algorithm given in Table 2.4;
4:	Find the channel m' with the minimum $U_k^{m'}$ value: $m' = \arg\min_{m \in \mathcal{A}_k(t)} U_k^m(\vec{P})$;
5:	Set $x_k^{m'} = 0$ and remove m' from $\mathcal{A}_k(t)$;
6:	END WHILE
7:	Run the power allocation algorithm given in Table 2.4 with all the x_k^m's determined.

The distributed algorithm consists of two tiers: (i) the upper tier is a *channel assignment algorithm*, which decides which channel to access for a CR transmitter, and (ii) the low tier is a *power allocation algorithm*, which decides how much power can be allocated to transmit on each available channel. In the channel assignment algorithm, we assume the transmit powers have been allocated to each available channel; the power allocation, denoted by an $M \times K$ vector $\vec{P}$, can be obtained from the power allocation algorithm. Define the capacity of CR link k if it uses channel m as

$$U_k^m = \log_2(1 + \upsilon_k^m). \tag{2.61}$$

Then in each loop, the channel with the lowest $U_k^m(\vec{P})$ is removed from the available channel set $\mathcal{A}_k(t)$ and the corresponding x_k^m is set to 0, until only one available channel is left. The complete channel assignment algorithm is presented in Table 2.3. With the algorithm, initially we assume CR transmitter k uses all the channels in $\mathcal{A}(t)$ in Step 1. Then in Steps $2-6$, we iteratively remove the channels with the minimum capacity gain, until only one channel is left. Finally, the transmit power is determined for the chosen channel in Step 7.

In the power allocation algorithm, the main idea is to iteratively allocate a small amount of power Δ to the CR link that can achieve the largest increase in (2.52). The algorithm is presented in Table 2.4. Let $\vec{\Delta}_k^m$ be a vector whose $[(k-1) \times M + m]$-th element is Δ and all other elements are 0, indicating that power Δ is allocated to CR link k on channel m. Obviously, if CR link k is allocated with the additional power Δ, the throughput of this CR link will increase; if another CR

Table 2.4: Power Allocation Algorithm for CR Link k

1:	Initialize $P_k^m = 0$ for all $k \in \Phi$ and $m \in \mathcal{A}_k(t)$;
2:	Calculate $EARN_k^m$ for all $m \in \mathcal{A}_k(t)$;
3:	IF ($\sum_{m\in\mathcal{A}(t)} P_k^m + \Delta \le \Gamma$) & ($B_j^m(\vec{P} + \vec{\Delta}_k^m) \le \Omega$ for $j \in \mathcal{P}_m$)
4:	$EARN_k^m = \sum_{n\in\mathcal{A}(t)}[U_k^n(\vec{P} + \vec{\Delta}_k^m) - U_k^n(\vec{P})]$;
5:	ELSE
6:	$EARN_k^m = 0$;
7:	END IF
8:	Calculate $COST_{k',k}^m$ for all $k' \neq k$ and $m \in \mathcal{A}_{k'}(t)$, where $Cost_{k',k}^m = \sum_{n\in\mathcal{A}(t)}[U_k^n(\vec{P} + \vec{\Delta}_{k'}^m) - U_k^n(\vec{P})]$;
9:	Broadcast all $EARN_k^m$ and $COST_{k',k}^m$ to all other CR links;
10:	Calculate $PROFIT_k^m$ for all $k \in \Phi$ and $m \in \mathcal{A}_k(t)$, where $PROFIT_k^m = EARN_k^m + \sum_{k\neq k'} COST_{k,k'}^m$;
11:	Find $\{m', k'\} = \arg\max_{k\in\Phi \& m\in\mathcal{A}_k(t)} PROFIT_k^m$;
12:	IF ($PROFIT_{k'}^{m'} > 0$)
13:	$\vec{P} = \vec{P} + \vec{\Delta}_{k'}^{m'}$;
14:	Go to Step 2;
15:	ELSE.
16:	The algorithm is terminated with solution $\vec{P}$;
17:	END IF

link $k' \neq k$ is allocated with the additional power Δ, the throughput of CR link k will decrease. The increase and decrease of the throughput of the CR link are termed *earning* and *cost*, respectively: $EARN_k^m$ is the throughput *increase* for CR link k on channel m if it gets the additional power Δ; $COST_{k',k}^m$ is the throughput *decrease* for CR link k if another CR link $k' \neq k$ on channel m wins the additional power allocation Δ. In Steps $3-7$, we calculate $EARN_k^m$, but set it to 0 if either (2.54) or (2.55) is not satisfied. In Steps $10-11$, the net throughput gains (or, *profit*) of all possible power allocations are calculated and the combination with the largest profit is selected. In Steps $12-13$, the CR link with the largest positive profit wins the additional power allocation Δ, if its profit is positive. Otherwise, the power allocation algorithm is terminated with solution $\vec{P}$, because no further power allocation can improve the total throughput.

A Simple Example To better explain our distributed algorithm, consider a simple example with two CR users 1 and 2. As shown in Fig. 2.13, for each channel m, CR transmitter 1 calculates

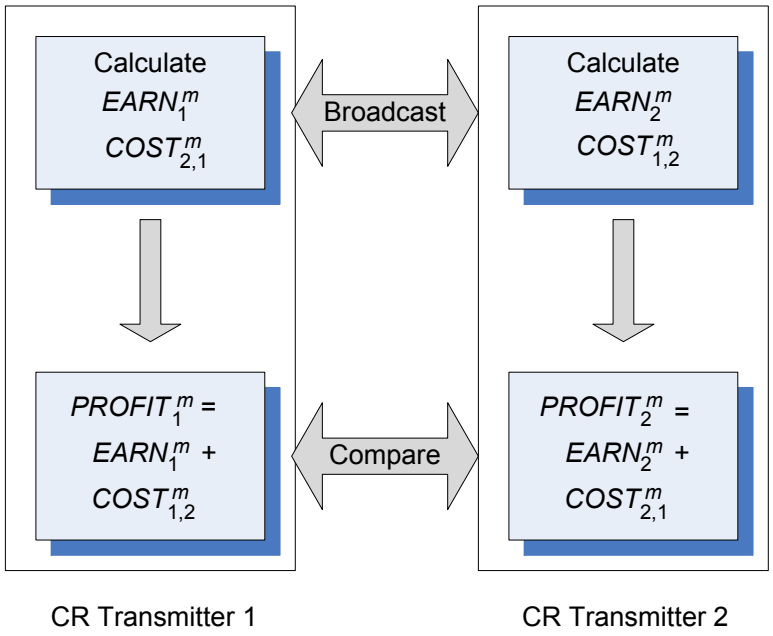


Figure 2.13: An example of the distributed algorithm operation with two CR users.

$EARN_1^m$ and $COST_{2,1}^m$ and CR transmitter 2 calculates $EARN_2^m$ and $COST_{1,2}^m$. The two nodes then broadcast the values to each other. Once a node receives the $COST$ from the other node, it calculates its $PROFIT$ as shown in the figure. The power $\vec{\Delta}_k^m$ is allocated to the node with the larger $PROFIT_k^m$. This power allocation algorithm is terminated if both $PROFIT_k^m$ from the two nodes are non-positive. Once the power allocation algorithm is terminated, the channel with the lowest $U_k^m(\vec{P})$ is removed from the available channel set $\mathcal{A}_k(t)$, until only one available channel is left, as given in Table 2.3.

2.5.3 Performance Evaluation

We evaluate the performance of the proposed algorithms using MATLAB (for solving the LP relaxations). For the results reported in this section, there are $M = 6$ licensed channels (unless otherwise specified) with identical transition probabilities $P_m^{01} = 0.4$ and $P_m^{10} = 0.3$ for all m. The maximum collision probability with primary users is $\gamma_m = 0.2$ for all m. The transmit power of primary base station is 30 dBm and the maximum acceptable interference for the primary users is $\Omega = 10$ dBm. There are $K = 6$ transmitter and receiver pairs in the CR network. The power of

CR transmitter is limited to $\Gamma_m = 27$ dBm for all m. The false alarm probability is $\epsilon_n^m = 0.3$ and the miss detection probability is $\delta_n^m = 0.3$ for all m and n, unless otherwise specified. Rayleigh block fading channels are used in the simulations.

We consider four types of results: (i) the upper bound obtained by solving the RLT relaxation as presented in Section 2.5.2; (ii) the centralized SF algorithm solution given in Table 2.2 (a lower bound); (iii) the distributed greedy algorithm given in Tables 2.3 and 2.4; and (iv) a simple centralized heuristic algorithm. With the centralized heuristic algorithm, each CR transmitter chooses the best available channel to access to exploit multiuser diversity gain. When the channels are assigned and all the x_k^m's are fixed. Then it solves the reduced problem (2.52) with MATLAB Optimization Toolbox to find a near-optimal power allocation. Each point in the curves is the average of 10 simulations with different random seeds. The 95% confidence intervals are plotted as error bars, which are negligible in all the cases.

We first examine the impact of the number of channels M on the overall throughput of the CR network. In Fig. 2.14, we increase M from 4 to 8, and plot the total throughput of the CR network. As expected, the more licensed channels, the more spectrum opportunities for CR users and the higher the network throughput. The curves of both SF and the heuristic algorithm have lower slop than that of the distributed greedy algorithm. It implies that the greedy algorithm is more efficient in exploiting the addition spectrum opportunities for CR transmissions. We find the upper bound quite loose, while the lower bound is reasonably tight.

In Fig. 2.15, we investigate the impact of primary user channel utilization η on the CR network throughput. The throughput curves achieved by the algorithms are plotted when η is increased from 0.3 to 0.7. Clearly, a smaller η allows more spectrum opportunities for CR transmissions. When the primary users get more busy, the spectrum opportunities for CR users decreases and the throughput of all the three algorithms decreases. It can be seen from the figure that all the three curves decrease as η gets larger. The CR network throughput of the distributed greedy algorithm is better than that of the simple heuristic algorithm and that of the centralized SF algorithm. In particular, when $\eta = 0.3$, the distributed greedy algorithm achieves a normalized throughput gain

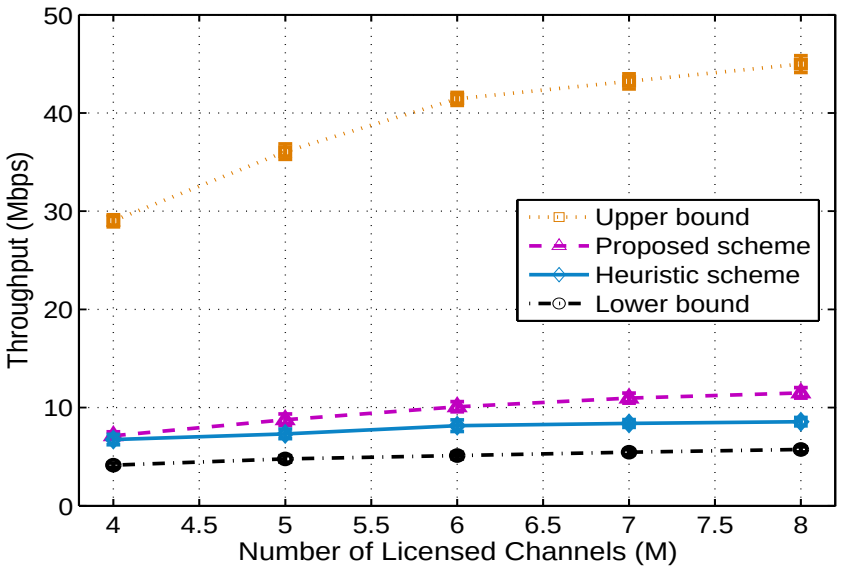


Figure 2.14: CR network throughput versus the number of licensed channels.

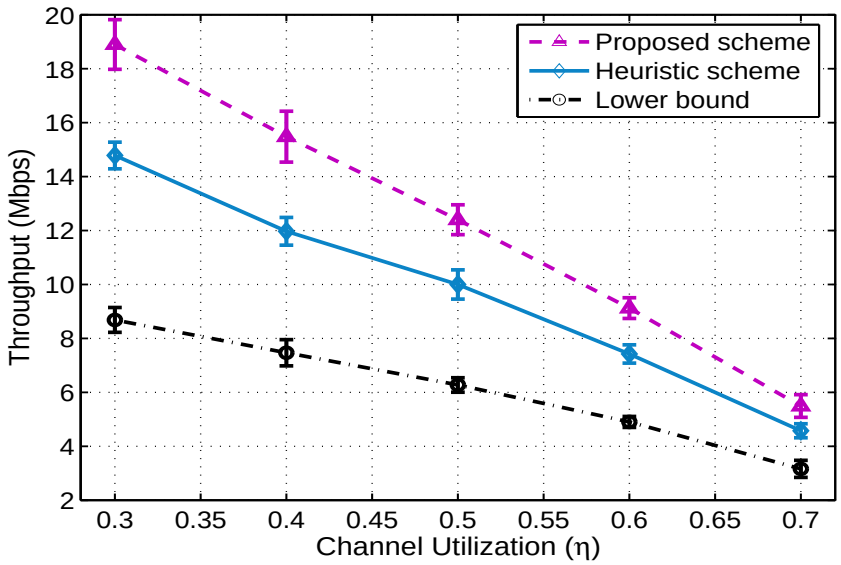


Figure 2.15: CR network throughput versus primary user channel utilization.

of 27.84% over the simple heuristic algorithm, and a normalized throughput gain of 117.62% over SF. When $\eta = 0.7$, the distributed algroithm achieves normalized throughput gains of 19.98% and 73.67% over the simple heuristic and SF, respectively.

Next we examine the impact of spectrum sensing errors on the CR network throughput. In Fig. 2.16, we test five pairs of $\{\epsilon, \delta\}$ values as follows: $\{0.2, 0.48\}$, $\{0.24, 0.38\}$, $\{0.3, 0.3\}$,

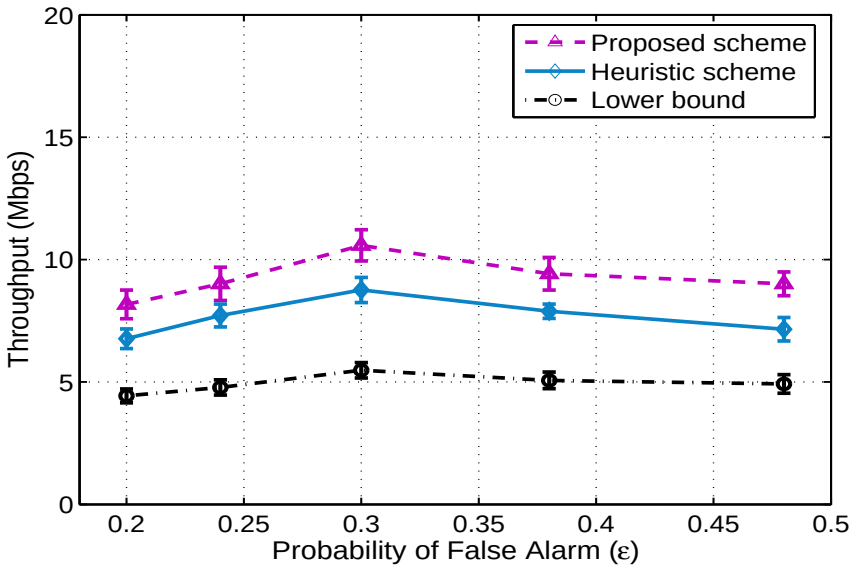


Figure 2.16: CR network throughput versus spectrum sensing error probabilities.

$\{0.38, 0.24\}$, and $\{0.48, 0.2\}$. The CR network throughputs achieved by the algorithms are plotted in the figure. It is interesting to see that the throughput performance gets worse when the probability of one of the two sensing errors gets large. We can trade-off between false alarm and miss detection probabilities to find the optimal operating point for spectrum sensing. Again, the throughput performance of the greedy algorithm is superior to that of the heuristic algorithm and doubles that of the SF algorithm.

We then investigate the impact of the ACI factor β on the CR network throughput. The simulation results are presented in Fig. 2.17, where β is increased from 0 to 0.5. As expected, the CR network throughput is degraded by the presence of ACI. The severer the ACI, the lower the CR network throughput. When β is increased from 0 to 0.5, the throughput degradations are 4.0647 Mbps, 3.8068 Mbps, and 3.3793 Mbps for the distributed algorithm, the simple heuristic, and SF, respectively. The distributed greedy algorithm outperforms both the simple heuristic algorithm and SF with considerable gaps for the entire range of β considered.

We also measure both types of interference in the simulations and exam the impact of the ACI factor β on channel interference. In Fig. 2.18, we increase β from 0 to 0.5 with step 0.1 and plot the measured average interference in the plots. The total average interference for each licensed

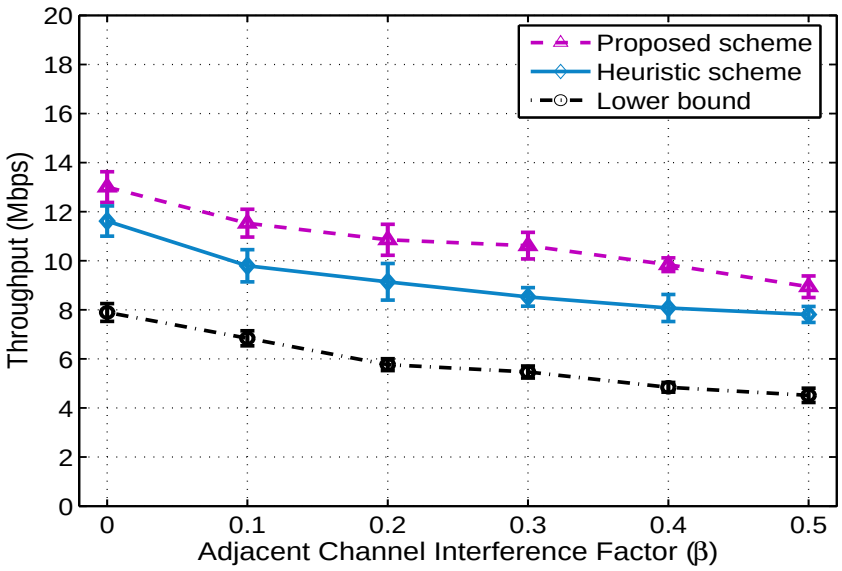
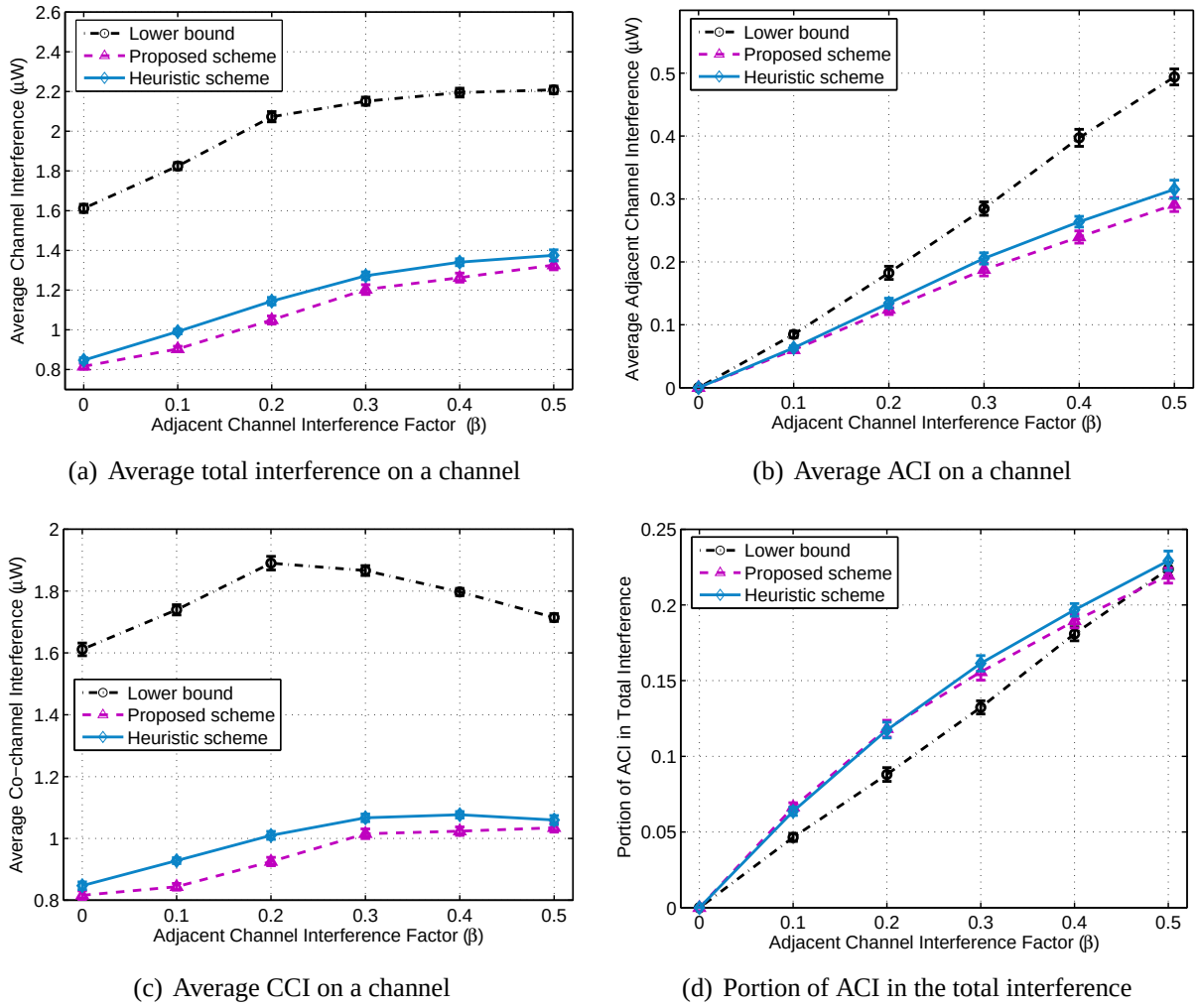


Figure 2.17: CR network throughput versus the ACI factor.

channel is shown in Fig. 2.18(a), which consists of both ACI and CCI. It can be seen that the total interference increases as β gets larger, since there is more power leakage from adjacent channels. The ACI and CCI components are plotted in Figs. 2.18(b) and 2.18(c), respectively. It can be seen that ACI almost linearly increases with β. When $\beta = 0$, ACI is zero for all the three schemes since there is no power leakage from neighboring channels. When $\beta = 0.5$, the ACI of the proposed distributed scheme is about 92.32% of that of the simple heuristic and 58.93% of that of SF. The proposed distributed algorithm curve has the lowest slop among the three schemes, indicating more effective control of ACI as β increases. The fractions of ACI in the total average interference are plotted in Fig. 2.18(d) for the three schemes. The fraction increases as β gets larger, from 0% to about 22%. Clearly ACI should be considered in the resource allocation and protocol design of CR networks.

Finally, we validate our proposed spectrum sensing and access scheme. We set the maximum allowable collision probability γ to be 0.2 and increase channel utilization η from 0.3 to 0.7 in steps of 0.1. In Fig. 2.19, the measured collision rates with primary uses are plotted, along with the $\gamma = 0.2$ curve. It can be seen that the measured collision rate is always kept below γ, showing that



(a) Average total interference on a channel

(b) Average ACI on a channel

(c) Average CCI on a channel

(d) Portion of ACI in the total interference

Figure 2.18: Composition of the total interference measured in the simulations as a function of ACI factor β.

the proposed spectrum sensing and access scheme is quite effective with regard to primary user protection.

2.6 Conclusions

In this chapter, we first studied the problem of design and analysis of MAC protocols for CR networks in this chapter. In particular, we proposed and analyzed two opportunistic multi-channel MAC protocols, adopting a memoryless sensing policy and an improved sensing policy, respectively. The impact of imperfect sensing (in the forms of miss detection and false alarm)

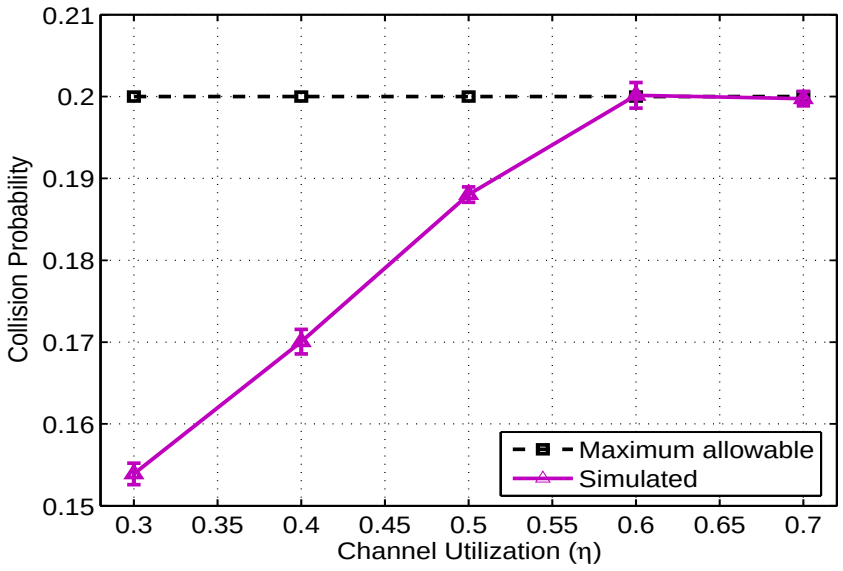


Figure 2.19: Interference.

are explicitly considered in the CR MAC design. We developed analytical models to evaluate the performance of the proposed protocols. Our simulation study demonstrates the accuracy of the analysis, as well as the superior throughput performance of the proposed CR MACs over existing approaches.

Then, we investigated the problem of CCI and ACI mitigation via channel assignment and power allocation in CR networks. The objective was to maximize the total CR network throughput while keeping both collision rate and interference with primary users below tolerance thresholds. We proposed an RLT-based centralized SF algorithm that computes near-optimal solutions, and a distributed greedy algorithm that only uses local channel gain information. The proposed algorithms are evaluated with simulations. The distributed greedy algorithm is shown to outperform both the centralized SF algorithm and a centralized heuristic algorithm with considerable gains.

Chapter 3

Video over CR Networks

3.1 Introduction

Video content delivery over wireless networks is expected to grow drastically in the coming years. The compelling need for ubiquitous video content access will significantly stress the capacity of existing and future wireless networks. To meet this critical demand, the Cognitive Radio (CR) technology provides an effective solution that can effectively exploit co-deployed networks and aggregate underutilized spectrum for future video-aware wireless networks.

The high potential of CRs has attracted substantial interest. The mainstream CR research has focused on developing effective spectrum sensing and access techniques (eg., see [14, 15]). Although considerable advances have been achieved, the important problem of guaranteeing application performance has not been well studied. We find video streaming can make excellent use of the enhanced spectrum efficiency in CR networks. Unlike data, where each bit should be delivered, video is loss-tolerant and rate-adaptive [49, 50]. They are highly suited for CR networks, where the available bandwidth depends on primary user transmission behavior. Graceful degradation of video quality can be achieved as spectrum opportunities evolve over time.

CR is an evolving concept with various network models and levels of cognitive functionality [14, 15]. IEEE 802.22 Wireless Regional Area Networks (WRAN) is the first CR standard for reforming broadcast TV bands, where a base station (BS) controls medium access for customer-premises equipments (CPEs) [51]. Therefore, we first consider multicasting scalable videos in such an infrastructure-based CR network. The spectrum consists of multiple channels, each allocated to a primary network. The CR network is co-located with the primary networks, where a CR BS seeks spectrum opportunities for multicasting multiple video streams, each to a group of secondary subscribers. The problem is to exploit spectrum opportunities for minimizing video

distortion, while keeping the collision rate with primary users below a prescribed threshold. We consider scalable video coding, such as fine-grained-scalability (FGS) and medium grain scalable (MGS) videos [52, 53]. We model the problem of CR video multicast over the licensed channels as a mixed integer nonlinear programming (MINLP) problem, and then develop a sequential fixing algorithm and a greedy algorithm to solve the MINLP, while the latter has a low computational complexity and a proved optimality gap [8].

We then tackle the problem of video over multi-hop CR networks, e.g., a wireless mesh network with CR-enabled nodes. This problem is more challenging than the problem above due to the lack of infrastructure support. We assume each secondary user is equipped with two transceivers. To model and guarantee end-to-end video performance, we adopt the amplify-and-forward approach for video data transmission, which is well-studied in the context of cooperative communications [54]. This is equivalent to setting up a "virtual tunnel" through a multi-hop multi-channel path. The challenging problem, however, is how to set up the virtual tunnels, while the available channels at each relay evolve over time due to primary user transmissions. The formulated MINLP problem is first solved using a centralized sequential fixing algorithm, which provides upper and lower bounds for the achievable video quality. We then apply dual decomposition to develop a distributed algorithm and prove its optimality as well as the convergence condition [9].

The rest of the chapter is organized as follows. We review related work in Section 3.2 and present preliminaries in Section 3.3. We examine video over infrastructure-based CR networks in Section 3.4 and over multi-hop CR networks in Section 3.5. We concludes the chapter in Section 3.6 with a discussion of open problems.

3.2 Background and Related Work

The high potential of CRs has attracted considerable interest form both industry, government and academia [1, 14]. The mainstream CR research has been focused on spectrum sensing and dynamic spectrum access issues. For example, the impact of spectrum sensing errors on the design of spectrum access schemes has been addressed in several papers [6, 24, 55, 56]. The approach

of iteratively sensing a selected subset of available channels has been developed in the design of CR MAC protocols [6, 20, 57]. The optimal trade-off between the two kinds of sensing errors is investigated comprehensively and addressed in depth in [24].

The important issue of QoS provisioning in CR networks has been studied only in a few papers [20, 58], where the objective is still focused on the so-called "network-centric" metrics such as maximum throughput and delay [20, 55]. In [55], an interesting delay throughput trade-off for a multi-cell cognitive radio network is derived, while the goal of primary user protection is achieved by stabilizing a virtual "collision queue". In [58], a game-theoretic framework is described for resource allocation for multimedia transmissions in spectrum agile wireless networks. In this interesting work, each wireless station participates in a resource management game, which is coordinated by a network moderator. A mechanism-based resource management scheme determines the amount of transmission opportunities to be allocated to various users on different frequency bands such that certain global system metrics are optimized.

The problem of video over CR networks has been addressed only in a few recent papers. In [59], a priority virtual queue model is adopted for wireless CR users to select channel and maximize video qualities. In [60], the impact of system parameters residing in different network layers are jointly considered to achieve the best possible video quality for CR users. The problem is formulated as a Mini-Max problem and solved with a dynamic programming approach. In [61], Ali and Yu jointly optimize video parameter with spectrum sensing and access strategy. A rate-distortion model is adopted to optimize the intra-mode selection and source-channel rate with a partially observable Markov decision process (POMDP) formulation. In [62], video encoding rate, power control, relay selection and channel allocation are jointly considered for video over cooperative CR networks. The problem is formulated as a mixed-integer nonlinear problem and solved by a solution algorithm based on a combination of the branch and bound framework and convex relaxation techniques.

Video multicast, as one of the most important multimedia services, has attracted considerable interest from the research community. Layered video multicast has been researched in the mobile

ad hoc networks [63, 64] and infrastructure-based wireless networks [52, 65]. A greedy algorithm is presented in [65] for layered video multicast in WiMAX networks with a proven optimality gap.

A few recent works [47, 66, 67] have studied multi-hop CR networks. The authors formulate cross-layer optimization problem considering factors from the PHY up to the transport layer. The dual decomposition technique [68, 69] is adopted to develop distributed algorithm. We choose similar methodology in our work and apply it to the more challenging problem of real-time video streaming.

3.3 System Model and Preliminaries

3.3.1 Primary Network

We consider a spectrum band consisting of M orthogonal channels with identical bandwidth [21]. We assume that the M channels are allocated to K primary networks, which cover different service areas. A primary network can use any of the M channels without interfering with other primary networks. We further assume that the primary systems use a synchronous slot structure as in prior work [14, 20]. Due to primary user transmissions, the occupancy of each channel evolves following a discrete-time Markov process, as validated by recent measurement studies [14, 16, 20].

In primary network k, the status of channel m in time slot t is denoted by $S_m^k(t)$ with idle (i.e., $S_m^k(t) = 0$) and busy (i.e., $S_m^k(t) = 1$) states. Let λ_m^k and μ_m^k be the transition probability of remaining in state 0 and that from state 1 to 0, respectively, for channel m in primary network k. As discussed in Section 2.3.1, the utilization of channel m in primary network k, denoted by $\eta_m^k = \Pr(S_m^k = 1)$, is

$$\eta_m^k = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} S_m^k(t) = \frac{1-\lambda_m^k}{1-\lambda_m^k+\mu_m^k}. \tag{3.1}$$

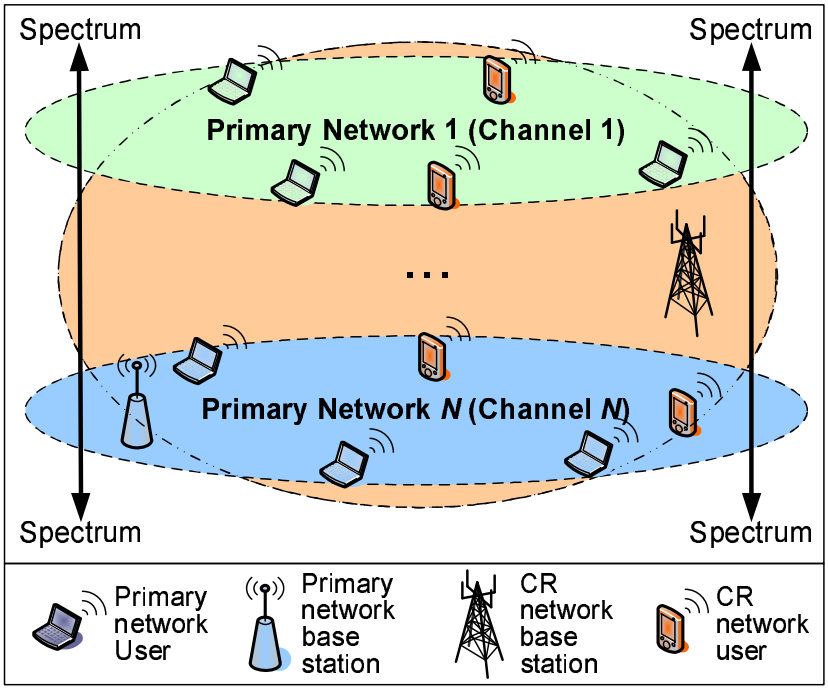


Figure 3.1: An infrastructure-based CR network collocated with N primary networks.

Note that in infrastructure-based CR networks and cooperative CR networks, we assume there is only one $K = 1$ primary network. In infrastructure-based CR networks introduced in the section 3.4, we adopt N as the number of licensed channels since M is denoted as the number of modulation-coding schemes.

3.3.2 Infrastructure-based CR Networks

As shown in Fig. 3.1, we consider a CR base station multicasts G real-time videos to G multicast groups, each of which have N_g users, $g = 1, 2, \cdots, G$. The base station seeks spectrum opportunities in the N channels to serve CR users. In each time slot t, the base station selects a set of channels $\mathcal{A}_1(t)$ to sense and a set of channels $\mathcal{A}_2(t)$ to access. Without loss of generality, the base station has $|\mathcal{A}_1(t)|$ transceivers such that it can sense $|\mathcal{A}_1(t)|$ channels simultaneously. Note that a time slot and channel combination, termed a *tile*, is the minimum unit for resource allocation.

We adopt the same time-slot structure as in [14, 57]. , which is illustrated in Fig. 2.4. At the beginning of each time slot, the base station senses channels in $\mathcal{A}_1(t)$ and then chooses a set

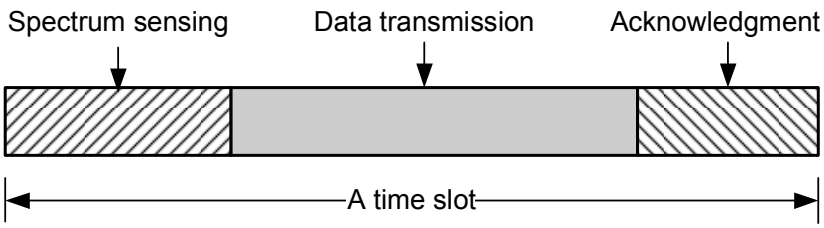


Figure 3.2: The structure of a time slot.

of available channels for opportunistic transmissions based on sensing results. After a successful transmission, the base station will receive an ACK from the user with the highest SNR in the target multicast group. Without loss of generality, we assume that each CR network user can access all the available channels with the channel bonding/aggregation techniques [44, 70].

3.3.3 Multi-hop CR networks

As shown in Fig. 3.3, we also consider a multi-hop CR network that is co-located with the primary networks, within which $\mathcal{S}$ real-time videos are streamed among N CR nodes. Let $\mathcal{U}^k$ denote the set of CR nodes that are located within the coverage of primary network k. A video session l may be relayed by multiple CR nodes if source z_l is not a one-hop neighbor of destination d_l. We assume a *common control channel* for the CR network [20]. We also assume the timescale of the primary channel process (or, the time slot durations) is much larger than the broadcast delays on the control channel, such that feedbacks of channel information can be received at the source nodes in a timely manner.

The time slot structure is the same as that in infrastructure-based CR networks. In the sensing phase, one transceiver of a CR node is used to sense one of the M channels, while the other is tuned to the control channel to exchange channel information with other CR users. Each video source computes the optimal path selection and channel scheduling based on sensing results. In the transmission phase, the channels assigned to a video session l at each link along the path form a virtual "tunnel" connecting source z_l and destination d_l. As illustrated in Fig. 3.4, each node can use one or more than one channels to communicate with other nodes using the channel bonding/aggregation techniques [44, 70]. When multiple channels are available on all the links

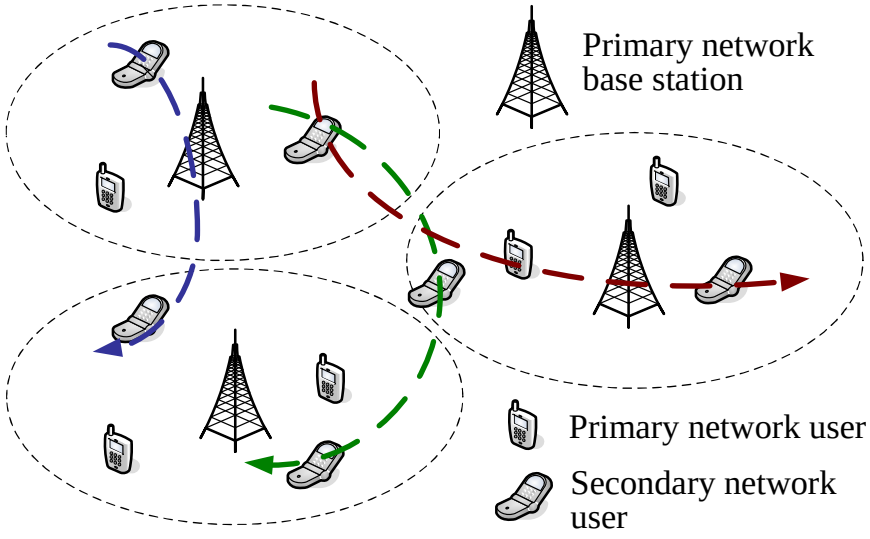


Figure 3.3: Illustration of the multi-hop video CR network architecture.

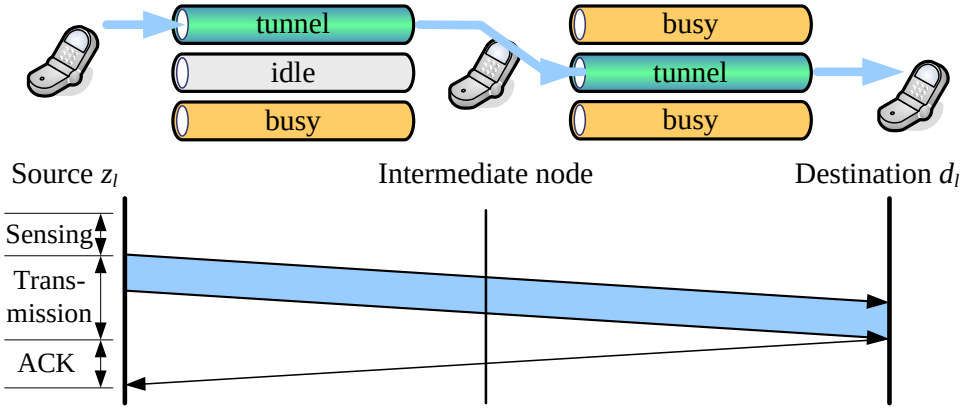


Figure 3.4: The cut-through switching model for video data.

along a path, multiple tunnels can be established and used simultaneously for a video session. In the acknowledgment phase, the destination sends ACK to the source for successfully received video packets through the same tunnel.

We adopt amplify-and-forward for video transmission [54]. During the transmission phase, one transceiver of the relay node receives video data from the upstream node on one channel, while the other transceiver of the relay node amplifies and forwards the data to the downstream node on a different, orthogonal channel. There is no need to store video packets at the relay nodes. Error detection/correction will be performed at the destination node. As a result, we can transmit through the tunnel a block of video data with minimum delay and jitter in one time slot.

3.3.4 Spectrum Sensing

As discussed in Section 2.3.3, two types of sensing errors may occur during the sensing process. A *false alarm* may lead to waste a spectrum opportunity and a *miss detection* may causes collision with primary users. In a multi-hop CR network, the sensing results from various users may be different. Denote H_0 as the hypothesis that channel m in primary network k is idle, and H_1 the hypothesis that channel m in primary network k is busy in time slot t. The conditional probability that channel m is available in primary network k, denoted by $a_m^k(t)$, can be derived as [8],

$$\begin{aligned} a_m^k(t) &= \Pr(H_0|W_i^m = \theta_i^m,\ i \in \mathcal{U}_m^k, \pi_m^k) \\ &= \left[1 + \left(\varphi_m^k\right)^{u_m^k} \left(\phi_m^k\right)^{|\mathcal{U}_m^k| - u_m^k} \frac{\Pr(H_1|\pi_m^k)}{\Pr(H_0|\pi_m^k)}\right]^{-1}. \end{aligned} \tag{3.2}$$

where θ_i^m represents a specific sensing result (0 or 1), $\mathcal{U}_m^k$ is the subset of users in $\mathcal{U}^k$ (i.e., the set of CR nodes that are located within the coverage of primary network k) that sense channel m, u_m^k is the number of users in $\mathcal{U}_m^k$ observing channel m is idle, π_m^k represents the history of channel m in primary network k, and φ_m^k and ϕ_m^k are defined as:

$$\begin{cases} \varphi_m^k = \frac{P(W_i^m=0|H_1)}{P(W_i^m=0|H_0)} = \frac{\delta_m}{1-\epsilon_m}, & \text{when } \theta_i^m = 0 \\ \phi_m^k = \frac{P(W_i^m=1|H_1)}{P(W_i^m=1|H_0)} = \frac{1-\delta_m}{\epsilon_m}, & \text{when } \theta_i^m = 1. \end{cases} \tag{3.3}$$

Based on the Markov chain channel model, we have (3.4), which can be recursively expanded:

$$\begin{cases} \Pr(H_0|\pi_m^k) = \lambda_m^k a_m^k(t-1) + \mu_m^k \left[1 - a_m^k(t-1)\right] \\ \Pr(H_1|\pi_m^k) = 1 - \Pr(H_0|\pi_m^k). \end{cases} \tag{3.4}$$

3.3.5 Video Performance Measure

Both FGS and MGS videos are highly suited for dynamic CR networks. With FGS or MGS coding, each video l is encoded into one base layer with rate R_l^b and one enhancement layer with rate R_l^e. The total bit rate for video l is $R_l = R_l^b + R_l^e$.

We consider peak-signal-noise-ratio (PSNR) (in dB) of reconstructed videos. As in prior work [8, 52], the average PSNR of video l, denoted as Q_l, can be estimated as:

$$Q_l(R_l) = Q_l^b + \beta_l(R_l - R_l^b) = Q_l^0 + \beta_l R_l, \tag{3.5}$$

where Q_l^b is the resulting PSNR when the base layer is decoded alone, β_l a constant depending on the video sequence and codec setting, and $Q_l^0 = Q_l^b - \beta_l R_l^b$. We verified the model (3.5) with several test video sequences using the MPEG-4 FGS codec and the H.264/SVC MGS codec and found it is highly accurate.

Due to the real-time nature, we assume that each *group of pictures* (GOP) must be delivered during the next GOP window, which consists of N_G time slots. Beyond that, overdue data from the current GOP will be useless and will be discarded. In infrastructure-based network, G video stream are multicast to G groups of CR user, so we choose the group index g instead of video session index l.

3.4 Video over Infrastructure Based CR Networks

In this section, we examine the problem of video over infrastructure-based CR networks. We consider cross-layer design factors such as scalable video coding, spectrum sensing, opportunistic spectrum access, primary user protection, scheduling, error control and modulation. We propose efficient optimization and scheduling algorithms for highly competitive solutions, and prove the complexity and optimality bound of the proposed greedy algorithm.

3.4.1 Network Model

Spectrum Access

At the beginning of each time slot t, the CR BS senses the M channels and compute $a_n(t)$ for each channel n. Based on spectrum sensing results, the base station determines which channels to access for video streaming. We adopt an opportunistic spectrum access approach, aiming to exploit unused spectrum while probabilistically bounding the interference to primary users.

Let $\gamma_n \in (0,1)$ be the *maximum allowed collision probability* with primary users on channel n, and $p_n^{tr}(t)$ the *transmission probability* on channel n for the base station in time slot t. The probability of collision caused by the base station should be kept below γ_n, i.e., $p_n^{tr}(t)\,[1-a_n(t)] \leq \gamma_n$. In addition to primary user protection, another important objective is to exploit unused spectrum as much as possible. The transmission probability can be determined by jointly considering both objectives, as

$$p_n^{tr}(t) = \begin{cases} \min\left\{1, \frac{\gamma_n}{1-a_n(t)}\right\}, & \text{if } 0 \leq a_n(t) < 1 \\ 1, & \text{if } a_n(t) = 1. \end{cases} \tag{3.6}$$

If $p_n^{tr}(t) = 1$, channel n will be accessed deterministically. If $p_n^{tr}(t) = \gamma_n/[1-a_n(t)] < 1$, channel n will be accessed opportunistically with probability $p_n^{tr}(t)$.

Modulation-Coding Schemes

At the PHY layer, we consider various modulation and channel coding combination schemes. Without loss of generality, we assume several choices of modulation schemes, such as QPSK, 16-QAM and 64-QAM, combined with several choices of forward error correction (FEC) schemes, e.g., with rates 1/2, 2/3, and 3/4. We consider M unique combinations of modulation and FEC schemes, termed *Modulation-Coding* (MC) schemes, in this paper.

Under the same channel condition, different MC schemes will achieve different data rates and symbol error rates. Adaptive modulation and channel coding allow us to exploit user channel variations to maximize video data rate under a given residual bit error rate constraint. When a user

has a good channel, it should adopt an MC scheme that can support a higher data rate. Conversely, it should adopt a low-rate MC scheme when the channel condition is poor. Let $\{MC_m\}_{m=1,\cdots,M}$ be the list of available MC schemes indexed according to their data rates in the increasing order. We assume slow fading channels with coherence time larger than a time slot. Each CR user measures its own channel and feedbacks measurements to the base station when its channel quality changes. At the beginning of a time slot, the base station is able to collect the number $n_{g,m}$ of users in each multicast group g who can successfully decode MC_m signals for $m = 1, 2, \cdots, M$.

Since the base layer carries the most important data, the most reliable MC scheme $MC_{b(g)}$ should be used, where $b(g) = \max_i\{i : n_{g,i} = N_g\}$, for all g. Without loss of generality, we assume that the base layer is always transmitted using MC_1. If a user's channel is so poor that it cannot decode the MC_1 signal, we consider it disconnected from the CR network. We further divide the enhancement layer into M sub-layers, where sub-layer m has rate $R^e_{g,m}$ and uses MC_m. Assuming that MC_m can carry $b_{g,m}$ bits of video g in one tile, we denote the number of tiles for sub-layer m of video g as $l_{g,m} \geq 0$. We have

$$R^e_g = \sum_{m=1}^{M} R^e_{g,m} = \sum_{m=1}^{M} b_{g,m} l_{g,m}. \tag{3.7}$$

Proportional Fair Allocation

Since we consider video quality in this paper, we define the utility for user i in group g as $U_{g,i} = \log Q_{g,i} = \log\left(Q^b_g + \beta_g R^e_g(i)\right)$, where $R^e_g(i)$ is the received enhancement layer rate of user i in group g.

The total utility for group g is $U_g = \sum_{i=1}^{N_g} U_{g,i}$. Intuitively, a lower layer should use a lower (i.e., more reliable) MC scheme. This is because if a lower layer is lost, a higher layer cannot be used at the decoder even if it is correctly received. Considering the user classification based on their MC schemes, we can rewrite U_g as follows [65]:

$$U_g = \sum_{k=1}^{M} (n_{g,k} - n_{g,k+1}) \log\left(Q^b_g + \beta_g \sum_{m=1}^{k} R^e_{g,m}\right), \tag{3.8}$$

where $n_{g,M+1} = 0$. The utility function of the entire CR video multicast system is

$$U = \sum_{g=1}^{G} U_g. \tag{3.9}$$

Maximizing U will achieve *proportional fairness* among the video sessions [71]

3.4.2 Optimized Video Multicast in CR Networks

Outline of the Proposed Approach

As discussed, the CR video multicast problem is highly challenging since a lot of design choices are tightly coupled. First, as users see different channels, such heterogeneity should be accommodated so that a user can receive a video quality commensurate to its channel quality. Second, we need to determine the video rates before transmission, which, however, depend on future channel evolution and choice of MC schemes. Third, the trade-off between primary user protection and spectrum utilization should guide the scheduling of video packets to channels. Finally, all the optimization decisions should be made in real-time. Low-complexity, but efficient algorithms are needed, while theoretical optimality bounds would be highly appealing.

To address heterogeneous user channels, we adopt FGS to produce a base layer with rate R_g^b and an enhancement layer with rate $\bar{R}_g^e$. Without loss of generality, we assume R_g^b is prescribed for an acceptable video quality, while $\bar{R}_g^e$ is set to a large value that is allowed by the codec. During transmission, we determine the *effective rate* for each enhancement layer $R_g^e \leq \bar{R}_g^e$ depending on channel availability, sensing, and MC schemes.[1] The optimal partition of the enhancement layer should be determined such that each sub-layer uses a different MC scheme.

We determine the optimal partition of enhancement layers, the choices of MC schemes, and video packet scheduling as follows. First, we solve the optimal partition problem for every GoP based on an estimated (i.e., average) number of available tiles T_e in the next GoP window that can be used for the enhancement layer, using algorithm GRD1 with complexity $O(MGT_e)$. The

[1]The proposed approach can also be used for streaming stored FGS video.

tile allocations are then dynamically adjusted in each time slot according to more recent (and thus more accurate) channel status using algorithm GRD2, with complexity $O(MGK)$, where $K \ll T_e$. Second, during each time slot, video packets are scheduled to the available channels such that the overall system utility is maximized. The TSA algorithm has complexity $O(N \log N)$. Both GRD2 and TSA have low complexity and are suitable for execution in each time slot.

In real-time video, overdue packets generally do not contribute to improving the received quality. We assume that the data from a GoP should be be delivered in the next GoP window consisting of T_{GoP} time slots.[2] Since the base layer is essential for decoding a video, we assume that the base layers of all the videos are coded using MC_1. For the M sub-layers of the enhancement layer, a more important sub-layer will be coded using a more reliable (i.e., lower rate) MC scheme. At the beginning of each GoP window, all the base layers are transmitted using the available tiles. *Retransmissions* will be scheduled if no ACK is received for a base layer packet. After the base layers are transmitted, we allocate the remaining available tiles in the GoP window for the enhancement layer. The same rule applies to the enhancement sub-layers, such that a higher sub-layer will be transmitted if and only if all the lower sub-layers are acknowledged. This is due to the decoding dependency of layered video.

In each time slot t, the base station opportunistically access every channel n with probability $p_n^{tr}(t)$ given in (3.6). Specifically, for each channel n, the base station generates a random number $x_n(t)$, which is independent of the channel history $\theta_n(t)$ and uniformly distributed in [0,1]. If $x_n(t) \leq p_n^{tr}(t)$, the most important packet among those not ACKed in the previous GoP will be transmitted on channel n. If an ACK is received for this packet at the end of time slot t, this packet is successfully received by at least one of the users and will be removed from the transmission buffer. Otherwise, there is a collision with primary user and this packet will remain in the transmission buffer and will be retransmitted.

In the following, we describe in detail the three algorithms.

[2]The proposed approach also works for the more general delay requirements that are multiple GoP windows.

Enhancement Layer Partitioning and Tile Allocation

As a first step, we need to determine the effective rate for each enhancement layer $R_g^e \leq \bar{R}_g^e$. We also need to determine the optimal partition of each enhancement layer. Clearly, the solutions will be highly dependent on the channel availability processes and sensing results.

Recall that the base layers are transmitted using MC_1 first in each GoP window. The *remaining* available tiles can then be allocated to the enhancement layers. We assume that the number of tiles used for the enhancement layers in a GoP window, T_e, is known at the beginning of the GoP window. For example, we can estimate T_e by computing the total average "idle" intervals of all the N channels based on the channel model, decreased by the number of tiles used for the base layers (i.e., $R_g^b/b_{g,1}$). We then split the enhancement layer of each video g into M sub-layers, each occupying $l_{g,m}$ tiles when coded with MC_m, $m = 1, 2, \cdots, M$.

Letting $\vec{l} = [l_{1,1}, l_{1,2}, \cdots, l_{1,M}, l_{2,1}, \cdots l_{G,M}]$ denote the *tile allocation vector*, we formulate an optimization problem OPT-Part as follows.

$$\text{maximize: } U(\vec{l}) = \sum_{g=1}^{G}\sum_{k=1}^{M}(n_{g,k} - n_{g,k+1}) \times \log\left[Q_g^b + \beta_g \sum_{m=1}^{k} b_{g,m} l_{g,m}\right] \tag{3.10}$$

$$\text{subject to: } \sum_{g=1}^{G}\sum_{m=1}^{M} l_{g,m} \leq T_e \tag{3.11}$$

$$\sum_{m=1}^{M} b_{g,m} l_{g,m} \leq \bar{R}_g^e,\ g \in [1, \cdots, M] \tag{3.12}$$

$$l_{g,m} \geq 0,\ m \in [1, \cdots, M], g \in [1, \cdots, G]. \tag{3.13}$$

OPT-Part is solved at the beginning of each GoP window to determine the optimal partition of the enhancement layer. The objective is to maximize the overall system utility by choosing optimal values for the $l_{g,m}$'s. We can derive the effective video rates as $R_g^e = \sum_{m=1}^{M} b_{g,m} l_{g,m}$. The formulated problem is a MINLP problem, which is NP-hard [65]. In the following, we present two algorithms for computing near-optimal solutions to problem OPT-Part: (i) a *sequential fixing* (SF)

algorithm based on a linear relaxation of (3.10), and (ii) a *greedy algorithm* GRD1 with proven optimality gap.

A Sequential Fixing Algorithm With this algorithm, the original MINLP is first linearized to obtain a linear programming (LP) relaxation. Then we iteratively solve the LP, while fixing one integer variable in every iteration [47,72]. We use the *Reformulation-Linearization Technique* (RLT) to obtain the LP relaxation [45]. RLT is a technique that can be used to produce LP relaxations for a nonlinear, nonconvex polynomial programming problem. This relaxation will provide a tight upper bound for a maximization problem. Specifically, we linearize the logarithm function in (3.10) over some suitable, tightly-bounded interval using a polyhedral outer approximation comprised of a convex envelope in concert with several tangential supports. We further relax the integer constraints, i.e., allowing the $l_{g,m}$'s to take fractional values. Then we obtain an upper-bounding LP relaxation that can be solved in polynomial time. Due to lack of space, we refer interested readers to [45] for a detailed description of the technique.

We next solve the LP relaxation iteratively. During each iteration, we find the $l_{\hat{g},\hat{m}}$ which has the minimum value for $(\lceil l_{\hat{g},\hat{m}} \rceil - l_{\hat{g},\hat{m}})$ or $(l_{\hat{g},\hat{m}} - \lfloor l_{\hat{g},\hat{m}} \rfloor)$ among all fractional $l_{g,m}$'s, and round it up or down to the nearest integer. We next reformulate and solve a new LP with $l_{\hat{g},\hat{m}}$ fixed. This procedure repeats until all the $l_{g,m}$'s are fixed. The complete SF algorithm is given in Table 3.1. The complexity of SF depends on the specific LP algorithm (e.g., the *simplex method* with polynomial-time average-case complexity).

A Greedy Algorithm Although SF can compute a near-optimal solution in polynomial time, it does not provide any guarantee on the optimality of the solution. In the following, we describe a greedy algorithm, termed GRD1, which exploits the inherent priority structure of layered video and MC schemes and has a proven optimality bound.

The complete greedy algorithm is given in Table 3.2, where $R = \sum_{g=1}^{G} \bar{R}_g^e$ is the total rate of all the enhancement layers and $\vec{e}_i$ is a *unit vector* with "1" at the i-th location and "0" at all other locations. In GRD1, all the $l_{g,m}$'s are initially set to 0. During each iteration, one tile is

Table 3.1: The Sequential Fixing (SF) Algorithm

1:	Use RLT to linearize the original problem
2:	Solved the LP relaxation
3:	Suppose $l_{\hat{g},\hat{m}}$ is the integer variable with the minimum $(\lceil l_{\hat{g},\hat{m}} \rceil - l_{\hat{g},\hat{m}})$ or $(l_{\hat{g},\hat{m}} - \lfloor l_{\hat{g},\hat{m}} \rfloor)$ value among all $l_{g,m}$ variables that remain to be fixed, round it up or down to the nearest integer
4:	If all $l_{g,m}$'s are fixed, got to Step 6
5:	Otherwise, reformulate a new relaxed LP with the newly fixed $l_{g,m}$ variables, and go to Step 2
6:	Output all fixed $l_{g,m}$ variables and $R_g^e = \sum_{m=1}^{M} b_{g,m} l_{g,m}$

Table 3.2: The Greedy Algorithm (GRD1)

1:	Initialize $l_{g,m} = 0$ for all g and m
2:	Initialize $A = \{1, 2, \cdots, G\}$
3:	WHILE $\left(\sum_{g=1}^{G} \sum_{m=1}^{M} l_{g,m} \leq T_e \text{ and } A \text{ is not empty}\right)$
4:	Find $l_{\hat{g},\hat{m}}$ that can be increased by one: $\vec{e}_{\hat{g},\hat{m}} = \arg\max_{g \in A, m \in [1,\cdots,M]} \left\{ \frac{U(\vec{l}+\vec{e}_{g,m}) - U(\vec{l})}{b_{g,m} + R/T_e} \right\}$
5:	$\vec{l} = \vec{l} + \vec{e}_{\hat{g},\hat{m}}$
6:	IF $\left(\sum_{m} b_{\hat{g},m} l_{\hat{g},m} > \bar{R}_g^e\right)$
7:	$\vec{l} = \vec{l} - \vec{e}_{\hat{g},\hat{m}}$
8:	Delete $\hat{g}$ from A
9:	END IF
10:	END WHILE

allocated to the $\hat{m}$-th sub-layer of video $\hat{g}$. In Step 4, $l_{\hat{m},\hat{g}}$ is chosen to be the one that achieves the largest increase in terms of the "normalized" utility (i.e., $[U(\vec{l}+\vec{e}_{g,m}) - U(\vec{l})]/[b_{g,m} + R/T_e]$) if it is assigned with an additional tile. Lines 6, 7, and 8 check if the assigned rate exceeds the maximum rate $\bar{R}_g^e$. GRD1 terminates when either all the available tiles are used or when all the video data are allocated with tiles. In the latter case, all the videos are transmitted at full rates. We have the following Theorem for GRD1.

Theorem 3.1. *The greedy algorithm GRD1 shown in Table 3.2 has a complexity $O(MGT_e)$. It guarantees a solution that is within a factor of $(1 - e^{-1/2})$ of the global optimal solution.*

Proof. (i) *Complexity*: In Step 4 in Table 3.2, it takes $O(MG)$ to solve for $\vec{e}_{\hat{g},\hat{m}}$. Since each iteration assigns one tile to sub-layer $\hat{m}$ of group $\hat{g}$, it takes T_e iterations to allocate all the available tiles in a GoP window. Therefore, the overall complexity of GRD1 is $O(MGT_e)$.

(ii) *Optimality Bound*: This proof is extended from a result first shown in [65] for layered videos. We first show a property of group utility $U_g(\vec{l})$, which will be used in the proof of the optimality gap. For two vectors $\vec{l}_g^1$ and $\vec{l}_g^2$, letting $\Delta = U_g(\vec{l}_g^1) - U_g(\vec{l}_g^2)$, we have

$$\begin{aligned}
\Delta &= \sum_{k=1}^{M}(n_{g,k} - n_{g,k+1}) \times \log\left(1 + \frac{\sum_{m=1}^{k}\beta_g b_{g,m}(l_{g,m}^1 - l_{g,m}^2)}{Q_g^b + \sum_{m=1}^{k}\beta_g b_{g,m} l_{g,m}^2}\right) \\
&\le \sum_{k=1}^{M}\sum_{m=1}^{k}(l_{g,m}^1 - l_{g,m}^2)^+(n_{g,k} - n_{g,k+1}) \times \log\left(1 + \beta_g b_{g,m} / \left[Q_g^b + \sum_{m=1}^{k}\beta_g b_{g,m} l_{g,m}^2\right]\right) \\
&\le \sum_{k=1}^{M}\sum_{m=1}^{M}(l_{g,m}^1 - l_{g,m}^2)^+(n_{g,k} - n_{g,k+1}) \times \log\left(1 + \beta_g b_{g,m} / \left[Q_g^b + \sum_{m=1}^{k}\beta_g b_{g,m} l_{g,m}^2\right]\right) \\
&= \sum_{m=1}^{M}(l_{g,m}^1 - l_{g,m}^2)^+ \left[U_g(\vec{l}_g^2 + b_{g,m}) - U(\vec{l}_g^2)\right], \qquad (3.14)
\end{aligned}$$

where $y^+ = \max\{0, y\}$. The first inequality is due to the concavity of logarithm functions.

Next we prove the optimality bound. Let $\vec{l}_t$ be the output of GRD1 after t iterations. Let the utility gap between the optimal solution and the GRD1 solution be $F_t = U(\vec{l}^*) - U(\vec{l}_t)$, and $\vec{e}_{\hat{g},\hat{m}}(t)$ the argument found in Step 4 of GRD1 after t iterations. We have $\vec{l}_t = \vec{l}_{t-1} + \vec{e}_{\hat{g},\hat{m}}(t)$ and

$$\begin{aligned}
& F_{t-1} = U(\vec{l}^*) - U(\vec{l}_{t-1}) \\
\le & \sum_g \sum_m (l_{g,m}^* - l_{g,m})^+ [U(\vec{l}_{t-1} + \vec{e}_{g,m}(t)) - U(\vec{l}_{t-1})] \\
\le & \sum_g \sum_m (l_{g,m}^* - l_{g,m})^+ [U(\vec{l}_{t-1} + \vec{e}_{\hat{g},\hat{m}}(t)) - U(\vec{l}_{t-1})] \frac{b_{g,m} + R/T_e}{b_{\hat{g},\hat{m}}(t) + R/T_e} \\
\le & \frac{U(\vec{l}_t) - U(\vec{l}_{t-1})}{b_{\hat{g},\hat{m}}(t) + R/T_e} \sum_g \sum_m [l_{g,m}^*(b_{g,m} + R/T_e)].
\end{aligned}$$

The first inequality is due to (3.14) and the second inequality follows Step 4 of GRD1. It follows (3.11) that $\sum_g \sum_m l^*_{g,m} \leq T_e$ and $\sum_g \sum_m b_{g,m} l^*_{g,m} \leq R$. We have $F_{t-1} \leq (F_{t-1} - F_t) \frac{2R}{b_{\hat{g},\hat{m}}(t)+R/T_e}$. Solving for F_t, we have $F_t \leq F_{t-1} \{1 - [b_{\hat{g},\hat{m}}(t) + R/T_e] / (2R)\}$.

Suppose the *WHILE* loop in Table 3.2 has been executed k times when the solution is obtained.

$$\begin{aligned} F_k &\leq F_{k-1} \{1 - [b_{\hat{g},\hat{m}}(k) + R/T_e] / (2R)\} \\ &\leq F_0 \prod_{t=1}^{k} \{1 - [b_{\hat{g},\hat{m}}(t) + R/T_e] / (2R)\} \\ &\leq F_0 \left\{ 1 - 1/(2kR) \sum_{t=1}^{k} [b_{\hat{g},\hat{m}}(t) + R/T_e] \right\}^k . \end{aligned}$$

The *WHILE* loop exits when one or both of two constraints are violated. If $\sum_g \sum_m l_{g,m} \leq T_e$ is violated, there is no tile that can be used. Therefore $k \geq T_e$ and $\sum_{t=1}^{k} R/T_e \geq R$. If constraint "$A$ is not empty" is violated, all the videos have been allocated sufficient number of tiles and will be transmitted at full rates. We have $\sum_{t=1}^{k} b_{\hat{g},\hat{m}}(t) \geq R$ in this case. It follows that

$$\begin{aligned} F_k &\leq F_0 \left\{ 1 - 1/(2kR) \sum_{t=1}^{k} [b_{\hat{g},\hat{m}}(t) + R/T_e] \right\}^k \\ &\leq F_0 \left[1 - 1/(2k)\right]^k \leq F_0 e^{-1/2}. \end{aligned}$$

Since $F_0 = U(\vec{l}^*)$, we have $U(\vec{l}_k) \geq (1 - e^{-1/2}) U(\vec{l}^*)$. Therefore, we conclude that the GRD1 solution is bounded by $(1 - e^{-1/2}) U(\vec{l}^*)$ and $U(\vec{l}^*)$. □

A Refined Greedy Algorithm GRD1 computes $l_{g,m}$'s based on an estimate of network status $\vec{S}(t)$ in the next T_{GoP} time slots. Due to channel dynamics, the computed $l_{g,m}$'s may not be exactly accurate, especially when T_{GoP} is large. We next present a refined greedy algorithm, termed GRD2, which adjusts the $l_{g,m}$'s based on more accurate estimation of the channel status.

GRD2 is executed at the beginning of every time slot. It estimates the number of available tiles $T_e(t)$ in the next T_{est} successive time slots, where $1 \leq T_{est} \leq T_{GoP}$ is a design parameter depending on the coherence time of the channels. Such estimates are more accurate than that in

GRD1 since they are based on recently received ACKs and recent sensing results. Specifically, we estimate $T_e(t)$ using the belief vector $\vec{a}(t)$ in time slot t. Recall that $a_n(t)$'s are computed based on the channel model, feedback, sensing results, and sensing errors, as given in (3.2), and (3.4). For the next time slot, $a_n(t+1)$ can be estimated as $\hat{a}_n(t+1) = \lambda_n a_n(t) + \mu_n[1 - a_n(t)] = (\lambda_n - \mu_n)a_n(t) + \mu_n$. Recursively, we can derive $\hat{a}_n(t+\tau)$ for the next τ time slots.

$$\hat{a}_n(t+\tau) = (\lambda_n - \mu_n)^\tau a_n(t) + \mu_n \frac{1-(\lambda_n-\mu_n)^\tau}{1-(\lambda_n-\mu_n)}. \tag{3.15}$$

At the beginning section of a GoP window, all the base layers will be firstly transmitted. We start the estimation after all the base layers have been successfully received (possibly with retransmissions). The number of available tiles in the following T_{est} time slots can be estimated as $T_e(t) = \sum_{n=1}^{N} \sum_{\tau=0}^{t_{min}} \hat{a}_n(t+\tau)$, where $\hat{a}_n(t+0) = a_n(t)$ and $t_{min} = \min\{T_{est} - 1, T_{GoP} - (t \bmod T_{GoP})\}$. $T_e(t)$ may not be an integer, but it does not affect the outcome of GRD2.

We then adjust the $l_{g,m}$'s based on $T_e(t)$ and $N_{ack}(t)$, the number of ACKs received in time slot t. If $T_e(t) + N_{ack}(t-1) > T_e(t-1) + N_{ack}(t-2)$, there are more tiles that can be allocated and we can increase some of the $l_{g,m}$'s. On the other hand, if $T_e(t) + N_{ack}(t-1) < T_e(t-1) + N_{ack}(t-2)$, we have to reduce some of the $l_{g,m}$'s. Due to layered videos, when we increase the number of allocated tiles, we only need to consider $l_{g,m}$ for $m = m', m'+1, \cdots, M$, where $MC_{m'}$ is the highest MC scheme used in the previous time slot. Similarly, when we reduce the number of allocated tiles, we only need to consider $l_{g,m}$ for $m = m', m'+1, \cdots, M$.

The refined greedy algorithm is given in Table 3.3. For time slot t, the complexity of GRD2 is $O(MGK)$, where $K = |N_{ack}(t-1) - N_{ack}(t-2) + T_e(t) - T_e(t-1)|$. Since $K \ll T_e$, the complexity of GRD2 is much lower than GRD1, suitable for execution in each time slot.

Tile Scheduling in a Time Slot

In each time slot t, we need to schedule the remaining tiles for transmission on the N channels. We define Inc(g, m, i) to be the increase in the group utility function $U(g)$ after the i-th tile in the

Table 3.3: The Refined Greedy Algorithm (GRD2) for Each Time Slot

Initialize $l_{g,m} = 0$ for all g and m
Initialize $A = \{1, 2, \cdots, G\}$
Initialize $N_{ack}(0) = 0$
Estimate $T_e(1)$ based on the Markov Chain channel model
Use GRD1 to find all $l_{g,m}$'s based on $T_e(1)$
WHILE $t = 2$ to T_{GoP}
  Estimate $T_e(t)$
  IF $[T_e(t) + N_{ack}(t-1) < T_e(t-1) + N_{ack}(t-2)]$
    WHILE $\left[\sum_{g=1}^{G}\sum_{m=1}^{M} l_{g,m} > T_e(t) + N_{ack}(t-2)\right]$
      Find $l_{\hat{g},\hat{m}}$ that can be reduced by 1:
        $\vec{e}_{\hat{g},\hat{m}} = \arg\min_{\forall g,m \in \{m',\cdots,M\}} \left\{ \frac{U(\vec{l}) - U(\vec{l} - \vec{e}_{g,m})}{b_{g,m} + R/T_e} \right\}$
      $\vec{l} = \vec{l} - \vec{e}_{\hat{g},\hat{m}}$
      IF $(\hat{g} \notin A)$
        Add $\hat{g}$ to A
      END IF
    END WHILE
  END IF
  IF $[T_e(t) + N_{ack}(t-1) > T_e(t-1) + N_{ack}(t-2)]$
    WHILE $\left[\sum_{g=1}^{G}\sum_{m=1}^{M} l_{g,m} \leq T_e(t) + N_{ack}(t-1)\right.$ and
     $\left.A \text{ is not empty}\right]$
      Find $l_{\hat{g},\hat{m}}$ that can be increased by 1
        $\vec{e}_{\hat{g},\hat{m}} = \arg\max_{g \in A, m \in \{m',\cdots,M\}} \left\{ \frac{U(\vec{l} + \vec{e}_{g,m}) - U(\vec{l})}{b_{g,m} + R/T_e} \right\}$
      $\vec{l} = \vec{l} + \vec{e}_{\hat{g},\hat{m}}$
      IF $\left(\sum_m b_{\hat{g},m} l_{\hat{g},m} > \bar{R}_g^e\right)$
        $\vec{l} = \vec{l} - \vec{e}_{\hat{g},\hat{m}}$
        Delete $\hat{g}$ from A
      END IF
    END WHILE
  END IF
  Update $N_{ack}(t-1)$
END WHILE

sub-layer using MC_m is successfully decoded. It can be shown that

$$\text{Inc}(g,m,i) = \sum_{k=m}^{M} (n_{g,k} - n_{g,k+1}) \times \log\left[1 + \frac{\beta_g b_{g,m}}{Q_g^b + \beta_g \sum_{u=1}^{m-1} b_{g,u} l_{g,u} + (i-1)\beta_g b_{g,m}}\right].$$

$\text{Inc}(g,m,i)$ can be interpreted as the *reward* if the tile is successfully received.

Table 3.4: Algorithm for Tile Scheduling in a Time Slot

1:	Initialize m_g to the lowest MC that has not been ACKed for all g
2:	Initialize i_g to the first packet that has not been ACKed for all g
3:	Sort $\{c_n(t)\}$ in decreasing order. Let the sorted channel list be indexed by j.
4:	While ($j = 1$ to N)
5:	Find the group having the maximum increase in $U(g)$: $\hat{g} = \arg\max_{\forall g} \text{Inc}(g, m_g, i_g)$
6:	Allocate the tile on channel j to group $\hat{g}$
7:	Update $m_{\hat{g}}$ and $i_{\hat{g}}$
8:	End while

Letting $c_n(t)$ be the probability that the tile is successfully received, then we have $c_n(t) = p_n^{tr}(t)a_n(t)$. Our objective of tile scheduling is to maximize the expected reward, i.e.,

$$\text{maximize:} \quad \text{E}[\text{Reward}(\vec{\xi})] = \sum_{n=1}^{N} c_n(t) \cdot \text{Inc}(\xi_n), \tag{3.16}$$

where $\vec{\xi} = \{\xi_n\}_{n=1,\cdots,N}$ and ξ_n is the tile allocation for channel n, i.e., representing the three-tuple $\{g, m, i\}$. The TSA algorithm is shown in Table 3.4, which solves the above optimization problem. The complexity of TSA is $O(N \log N)$. We have the following theorem for TSA.

Theorem 3.2. *E[Reward] is maximized if Inc*$(\xi_i) >$ *Inc*(ξ_j) *when* $c_i(t) > c_j(t)$ *for all* i *and* j.

Proof. Suppose there exists a pair of i and j where $\text{Inc}(\xi_i) > \text{Inc}(\xi_j)$ and $c_i(t) < c_j(t)$. We can further increase E[Reward] by switching the tile assignment, i.e., assign channel i to ξ_j and channel j to ξ_i. With this new assignment, the net increase in E[Reward] is

$$\begin{aligned} & c_j(t)\text{Inc}(\xi_i) + c_i(t)\text{Inc}(\xi_j) - c_i(t)\text{Inc}(\xi_i) - c_j(t)\text{Inc}(\xi_j) \\ = \;& [c_j(t) - c_i(t)][\text{Inc}(\xi_i) - \text{Inc}(\xi_j)] > 0. \end{aligned}$$

Therefore E[Reward] is maximized when the $\{\text{Inc}(\xi_i)\}$ and $\{c_i(t)\}$ are in the same order. □

3.4.3 Simulation Results

We evaluate the performance of the proposed CR video multicast framework using a customized simulator implemented with a combination of C and MATLAB. Specifically, the LPs are solved using the MATLAB Optimization Toolbox and the remaining parts are written in C. For the results reported in this section, we have $N = 12$ channels (unless otherwise specified). The channel parameters λ_n and μ_n are set between $(0, 1)$. The maximum allowed collision probability γ_n is set to 0.2 for all the channels unless otherwise specified.

The CR base station multicasts three Common Intermediate Format (CIF, 352×288) video sequences to three multicast groups, i.e., *Bus* to group 1, *Foreman* to group 2, and *Mother & Daughter* to group 3. The $n_{1,m}$'s are $\{42, 40, 36, 30, 22, 12\}$ (i.e., 42 users can decode MC_1 signal, 40 users can decode MC_2 signal, and so forth); the $n_{2,m}$'s are $\{51, 46, 40, 32, 23, 12\}$ and the $n_{3,m}$'s are $\{49, 44, 40, 32, 24, 13\}$. The number of bits carried in one tile using the MC schemes are 1 kb/s, 1.5 kb/s, 2 kb/s, 3 kb/s, 5.3 kb/s, and 6 kb/s, respectively. We choose T_{GoP}=150 and $T_{est} = 10$, sensing interval $W = 3$, false alarm probability $\epsilon_n = 0.3$ and miss detection probability $\delta_n = 0.25$ for all n, unless otherwise specified.

In every simulation, we compare three schemes: (i) a simple heuristic scheme that equally allocates tiles to each group (Equal Allocation); (ii) A scheme based on SF (Sequential Fixing), and (iii) a scheme based on the greedy algorithm GRD2 (Greedy Algorithm). These schemes have increasing complexity in the order of Equal Allocation, Greedy Algorithm, and Sequential Fixing. They differ on how to solve Problem OPT-Part, while the same tile scheduling algorithm and opportunistic spectrum access scheme are used in all the schemes. Each point in the figures is the average of 10 simulation runs, with 95% confidence intervals plotted. We observe that the 95% confidence intervals for Equal Allocation and Greedy Algorithm are negligible, while the 95% confidence intervals for Sequential Fixing is relatively larger. The C/MATLAB code is executed in a Dell Precision Workstation 390 with an Intel Core 2 Duo E6300 CPU working at 1.86 GHz and a 1066 MB memory. For number of channels ranging from N=3 to N=15, the execution times

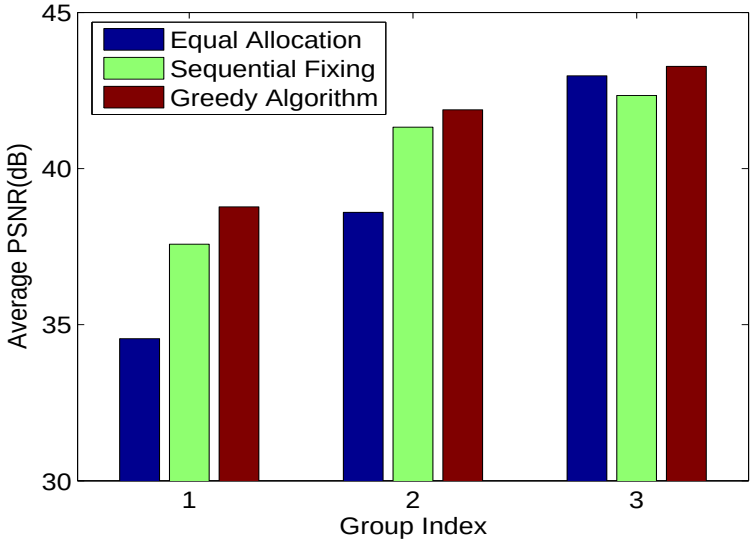


Figure 3.5: Average PSNR of all multicast users.

of Equal Allocation and Greedy Algorithm are about a few milliseconds, while Sequential Fixing takes about two seconds.

In Fig. 3.5 we plot the average PSNR among all users in each multicast group. For all the groups, Greedy Algorithm achieves the best performance, with up to 4.2 dB improvements over Equal Allocation and up to 0.6 dB improvements over Sequential Fixing. We find Sequential Fixing achieves a lower PSNR than Equal Allocation for group 3, but higher PSNRs for groups 1 and 2. This is because Equal Allocation does not consider channel conditions and fairness. It achieves better performance for group 3 at the cost of much lower PSNRs for groups 1 and 2. We also plot Frame 53 from the original *Bus* sequence and the decoded video at user 1 of group 1 in Fig 3.6. We choose this user since it is one of the users with lowest PSNR values. The average PSNR of this user is 29.54 dB, while the average PSNR of all group 1 users is 34.6 dB. Compared to the original frame (right), the reconstructed frame (left) looks quite good, although some details are lost.

In Fig. 3.7, we examine the impact of the maximum allowed collision probability γ_n. We increase γ_n from 0.1 to 0.3, and plot the average PSNR values among all the users. When γ_n gets larger, there will be higher chance of collision for the video packets, which hurts the received

Figure 3.6: The original (the right one) and decoded Frame 53 (the left one) at user 1 in group 1.

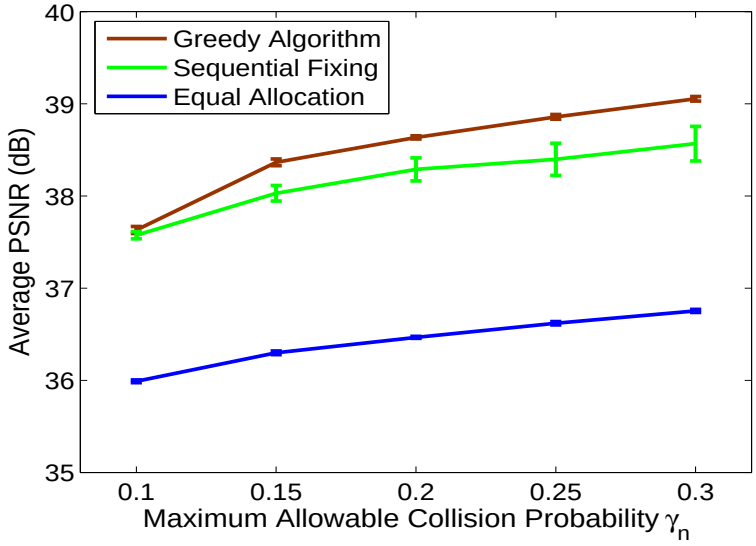


Figure 3.7: Average PSNR of all users versus γ_n (with 95% confidence intervals).

video quality. However, a higher γ_n also allows a higher transmission probability $p_n^{tr}(t)$ for the base station (see (3.6)), thus allowing the base station to grab more spectrum opportunities and achieve a higher video rate. The net effect of these two contradicting effects is improved video quality for the range of γ_n values considered in this simulation. This is illustrated in the figure where all the three curves increase as γ_n gets larger. We also observe that the curves for Sequential Fixing and Equal Allocation are roughly parallel to each other, while the Greedy Algorithm curve has a steeper slope. This indicates that Greedy Algorithm is more efficient in exploiting the additional bandwidth allowed by an increased γ_n.

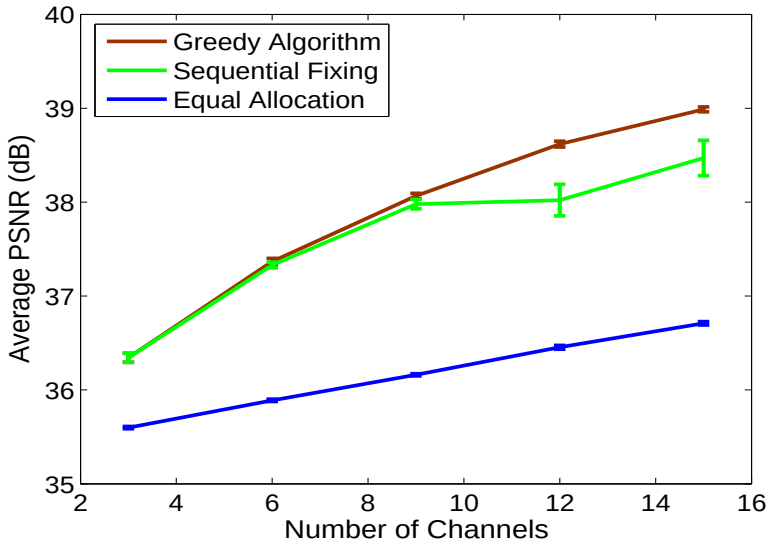


Figure 3.8: Average PSNR of all users versus N (with 95% confidence intervals).

In Fig. 3.8, we examine the impact of number of channels N. We increase N from 3 to 15 in steps of 3, and plot the average PSNR values of all multicast users. As expected, the more channels, the more spectrum opportunities for the CR networks, and the better the video quality. Again, we observe that the Greedy Algorithm curve has the steepest slope, implying it is more efficient in exploiting the increased spectrum opportunity for video transmissions.

We demonstrate the impact of sensing errors in Fig. 3.9. We test five sets of $\{\epsilon_n, \delta_n\}$ values as follows: $\{0.10, 0.38\}$, $\{0.30, 0.25\}$, $\{0.5, 0.17\}$, $\{0.70, 0.10\}$ and $\{0.9, 0.04\}$ [24], and plot the average PSNR values of all users. It is quite interesting to see that the video quality is not very sensitive to sensing errors. Even as ϵ_n is increased nine times from 10% to 90%, there is only 0.58 dB reduction (or a 1.5% normalized reduction) in average PSNR when Greedy Algorithm is used. The same can be observed for the other two curves. We conjecture that this is due to the opportunistic spectrum access approach adopted in all the three schemes. A special strength of the proposed approach is that it explicitly considers both types of sensing errors and mitigates the impact of both sensing errors. For example, when the false alarm rate is very high, the base station will not trust the sensing results and will access the channel relatively more aggressively, thus mitigating the negative effect of the high false alarm rate.

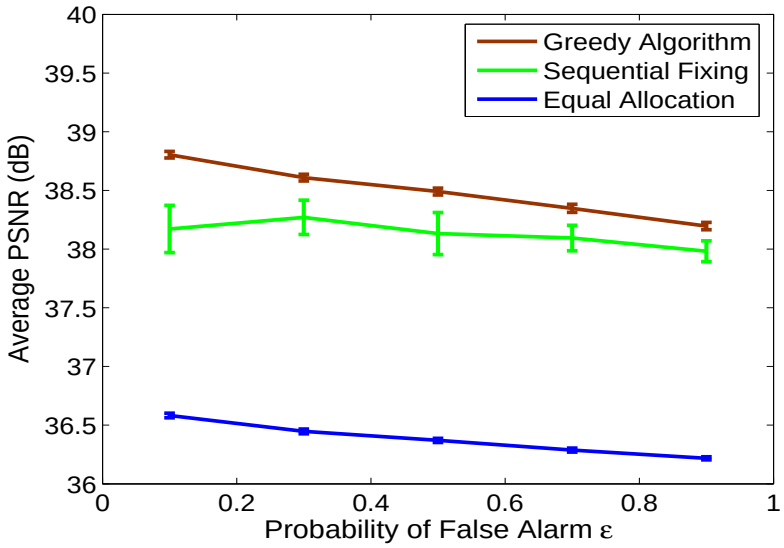


Figure 3.9: Average PSNR of all users for various $\{\epsilon_n, \delta_n\}$ values (with 95% confidence intervals).

Finally, we demonstrate the impact of user channel variations (i.e., due to mobility). We chose a tagged user in group 1 and assume that its channel condition changes every 20 GoPs. The highest MC scheme that the tagged user can decode is changed according to the following sequence: MC3, MC5, MC4, MC6, MC5 and MC3. All other parameters remain the same as in the previous experiments. In Fig. 3.10, we plot the average PSNRs for each GoP at this user that are obtained using the three algorithms. We observe that both Greedy Algorithm and Sequential Fixing can quickly adapt to changing channel conditions. Both algorithms achieve received video qualities commensurate with the channel quality of the tagged user. We also find the video quality achieved by Greedy Algorithm is more stable than that of Sequential Fixing, while the latter curve has some deep fades from time to time. This is due to the fact that Greedy Algorithm has a proven optimality bound, while Sequential Fixing does not provide any guarantee. The Equal Allocation curve is relative constant for the entire period since it does not adapt to channel variations. Although being simple, it does not provide good video quality in this case.

For optimization-driven multimedia systems, there is a trade-off between (i) grabbing all the available resource to maximize media quality and (ii) be less adaptive to network dynamics for a smooth playout. The main objective of this paper is to demonstrate the feasibility and layout the

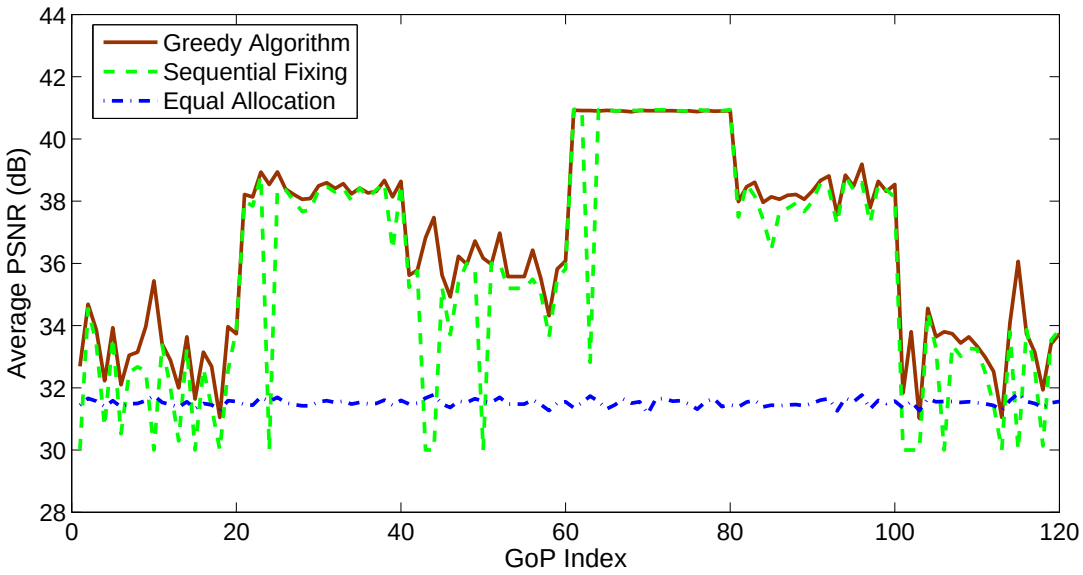


Figure 3.10: GoP average PSNRs of a tagged user in Group 1, when its channel condition varies over time.

framework for video streaming over infrastructure-based CR networks, using an objective function of maximizing the overall user utility. We will investigate the interesting problem of trading off resource utilization and smoothness in our future work.

3.5 Video over Multi-hop CR Networks

In this section, we examine the problem of video over multi-hop CR networks. We model streaming of concurrent videos as an MINLP problem, aiming to maximize the overall received video quality and fairness among the video sessions, while bound the collision rate with primary users under spectrum sensing errors. We solve the MINLP problem using a centralized sequential fixing algorithm, and derive upper and lower bounds for the objective value. We then apply dual decomposition to develop a distributed algorithm and prove its optimality and convergence conditions.

3.5.1 Network Model

Spectrum Access

During the transmission phase of a time slot, a CR user determines which channel(s) to access for transmission of video data based on spectrum sensing results. Let κ_m^k be a threshold for spectrum access: channel m is considered idle if the estimate a_m^k is greater than the threshold, and busy otherwise. The availability of channel m in primary network k, denoted as A_m^k, is

$$A_m^k = \begin{cases} 0, & a_m^k \geq \kappa_m^k \\ 1, & \text{otherwise.} \end{cases} \tag{3.17}$$

For each channel m, we can calculate the probability of collision with primary users as:

$$\Pr(A_m^k = 0|H_1) = \sum_{i \in \psi_m^k} \binom{|\mathcal{U}_m^k|}{i} (1-\delta_m)^{|\mathcal{U}_m^k|-i} (\delta_m)^i, \tag{3.18}$$

where set ψ_m^k is defined as:

$$\psi_m^k = \left\{ i \,\middle|\, \left[1 + \varphi_m^i \phi_m^{|\mathcal{U}_m^k|-i} \frac{\Pr(H_1|\pi_m^k)}{\Pr(H_0|\pi_m^k)}\right]^{-1} \geq \kappa_m^k \right\}. \tag{3.19}$$

For non-intrusive spectrum access, the collision probability should be bounded with a prescribed threshold γ_m^k. A higher spectrum access threshold κ_m^k will reduce the potential interference with primary users, but increase the chance of wasting transmission opportunities. For a given collision tolerance γ_m^k, we can solve $\Pr(A_m^k = 0|H_1) = \gamma_m^k$ for κ_m^k. The objective is to maximize CR users' spectrum access without exceeding the maximum collision probability with primary users.

Let $\Omega_{i,j}$ be the set of available channels at link $\{i, j\}$. Assuming $i \in \mathcal{U}^k$ and $j \in \mathcal{U}^{k'}$, we have

$$\Omega_{i,j} = \left\{ m \,\middle|\, A_m^k = 0 \text{ and } A_m^{k'} = 0 \right\}. \tag{3.20}$$

Link and Path Statistics

Due to the amplify-and-forward approach for video data transmission, there is no queueing delay at intermediate nodes. Assume each link has a fixed delay $\omega_{i,j}$ (i.e., processing and propagation delays). Let $\mathcal{P}_l^A$ be the set of all possible paths from z_l to d_l. For a given delay requirement T_{th}, the set of feasible paths $\mathcal{P}_l$ for video session l can be determined as:

$$\mathcal{P}_l = \left\{ \mathcal{P} \left| \sum_{\{i,j\}\in\mathcal{P}} \omega_{i,j} \le T_{th},\ \mathcal{P} \in \mathcal{P}_l^A \right. \right\}. \tag{3.21}$$

Let $p_{i,j}^m$ be the packet loss rate on channel m at link $\{i,j\}$. A packet is successfully delivered over link $\{i,j\}$ if there is no loss on all the channels that were used for transmitting the packet. The link loss probability $p_{i,j}$ can be derived as:

$$p_{i,j} = 1 - \prod_{m\in\mathcal{M}} (1 - p_{i,j}^m)^{I_m}, \tag{3.22}$$

where $\mathcal{M}$ is set of licensed channels and I_m is an indicator: $I_m = 1$ if channel m is used for the transmission, and $I_m = 0$ otherwise. Assuming independent link losses, the end-to-end loss probability for path $\mathcal{P}_l^h \in \mathcal{P}_l$ can be estimated as:

$$p_l^h = 1 - \prod_{\{i,j\}\in\mathcal{P}_l^h} (1 - p_{i,j}). \tag{3.23}$$

3.5.2 Problem Statement

We also aim to achieve fairness among the concurrent video sessions. It has been shown that *proportional fairness* can be achieved by maximizing the sum of logarithms of video PSNRs (i.e., utilities). Therefore, our objective is to maximize the overall system utility, i.e.,

$$\text{maximize:}\ \sum_l U_l(R_l) = \sum_l \log(Q_l(R_l)). \tag{3.24}$$

Multi-hop CR Network Video Streaming Problem

For the system described in Section 2.4.1, the problem of video over multi-hop CR networks consists of path selection for each video session and channel scheduling for each CR node along the chosen paths. We define two sets of index variables. For channel scheduling, we have

$$x_{i,j,m}^{l,h,r} = \begin{cases} 1, & \text{at link } \{i,j\}, \text{ if channel } m \text{ is} \\ & \text{assigned to tunnel } r \text{ in path } \mathcal{P}_l^h \\ 0, & \text{otherwise.} \end{cases} \tag{3.25}$$

For path selection, we have

$$y_l^h = \begin{cases} 1, & \text{if video session } l \text{ selects path } \mathcal{P}_l^h \in \mathcal{P}_l \\ 0, & \text{otherwise,} \end{cases} \tag{3.26}$$

Note that the indicators, $x_{i,j,m}^{l,h,r}$ and y_l^h, are not independent. If $y_l^h = 0$ for path $\mathcal{P}_l^h$, all the $x_{i,j,m}^{l,h,r}$'s on that path are 0. If link $\{i,j\}$ is not on path $\mathcal{P}_l^h$, all its $x_{i,j,m}^{l,h,r}$'s are also 0. For link $\{i,j\}$ on path $\mathcal{P}_l^h$, we can only choose those available channels in set $\Omega_{i,j}$ to schedule video transmission. That is, we have $x_{i,j,m}^{l,h,r} \in \{0,1\}$ if $m \in \Omega_{i,j}$, and $x_{i,j,m}^{l,h,r} = 0$ otherwise. In the rest of the paper, we use $\mathbf{x}$ and $\mathbf{y}$ to represent the vector forms of $x_{i,j,m}^{l,h,r}$ and y_l^h, respectively.

As discussed, the objective is to maximize the expected utility sum at the end of N_G time slots, as given in (3.24). Since $\log(Q_l(\mathrm{E}[R_l(0)]))$ is a constant, (3.24) is equivalent to the sum of utility increments of all the time slots, as

$$\begin{aligned} &\sum_l \log(Q_l(\mathrm{E}[R_l(N_G)])) - \log(Q_l(\mathrm{E}[R_l(0)])) \\ =&\sum_t \sum_l \left\{\log(Q_l(\mathrm{E}[R_l(t)])) - \log(Q_l(\mathrm{E}[R_l(t-1)]))\right\}. \end{aligned} \tag{3.27}$$

Therefore, (3.24) will be maximized if we maximize the expected utility increment during each time slot, which can be written as:

$$\begin{aligned}
&\sum_{l} \log(Q_l(\mathrm{E}[R_l(t)])) - \log(Q_l(\mathrm{E}[R_l(t-1)])) \\
= &\sum_{l} \log\left(1+\beta_l \frac{\mathrm{E}[R_l(t)]-\mathrm{E}[R_l(t-1)]}{Q_l(\mathrm{E}[R_l(t-1)])}\right) \\
= &\sum_{l}\sum_{h\in\mathcal{P}_l} y_l^h \log\left(1+\sum_{r}\sum_{m}\frac{\beta_l L_p x^{l,h,r}_{z_l,z_l',m}}{N_G T_s Q_l^{t-1}}(1-p^r_{l,h})\right) \\
= &\sum_{l}\sum_{h\in\mathcal{P}_l} y_l^h \log\left(1+\rho_l^t \sum_{r}\sum_{m} x^{l,h,r}_{z_l,z_l',m}(1-p^r_{l,h})\right),
\end{aligned}$$

where z_l' is the next hop from z_l on path $\mathcal{P}_l^h$, $p^r_{l,h}$ is the packet loss rate on tunnel r of path $\mathcal{P}_l^h$, $Q_l^{t-1} = Q_l(\mathrm{E}[R_l(t-1)])$, and $\rho_l^t = \beta_l L_p/(N_G T_s Q_l^{t-1})$.

From (3.22) and (3.23), the end-to-end packet loss rate for tunnel r on path $\mathcal{P}_l^h$ is:

$$p^r_{l,h} = 1 - \prod_{\{i,j\}\in\mathcal{P}_l^h}\prod_{m\in\mathcal{M}} (1-p^m_{i,j})^{x^{l,h,r}_{i,j,m}}. \tag{3.28}$$

We assume that each tunnel can only include one channel on each link. When there are multiple channels available at each link along the path, a CR source node can set up multiple tunnels to exploit the additional bandwidth. We then have the following constraint:

$$\sum_{m} x^{l,h,r}_{i,j,m} \le 1, \quad \forall\, \{i,j\} \in \mathcal{P}_l^h. \tag{3.29}$$

Considering availability of the channels, we further have,

$$\sum_{r}\sum_{m} x^{l,h,r}_{i,j,m} \le |\Omega_{i,j}|, \quad \forall\, \{i,j\} \in \mathcal{P}_l^h, \tag{3.30}$$

where $|\Omega_{i,j}|$ is the number of available channels on link $\{i,j\}$ defined in (3.20).

As discussed, each node is equipped with two transceivers: one for receiving and the other for transmitting video data during the transmission phase. Hence a channel cannot be used to receive and transmit data simultaneously at a relay node. We have for each channel m:

$$\sum_r x_{i,j,m}^{l,h,r} + \sum_r x_{j,k,m}^{l,h,r} \leq 1, \quad \forall\, m, l, \forall\, h \in \mathcal{P}_l, \forall\, \{i,j\}, \{j,k\} \in \mathcal{P}_l^h. \tag{3.31}$$

Let n_l^h be the number of tunnels on path $\mathcal{P}_l^h$. For each source z_l and each destination d_l, the number of scheduled channels is equal to n_l^h. We have for each source node

$$\sum_r \sum_m x_{z_l,z_l',m}^{l,h,r} = n_l^h y_l^h, \quad \forall\, h \in \mathcal{P}_l, \forall\, l. \tag{3.32}$$

Let d_l' be the last hop to destination d_l on path $\mathcal{P}_l^h$, we have for each destination node

$$\sum_r \sum_m x_{d_l',d_l,m}^{l,h,r} = n_l^h y_l^h, \quad \forall\, h \in \mathcal{P}_l, \forall\, l. \tag{3.33}$$

At a relay node, the number of channels used to receive data is equal to that of channels used to transmit data, due to flow conservation and amplify-and-forward. At relay node j for session l, assume $\{i,j\} \in \mathcal{P}_l^h$ and $\{j,k\} \in \mathcal{P}_l^h$. We have,

$$\sum_r \sum_m x_{i,j,m}^{l,h,r} = \sum_r \sum_m x_{j,k,m}^{l,h,r}, \quad \forall\, h \in \mathcal{P}_l, \forall\, l, \forall\, \{i,j\}, \{j,k\} \in \mathcal{P}_l^h. \tag{3.34}$$

We also consider hardware-related constraints on path selection. We summarize such constraints in the following general form for ease of presentation:

$$\sum_l \sum_{h \in \mathcal{P}_l} w_{l,h}^g y_l^h \leq 1, \forall\, g. \tag{3.35}$$

To simplify exposition, we choose at most one path in $\mathcal{P}_l$ for video session l. Such a single path routing constraint can be expressed as $\sum_h y_l^h \leq 1$, which is a special case of (3.35) where $w_{l,h}^1 = 1$ for all h, and $w_{l',h}^g = 0$ for all $g \neq 1$, $l' \neq l$, and h. We can also have $\sum_h y_l^h \leq \xi$ to allow up

to ξ paths for each video session. In order to achieve optimality in the general case of multi-path routing, an optimal scheduling algorithm should be designed to dispatch packets to paths with different conditions (e.g., different number of tunnels and delays).

There are also disjointedness constraints for the chosen paths. This is because each CR node is equipped with two transceivers and both will be used for a video session if it is included in a chosen path. Such disjointedness constraint is also a special case of (3.35) with the following definition for $w_{l,h}^g$ for each CR node g:

$$w_{l,h}^g = \begin{cases} 1, & \text{if node } g \in \text{path } \mathcal{P}_l^h \\ 0, & \text{otherwise}, \end{cases} \tag{3.36}$$

Finally we formulate the problem of multi-hop CR network video streaming (OPT-CRV) as:

$$\begin{aligned} \text{max:} \quad & \sum_l \sum_{h \in \mathcal{P}_l} y_l^h \log\left(1 + \rho_l^t \sum_r \sum_m x_{z_l, z_l', m}^{l,h,r} (1 - p_{l,h}^r)\right) \\ \text{subject to:} \quad & (3.25) \sim (3.35). \end{aligned} \tag{3.37}$$

Centralized Algorithm and Upper/Lower Bounds

Problem OPT-CRV is in the form of MINLP (without continuous variables), which is NP-hard in general. We first describe a centralized algorithm to derive performance bounds in this section, and then present a distributed algorithm based on dual decomposition in the next section.

We first obtain a relaxed *non-linear programming* (NLP) version of OPT-CRV. The binary variables $x_{i,j,m}^{l,h,r}$ and y_l^h are relaxed to take values in [0,1]. The integer variables n_l^h are treated as nonnegative real numbers. It can be shown that the relaxed problem has a concave object function and the constraints are convex. This relaxed problem can be solved using a constrained nonlinear optimization problem solver. If all the variables are integer in the solution, then we have the exact optimal solution. Otherwise, we obtain an infeasible solution, which produces an upper bound for the problem. This is given in Lines 1$\sim$2 in Table 3.5.

Table 3.5: The Sequential Fixing Algorithm (SF) for Problem OPT-CRV

```
1 :  Relax integer variables x_{i,j,m}^{l,h,r}, y_l^h, and n_l^h;
2 :  Solve the relaxed problem using a constrained NLP solver;
3 :  if (there is y_l^h not fixed)
4 :    Find the largest y_{l'}^{h'}, where [l', h'] = arg max{y_l^h},
       and fix it to 1;
5 :    Fix other y_l^h's according to constraint (3.35);
6 :    Go to Step 2;
7 :  end if
8 :  if (there is x_{i,j,m}^{l,h,r} not fixed)
9 :    Find the largest x_{i',j',m'}^{l',h',r'}, where [i', j', m', l', h', r'] =
       arg max{x_{i,j,m}^{l,h,r}}, and set it to 1;
10:    Fix other x_{i,j,m}^{l,h,r}'s according to the constraints;
11:    if (there is other variable that is not fixed)
12:      Go to Step 2;
13:    else
14:      Fix n_l^h's based on x and y;
15:      Exit with feasible solution {x, y, n};
16:    end if
17:  end if
```

We also develop a *sequential fixing algorithm* (SF) for solving OPT-CRV. The pseudo-code is given in Table 3.5. SF iteratively solves the relaxed problem, fixing one or more integer variables after each iteration [8, 47]. In Table 3.5, Lines 3∼7 fix the path selection variables y_l^h, and Lines 8∼16 fix the channel scheduling variables $x_{i,j,m}^{l,h,r}$ and tunnel variables n_l^h. The tunnel variables n_l^h can be computed using (3.32) after $x_{i,j,m}^{l,h,r}$ and y_l^h are solved. When the algorithm terminates, it produces a feasible solution that yields a lower bound for the objective value.

3.5.3 Dual Decomposition

SF is a centralized algorithm requiring global information. It may not be suitable for multi-hop wireless networks, although the upper and lower bounds provide useful insights on the performance limits. In this section, we develop a distributed algorithm for Problem OPT-CRV and analyze its optimality and convergence performance.

Decompose Problem OPT-CRV

Since the domains of $x_{i,j,m}^{l,h,r}$ defined in (3.29)$\sim$(3.34) for different paths do not intersect with each other, we can decompose Problem OPT-CRV into two subproblems. The first subproblem deals with channel scheduling for maximizing the expected utility on a chosen path $\mathcal{P}_l^h$. We have the *channel scheduling* problem (OPT-CS) as:

$$H_l^h = \max_{\mathbf{x}} \sum_r \sum_m x_{z_l,z_l',m}^{l,h,r}(1-p_{l,h}^r) \tag{3.38}$$
$$\text{subject to: } (3.29) \sim (3.34),\ x_{z_l,z_l',m}^{l,h,r} \in \{0,1\},\ \text{for all } l,h,r,m.$$

In the second part, optimal paths are selected to maximize the overall objective function. Letting $F_l^h = \log\left(1+\rho_l^T H_l^h\right)$, we have the following *path selection* problem (OPT-PS):

$$\begin{aligned} \text{maximize:} \quad & f(\mathbf{y}) = \sum_l \sum_h F_l^h y_l^h \\ \text{subject to:} \quad & \sum_l \sum_{h\in\mathcal{P}_l} w_{l,h}^g y_l^h \le 1,\ \text{for all } g \\ & y_l^h \in \{0,1\},\ \text{for all } l,h. \end{aligned} \tag{3.39}$$

Solve the Channel Scheduling Subproblem

We have the following result for assigning available channels at a relay node.

Theorem 3.3. *Consider three consecutive nodes along a path, denoted as nodes i, j, and k. Idle channels 1 and 2 are available at link $\{i,j\}$ and idle channels 3 and 4 are available at link $\{j,k\}$. Assume the packet loss rates of the four channels satisfy $p_{i,j}^1 > p_{i,j}^2$ and $p_{j,k}^3 > p_{j,k}^4$. To set up two tunnels, assigning channels $\{1, 3\}$ to one tunnel and channels $\{2, 4\}$ to the other tunnel achieves the maximum expectation of successful transmission on path section $\{i,j,k\}$.*

Proof. Let the success probabilities on the channels be $\tilde{p}_{i,j}^1 = 1-p_{i,j}^1$, $\tilde{p}_{i,j}^2 = 1-p_{i,j}^2$, $\tilde{p}_{j,k}^3 = 1-p_{j,k}^3$, and $\tilde{p}_{j,k}^4 = 1-p_{j,k}^4$. We have $\tilde{p}_{i,j}^1 < \tilde{p}_{i,j}^2$ and $\tilde{p}_{j,k}^3 < \tilde{p}_{j,k}^4$. Comparing the success probabilities of

the channel assignment given in Theorem 3.3 and that of the alternative assignment, we have $\tilde{p}^1_{i,j}\tilde{p}^3_{j,k} + \tilde{p}^2_{i,j}\tilde{p}^4_{j,k} - \tilde{p}^1_{i,j}\tilde{p}^4_{j,k} - \tilde{p}^2_{i,j}\tilde{p}^3_{j,k} = (\tilde{p}^1_{i,j} - \tilde{p}^2_{i,j})(\tilde{p}^3_{j,k} - \tilde{p}^4_{j,k}) > 0$. The result follows. □

According to Theorem 3.3, a greedy approach, which always chooses the channel with the lowest loss rate at each link when setting up tunnels along a path, produces the optimal overall success probability. More specifically, when there is only one tunnel to be set up along a path, the tunnel should consist of the most reliable channels available at each link along the path. When there are multiple tunnels to set up along a path, tunnel 1 should consist of the most reliable channels that are available at each link; tunnel 2 should consist of the second most reliable links available at each link; and so forth.

Define the set of loss rates of the available channels on link $\{i, j\}$ as $\Lambda_{i,j} = \{p^m_{i,j} | m \in \Omega_{i,j}\}$. The greedy algorithm is given in Table 3.6, with which each video source node solves Problem OPT-CS for each feasible path. Lines 2∼3 in Table 3.6 checks if there is more channels to assign and the algorithm terminates if no channel is left. In Lines 4∼10, links with only one available channel are assigned to tunnel r and the neighboring links with the same available channels are removed due to constraint (3.31). In Lines 11∼17, links with more than two channels are grouped to be assigned later. In Lines 18∼20, the available channel with the lowest packet loss rate is assigned to tunnel r at each unallocated link, according to Theorem 3.3. To avoid co-channel interference, the same channel on neighboring links is removed as in Lines 21∼33.

Solve the Path Selection Subproblem

To solve Problem OPT-PS, we first relax binary variables y^h_l to allow them take real values in [0,1] and obtain the following *relaxed path selection* problem (OPT-rPS):

$$\begin{aligned} \text{maximize:} \quad & f(\mathbf{y}) = \sum_l \sum_h F^h_l y^h_l \\ \text{subject to:} \quad & \sum_l \sum_{h \in \mathcal{P}_l} w^g_{l,h} y^h_l \leq 1, \text{ for all } g \\ & 0 \leq y^h_l \leq 1, \text{ for all } h, l. \end{aligned} \tag{3.40}$$

Table 3.6: The Greedy Algorithm for Channel Scheduling

1 : Initialization: tunnel $r = 1$, link $\{i, j\}$'s from z_l to d_l;
2 : **if** ($|\Lambda_{i,j}|$ == 0)
3 : Exit;
4 : **else if** ($|\Lambda_{i,j}|$ == 1)
5 : Assign the single channel in $\Lambda_{i,j}$, m', to tunnel r;
6 Check neighboring link $\{k, i\}$;
7 : **if** ($p_{k,i}^{m'} \in \Lambda_{k,i}$)
8 : Remove $p_{k,i}^{m'}$ from $\Lambda_{k,i}$,
$i \leftarrow k$, $j \leftarrow i$ and go to Step 2;
9 : **else**
10: Go to Step 13;
11: **end if**
12: **else**
13: Put $\Lambda_{i,j}$ in set Λ_l^h;
14: **if** (node j is not destination d_l)
15: $i \leftarrow j$, $j \leftarrow v$;
16: Go to Step 2;
17: **end if**
18: **end if**
19: **while** (Λ_l^h is not empty)
20: Find the maximum value $p_{i',j'}^{m'}$ in set Λ_l^h
$\{i', j', m'\} = \arg\min\{p_{i,j}^{m}\}$;
21: Assign channel m' to tunnel r;
22: Remove set $\Lambda_{i',j'}$ from set Λ_l^h;
23: Check neighboring link $\{k, i\}$ and $\{j, v\}$;
24: **if** ($p_{k,i}^{m'} \in \Lambda_{k,i}$ and $\Lambda_{k,i} \in \Lambda_l^h$)
25: Remove $p_{k,i}^{m'}$ from $\Lambda_{k,i}$;
26: **if** ($\Lambda_{k,i}$ is empty)
27: Exit;
28: **end if**
29: **end if**
30: **if** ($p_{j,v}^{m'} \in \Lambda_{j,v}$ and $\Lambda_{j,v} \in \Lambda_l^h$)
31: Remove $p_{j,v}^{m'}$ from $\Lambda_{j,v}$;
32: **if** ($\Lambda_{j,v}$ is empty)
33: Exit;
34: **end if**
35: **end if**
36: **end while**
37: Compute the next tunnel: $r \leftarrow r + 1$ and go to Step 2;

We then introduce positive Lagrange Multipliers e_g for the path selection constraints in Problem OPT-rPS and obtain the corresponding *Lagrangian function*:

$$\begin{aligned}\mathcal{L}(\mathbf{y},\mathbf{e}) &= \sum_l \sum_h F_l^h y_l^h + \sum_g e_g(1 - \sum_l \sum_h w_{l,h}^g y_l^h) \\ &= \sum_l \sum_h (F_l^h y_l^h - \sum_g w_{l,h}^g y_l^h e_g) + \sum_g e_g \\ &= \sum_l \sum_h \mathcal{L}_l^h(y_l^h,\mathbf{e}) + \sum_g e_g.\end{aligned} \tag{3.41}$$

Problem (3.41) can be decoupled since the domains of y_l^h's do not overlap. Relaxing the coupling constraints, it can be decomposed into two levels. At the lower level, we have the following subproblems, one for each path $\mathcal{P}_l^h$,

$$\max_{0\le y_l^h\le 1} \mathcal{L}_l^h(y_l^h,\mathbf{e}) = F_l^h y_l^h - \sum_g w_{l,h}^g y_l^h e_g. \tag{3.42}$$

At the higher level, by updating the dual variables e_g, we can solve the *relaxed dual problem*:

$$\min_{\mathbf{e}\ge 0} \quad q(\mathbf{e}) = \sum_l \sum_h \mathcal{L}_l^h \left(\left(y_l^h\right)^*, \mathbf{e} \right) + \sum_g e_g, \tag{3.43}$$

where $\left(y_l^h\right)^*$ is the optimal solution to (3.42). Since the solution to (3.42) is unique, the relaxed dual problem (3.43) can be solved using the following *subgradient method* that iteratively updates the Lagrange Multipliers [69]:

$$e_g(\tau+1) = \left[e_g(\tau) - \alpha(\tau)(1 - \sum_l \sum_h w_{l,h}^g y_l^h) \right]^+, \tag{3.44}$$

where τ is the iteration index, $\alpha(\tau)$ is a sufficiently small positive step size and $[x]^+$ denotes $\max\{x, 0\}$. The pseudo code for the distributed algorithm is given in Table 3.7.

Table 3.7: Distribution Algorithm for Path Selection

1:	Initialization: set $\tau = 0$, $e_g(0) > 0$ and step size $s \in [0, 1]$;
2:	Each source locally solves the lower level problem in (3.42); **if** $(F_l^h - \sum_g d_{l,h}^g e_g(\tau)) > 0)$ $y_l^h = y_l^h + s$, $y_l^h = \min\{y_l^h, 1\}$; **else** $y_l^h = y_l^h - s$, $y_l^h = \max\{y_l^h, 0\}$;
3:	Broadcast solution $y_l^h(\mathbf{e}(\tau))$;
4:	Each source updates $\mathbf{e}$ according to (3.44) and broadcasts $\mathbf{e}(\tau + 1)$ through the common control channel;
5:	$\tau \leftarrow \tau$+1 and go to Step 2 until termination criterion is satisfied;

Optimality and Convergence Analysis

The distributed algorithm in Table 3.7 iteratively updates the dual variables until they converge to stable values. In this section, we first prove that the solution obtained by the distributed algorithm is also optimal for the original path selection problem OPT-PS. We then derive the convergence condition for the distributed algorithm.

Fact 1 ([69])**.** *Consider a linear problem involving both equality and inequality constraints*

$$\begin{aligned} &\textit{maximize:} && \mathbf{a}'\mathbf{x} \\ &\textit{subject to:} && \mathbf{h}_1'\mathbf{x} = b_1, \ \cdots, \ \mathbf{h}_m'\mathbf{x} = b_m \\ & && \mathbf{g}_1'\mathbf{x} \leq c_1, \ \cdots, \ \mathbf{g}_r'\mathbf{x} \leq c_r, \end{aligned} \tag{3.45}$$

where $\mathbf{a}$*,* $\mathbf{h}_i$*, and* $\mathbf{g}_j$ *are column vectors in* $\mathcal{R}_n$*,* b_i*'s and* c_j*'s are scalars, and* $\mathbf{a}'$ *is the* transpose *of* $\mathbf{a}$*. For any feasible point* $\mathbf{x}$*, the set of* active *inequality constraints is denoted by* $\mathcal{A}(\mathbf{x}) = \{j|\mathbf{g}_j'\mathbf{x} = c_j\}$*. If* $\mathbf{x}^*$ *is a maximizer of inequality constrained problem (3.45),* $\mathbf{x}^*$ *is also a maximizer of the following equality constrained problem:*

$$\begin{aligned} &\textit{maximize:} && \mathbf{a}'\mathbf{x} \\ &\textit{subject to:} && \mathbf{h}_1'\mathbf{x} = b_1, \ \cdots, \ \mathbf{h}_m'\mathbf{x} = b_m \\ & && \mathbf{g}_j'\mathbf{x} = c_j, \forall \ j \in \mathcal{A}(\mathbf{x}). \end{aligned} \tag{3.46}$$

Lemma 3.1. *The optimal solution for the relaxed primal problem OPT-rPS in (3.40) is also feasible and optimal for the original Problem OPT-PS in (3.39).*

Proof. According to Fact 1, the linearized problem of OPT-PS, i.e., OPT-rPS, can be rewritten as an equality constrained problem in the following form:

$$\text{maximize:} \quad \mathbf{F}'\mathbf{y} \tag{3.47}$$

$$\text{subject to:} \quad \mathbf{w}_j'\mathbf{y} = 1, \;\; j \in \mathcal{A}(\mathbf{y}^*) \tag{3.48}$$

$$0 \leq y_l^h \leq 1, \text{ for all } h, l,$$

where $\mathbf{F}$, $\mathbf{w}_j$'s, and $\mathbf{y}$ are column vectors with elements F_l^h, $w_{l,h}^g$, and y_l^h, respectively. We apply *Gauss-Jordan elimination* to the constraints in (3.48) to solve for $\mathbf{y}$. Since there is not sufficient number of equations, some y_l^h's are free variables (denoted as y_i^f) and the rest are dependent variables (denoted as y_j^d). Assuming there are r free variables, the dependent variables can be written as linear combinations of the free variables after Gauss-Jordan elimination, as

$$y_j^d = \sum_{i=1}^{r} \bar{w}_j^i y_i^f + \bar{b}_j, \;\; j \in \mathcal{A}(y_i^*). \tag{3.49}$$

Due to Gauss-Jordan elimination and binary vectors $\mathbf{w}_j$'s, $\bar{w}_j^i$ and $\bar{b}_j$ in (3.49) are all integers. Therefore, if all the free variables y_i^f attain binary values, then all the dependent variables y_j^d computed using (3.49) will also be integers. Since $0 \leq y_j^d \leq 1$, being integers means that they are either 0 or 1, i.e., binaries. That is, such a solution will be feasible.

Next we substitute (3.49) into problem (3.47) to eliminate all the dependent variables. Then we obtain a unconstrained problem with only r free variables, as

$$\text{maximize:} \quad \sum_{i=1}^{r} \bar{F}_i y_i^f + \bar{b}_0 \tag{3.50}$$

Since the free variables y_i^f's take value in $\{0, 1\}$, this problem can be easily solved as follows. If the coefficient $\bar{F}_i > 0$, we set $y_i^f = 1$; otherwise, if $\bar{F}_i < 0$, we set $y_i^f = 0$. Thus (3.50) achieves its maximum objective value. Once all the free variables are determined with their optimal binary values, we computes the dependent variables using (3.49), which are also binary as discussed above. Thus we obtain a feasible solution, which is optimal. □

Lemma 3.2. *If the relaxed primal Problem OPT-rPS in (3.40) has an optimal solution, then the relaxed dual problem (3.43) also has an optimal solution and the corresponding optimal values of the two problems are identical.*

Proof. By definition, the problems in (3.41) and (3.43) are primal/dual problems. The primal problem always has an optimal solution because it is bounded. Since Problem OPT-rPS is an LP problem, the relaxed dual problem is also bounded and feasible. Therefore the relaxed dual problem also has an optimal solution. We have the *strong duality* if the primal problem is convex, which is the case here since Problem OPT-rPS is an LP problem. □

We have Theorem 3.4 on the optimality of the path selection solution, which follows naturally from Lemmas 3.1 and 3.2.

Theorem 3.4. *The optimal solution to the relaxed dual problem (3.42) and (3.43) is also feasible and optimal to the original path selection Problem OPT-PS given in (3.39).*

As discussed, the relaxed dual problem (3.43) can be solved using the *subgradient method* that iteratively updates the Lagrange Multipliers. We have the following theorem on the convergence of the distributed algorithm given in Table 3.7.

Theorem 3.5. *Let $\mathbf{e}^*$ be the optimal solution. The distributed algorithm in Table 3.7 converges if the step sizes $\alpha(\tau)$ in (3.44) satisfy the following condition:*

$$0 < \alpha(\tau) < \frac{2\,[q(\mathbf{e}(\tau)) - q(\mathbf{e}^*)\,]}{||G(\tau)||^2}, \quad \textit{for all } \tau, \tag{3.51}$$

where $G(\tau)$ is the gradient of $q(\mathbf{e}(\tau))$.

Proof. Since $q(\mathbf{e}(\tau))$ is a linear function, we have subgradient equality, as

$$q(\mathbf{e}(\tau)) - q(\mathbf{e}^*) = [\mathbf{e}(\tau) - \mathbf{e}^*]' \, G(\tau).$$

It then follows that

$$\begin{aligned} & ||\mathbf{e}(\tau) - \alpha(\tau)G(\tau) - \mathbf{e}^*||^2 \\ & = ||\mathbf{e}(\tau) - \mathbf{e}^*||^2 - 2\alpha(\tau)[\mathbf{e}(\tau) - \mathbf{e}^*]' G(\tau) + (\alpha(\tau))^2 ||G(\tau)||^2 \\ & = ||\mathbf{e}(\tau) - \mathbf{e}^*||^2 - 2\alpha(\tau)[q(\mathbf{e}(\tau)) - q(\mathbf{e}^*)] + (\alpha(\tau))^2 ||G(\tau)||^2. \end{aligned} \tag{3.52}$$

If $\alpha(\tau)$ satisfy (3.51), the sum of the last two terms in (3.52) is negative. It follows that, $||\mathbf{e}(\tau) - \alpha(\tau)G(\tau) - \mathbf{e}^*|| < ||\mathbf{e}(\tau) - \mathbf{e}^*||$. Since the projection operation is *nonexpansive*, we have,

$$\begin{aligned} ||\mathbf{e}(\tau+1) - \mathbf{e}^*|| &= ||[\mathbf{e}(\tau) - \alpha(\tau)G(\tau)]^+ - [\mathbf{e}^*]^+|| \\ &\leq \; ||\mathbf{e}(\tau) - \alpha(\tau)G(\tau) - \mathbf{e}^*|| < ||\mathbf{e}(\tau) - \mathbf{e}^*||, \end{aligned}$$

which states the conditional convergence of the algorithm. □

Since the optimal solution $\mathbf{e}^*$ is not known a priori, we use the following approximation in the algorithm: $\alpha(\tau) = \frac{q(\mathbf{e}(\tau)) - \hat{q}(\tau)}{||G(\tau)||^2}$, where $\hat{q}(\tau)$ is the current estimate for $q(\mathbf{e}^*)$. We choose the mean of the objective values of the relaxed primal and dual problems for $\hat{q}(\tau)$.

Practical Considerations

Our distributed algorithms are based on the fact that the computation is distributed on each feasible path. The OPT-CS algorithm requires information on channel availability and packet loss rates at the links of feasible paths. The OPT-PS algorithm computes the primal variable y_l^h for each

path and broadcasts Lagrangian multipliers over the control channel to all the source nodes. We assume a perfect control channel such that channel information can be effectively distributed and shared, which is not confined by the time slot structure [20].

We assume relatively large timescales for the primary network time slots, and small to medium diameter for the CR network, such that there is sufficient time for timely feedback of channel information to the video source nodes and for the distributed algorithms to converge. Otherwise, channel information can be estimated using (3.4) based on delayed feedback, leading to suboptimal solutions. If the time slot is too short, the distributed algorithm may not converge to the optimal solution (see Fig. 3.15). We focus on developing the CR video framework in this paper, and will investigate these issues in our future work.

3.5.4 Simulation Results

Methodology and Simulation Settings

We implement the proposed algorithms with a combination of C and MATLAB (i.e., for solving the relaxed NLP problems), and evaluate their performance with simulations. For the results reported in this section, we have $K = 3$ primary networks and $M = 10$ channels. There are 56, 55, and 62 CR users in the coverage areas of primary networks 1, 2, and 3, respectively. The $|\mathcal{U}_m^1|$'s are [5 4 6 4 8 7 5 6 7 4] (i.e., five users sense channel 1, four users sense channel 2, and so forth); the $|\mathcal{U}_m^2|$'s are [4 6 5 7 6 5 3 8 5 6], and the $|\mathcal{U}_m^3|$'s are [8 6 5 4 7 6 8 5 6 7]. The topology is shown in Fig. 3.11.

We choose $L_p = 100$, $T_s = 0.02$ and $N_G = 10$. The channel utilization is $\eta_m^k = 0.6$ for all the channels. The probability of false alarm is $\epsilon_m^k = 0.3$ and the probability of miss detection is $\delta_m^k = 0.2$ for all m and k, unless otherwise specified. Channel parameters λ_m^k and μ_m^k are set between $(0, 1)$. The maximum allowed collision probability γ_m^k is set to 0.2 for all the M channels in the three primary networks.

We consider three video sessions, each streaming a video in the Common Intermediate Format (CIF, 352×288), i.e., *Bus* to destination 1, *Foreman* to destination 2, and *Mother & Daughter* to

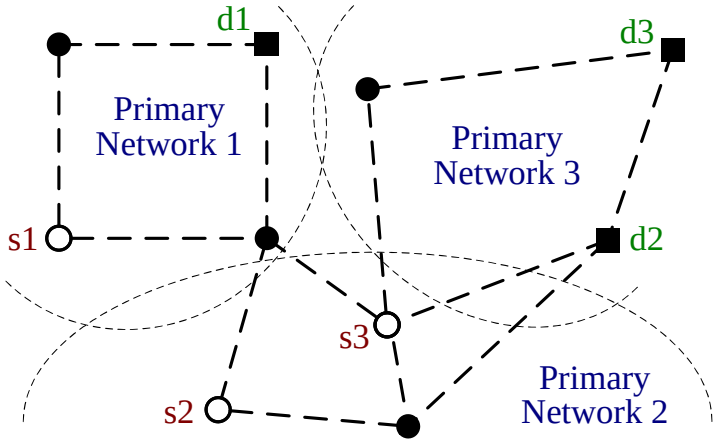


Figure 3.11: Topology of the multi-hop CR network. Note that only video source nodes, video destination nodes, and those nodes along the precomputed paths are shown in the topology.

destination 3. The frame rate is 30 fps, and a GOP consists of 10 frames. We assume that the duration of a time slot is 0.02 seconds and each GOP should be delivered in 0.2 seconds (i.e., 10 time slots).

We compare four schemes in the simulations: (i) the upper-bounding solution by solving the relaxed version of Problem OPT-CRV using an NLP solver, (ii) the proposed distributed algorithm in Tables 3.6 and 3.7, (iii) the sequential fixing algorithm given in Table 3.5, which computes a lower-bounding solution, and (iv) a greedy heuristic where at each hop, the link with the most available channels is used. Each point in the figures is the average of 10 simulation runs, with 95% confidence intervals plotted as error bars in the figures. The 95% confidence intervals are negligible in all the figures.

Simulation Results

Algorithm Performance To demonstrate the convergence of the distributed algorithm, we plot the traces of the four Lagrangian multipliers in Fig. 3.12. We observe that all the Lagrangian multipliers converge to their optimal values after 76 iterations. We also plot the control overhead as measured by the number of distinct broadcast messages for $e_i(\tau)$ using the y-axis on the right-hand side. The overhead curve increases linearly with the number of iterations and gets flat (i.e., no more broadcast message) when all the Lagrangian multipliers converge to their optimal values.

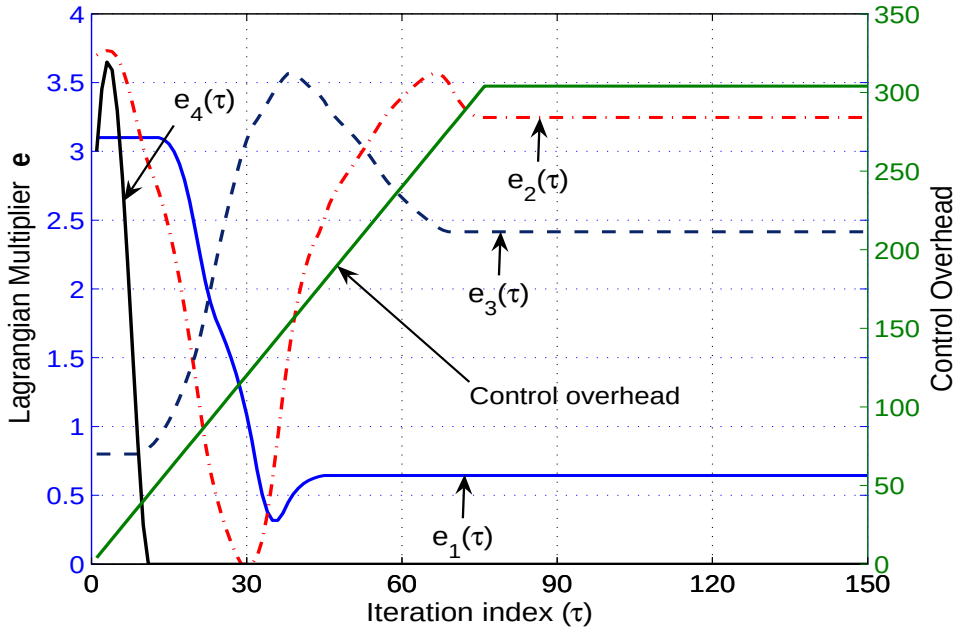


Figure 3.12: Illustrate the convergence of the distributed algorithm.

We examine the impact of spectrum sensing errors in Fig. 3.13. We test six sensing error combinations $\{\epsilon_m, \delta_m\}$ as follows: $\{0.1, 0.5\}$, $\{0.2, 0.3\}$, $\{0.3, 0.2\}$, $\{0.5, 0.11\}$, $\{0.7, 0.06\}$, and $\{0.9, 0.02\}$, and plot the average PSNR values of the Foreman session. It is interesting to see that the best video quality is achieved when the false alarm probability ϵ_m is between 0.2 and 0.3. Since the two error probabilities are correlated, increasing one will generally decrease the other. With a larger ϵ_m, CR users are more likely to waste spectrum opportunities that are actually available, leading to lower bandwidth for videos and poorer video quality, as shown in Fig. 3.13. On the other hand, a larger δ_m implies more aggressive spectrum access and more severe interference to primary users. Therefore when ϵ_m is lower than 0.2 (and δ_m is higher than 0.3), the CR nodes themselves also suffer from the collisions and the video quality degrades.

Impact of Primary Network Parameters In Fig. 3.14, we examine the impact of channel utilization η on received video quality. We focus on Session 2 with the Foreman sequence. The average PSNRs achieved by the four schemes are plotted when η is increased from 0.6 to 0.9 for all licensed channels. Intuitively, a smaller η allows more transmission opportunities for CR nodes, leading to improved video quality. This is illustrated in the figure where all the four curves

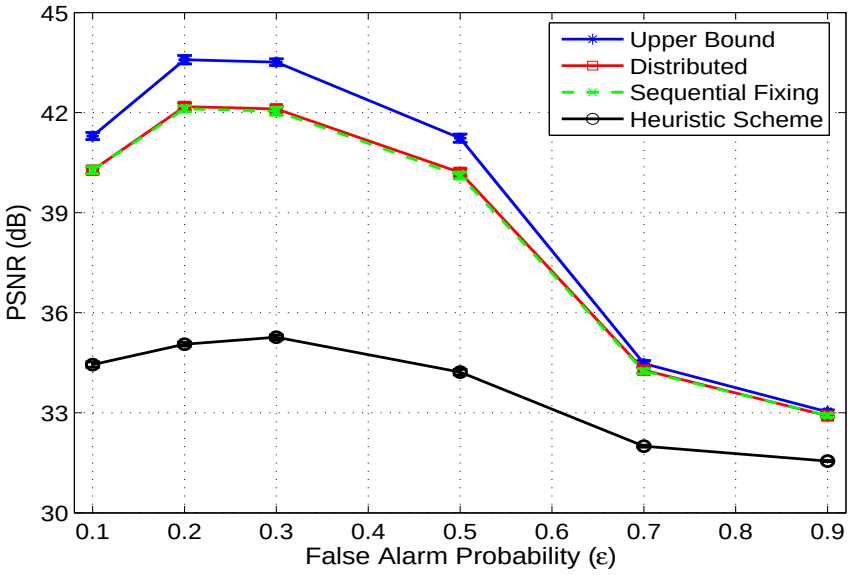


Figure 3.13: Video PSNRs versus spectrum sensing error.

decrease as η gets larger. The distributed scheme achieves PSNRs very close to that obtained by sequential fixing, and both of them are close to the upper bound. The heuristic scheme is inefficient in exploiting the available spectrum even when the channel utilization is low. As discussed, the time slot duration is also an important parameter that may affect the convergence of the distributed algorithm. In Fig. 3.15, we keep the same network and video session settings, while increasing the time slot duration as 4 ms, 10 ms, 20 ms, 40ms and 100 ms. For a given time slot duration, we let the distributed algorithm run for 5% of the time slot duration, starting from the beginning of the time slot, and then stop. The solution that the algorithm produces when it is stopped will be used for video transmission in the remainder of this time slot. It can be seen that when the time slot is 4 ms, the algorithm does not converge after 5%$\times$4=0.2 ms, and the PSNR produced by the distributed algorithm is low (but still higher than that of the heuristic algorithm). When the time slot duration is sufficiently large (e.g., over 10 ms), the algorithm can converge and the proposed algorithm produces very good video quality as compared to the upper bound and the lower bound given by the sequential fixing algorithm.

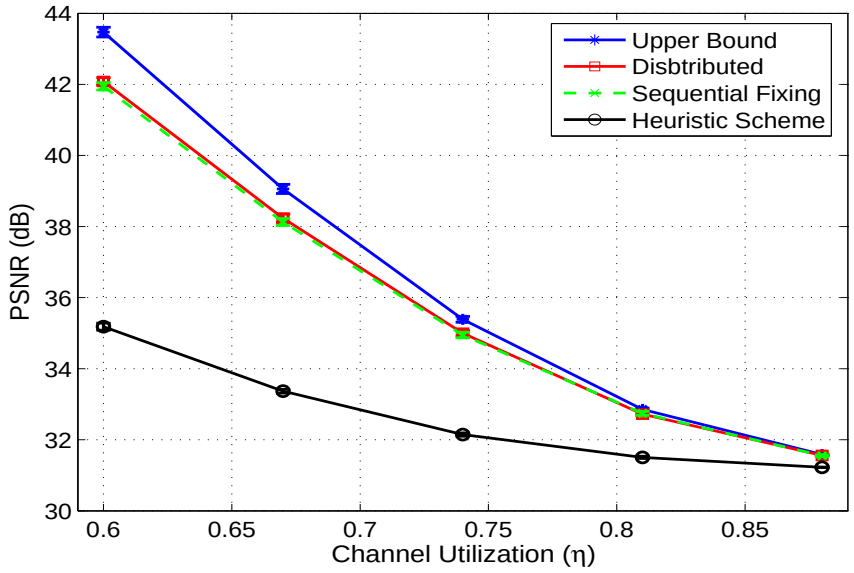


Figure 3.14: Video PSNRs versus primary user channel utilization η.

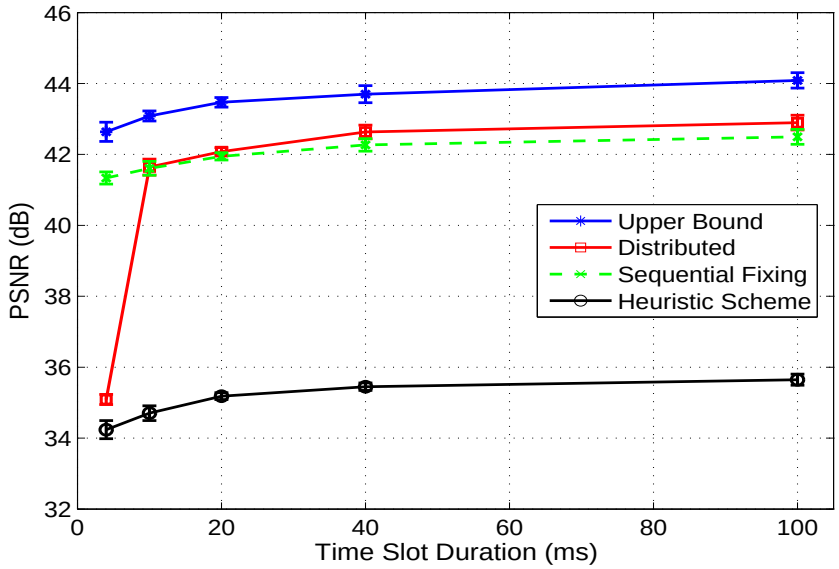


Figure 3.15: Impact of time slot duration on received video quality.

Comparison of MPEG-4 FGS and H.264/SVC MGS Videos Finally, we compare MPEG-4 FGS and H.264/SVC MGS videos, while keeping the same settings. It has been shown that H.264/SVC has better rate-distortion performance than MPEG-4 FGS due to the use of efficient hierarchical prediction structures, the inter-layer prediction mechanisms, improved drift control mechanism, and the efficient coding scheme in H.264/AVC [53]. Although MGS has Network

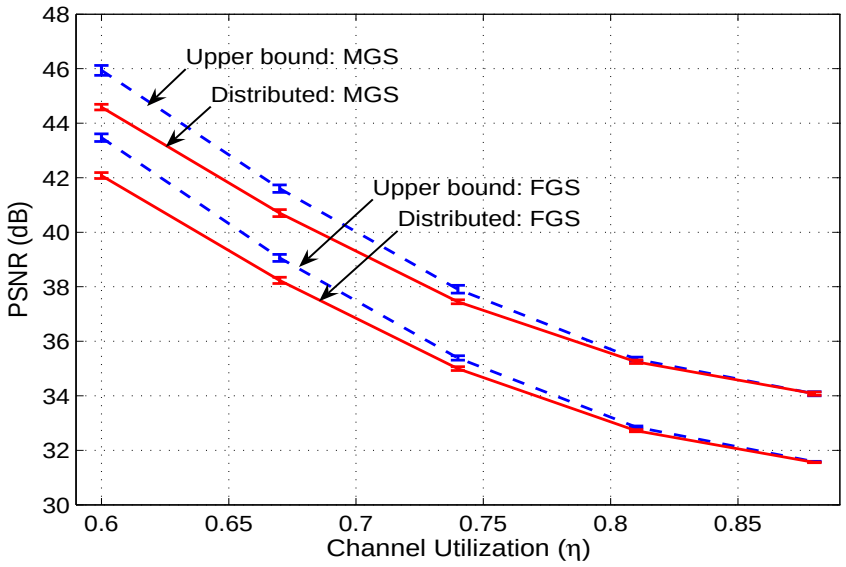


Figure 3.16: Comparison of MPEG-4 FGS video with H.264/SVC MGS video under various channel utilizations.

Abstraction Layer (NAL) unit-based granularity, it achieves similar rate-distortion performance as H.264/SVC FGS [53].

We plot the upper bounds and the distributed algorithm results in Figs. 3.16 and 3.17 for various channel utilizations and false alarm probabilities, respectively. From the figures, it can be observed that there is a gap about 2.5 dB between the H.264/SVC MGS and MPEG-4 FGS curves, which clearly demonstrates the rate-distortion efficiency of MGS over MPEG-4 FGS. The proposed algorithm can effectively handle both MGS and FGS videos, and the same trend is observed in both cases.

3.6 Conclusions

In this chapter, we first addressed the problem of multicasting FGS video in CR networks. The problem formulation took video quality and proportional fairness as objectives, while considering cross-layer design factors such as FGS coding, spectrum sensing, opportunistic spectrum access, primary user protection, scheduling, error control and modulation. We proposed efficient

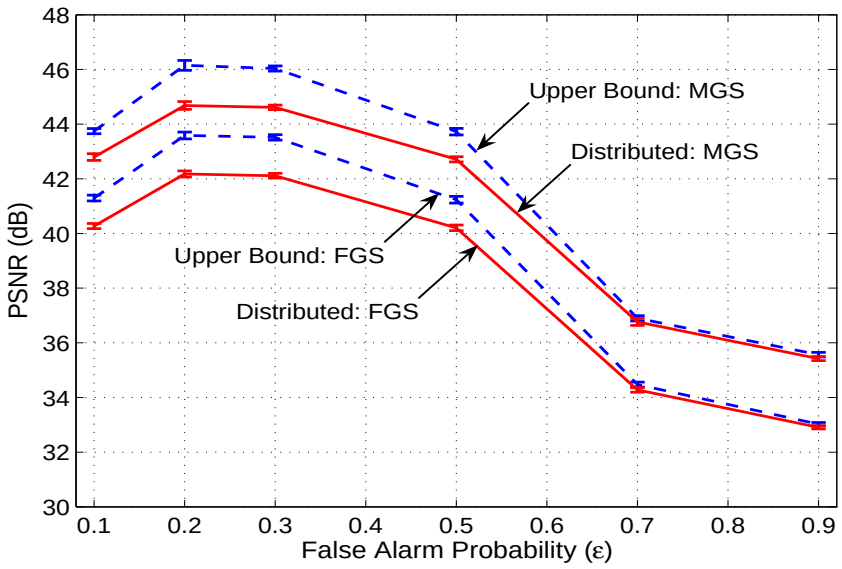


Figure 3.17: Comparison of MPEG-4 FGS video with H.264/SVC MGS video under various false alarm probabilities.

optimization and scheduling algorithms for highly competitive solutions, and proved the complexity and optimality bound of the proposed greedy algorithm. Our simulation results demonstrate not only the viability of video over CR networks, but also the efficacy of the proposed approach.

Then, we studied the challenging problem of streaming multiple scalable videos in a multi-hop CR network. The problem formulation considered spectrum sensing and sensing errors, spectrum access and primary user protection, video quality and fairness, and channel/path selection for concurrent video sessions. We first solved the formulated MINLP problem using a sequential fixing scheme that produces lower and upper bounds on the achievable video quality. We then applied dual decomposition to derive a distributed algorithm, and analyzed its optimality and convergence performance. Our simulations validated the efficacy of the proposed scheme.

Chapter 4

Cooperative CR Networks

4.1 Introduction

Cooperative relay in CR networks [3, 73] represents another new paradigm for wireless communications. It allows wireless CR nodes to assist each other in data delivery, with the objective of achieving greater reliability and efficiency than each of them could attain individually (i.e., to achieve the so-called *cooperative diversity*). Cooperation among CR nodes enables opportunistic use of energy and bandwidth resources in wireless networks, and can deliver many salient advantages over conventional point-to-point wireless communications.

Recently, there has been some interesting work on cooperative relay in CR networks [3, 73]. In [73], the authors considered the case of two single-user links, one primary and one secondary. The secondary transmitter is allowed to act as a "transparent" relay for the primary link, motivated by the rationale that helping primary users will lead to more transmission opportunities for CR nodes. In [3], the authors presented an excellent overview of several cooperative relay scenarios and various related issues. A new MAC protocol was proposed and implemented in a testbed to select a spectrum-rich CR node as relay for a CR transmitter/receiver pair.

We investigate cooperative relay in CR networks, using video as a reference application to make the best use of the enhanced network capacity. We consider a base station (BS) and multiple relay nodes (RN) that collaboratively stream multiple videos to CR users within the network. To support high quality video service in such a challenging environment, we assume a well planned relay network where the RNs are connected to the BS with high-speed wireline links. Therefore the video packets will be available at both the BS and the RNs before their scheduled transmission time, thus allowing advanced cooperative transmission techniques to be adopted for streaming videos. In particular, we consider interference alignment, where the BS and RNs simultaneously

transmit encoded signals to all CR users, such that undesired signals will be canceled and the desired signal can be decoded at each CR user [74, 75]. In [76], such cooperative sender-side techniques are termed *interference alignment*, while receiver-side techniques that use overheard (or exchanged via a wireline link) packets to cancel interference is termed *interference cancelllation*. We present a stochastic programming formulation of the problem of interference alignment for video streaming in cooperative CR networks and then a reformulation of the problem based on Linear Algebra theory [77], such that the number of variables and computational complexity can be greatly reduced. To address the formulated problem, we propose an optimal distributed algorithm with proven convergence and convergence rate, and then a greedy algorithm with a proven performance bound.

The remainder of this chapter is organized as follows. Related work is discussed in Section 4.2. In Section 4.3, we compare two cooperative relay strategies in CR networks. We investigate the problem of cooperative CR relay with interference alignment for MGS video streaming in Section 4.4. Section 4.5 concludes the chapter.

4.2 Background and Related Work

The theoretical foundation of relay channels was laid by the seminal work [78]. The capacities of the Gaussian relay channel and certain discrete relay channels are evaluated, and the achievable lower bound to the capacity of the general relay channel is established in this work. In [79, 80], the authors described the concept of cooperative diversity, where diversity gains are achieved via the cooperation of mobile users. In [54], the authors developed and analyzed low-complexity cooperative diversity protocols. Several cooperative strategies, including AF and DF, were described and their performance characterizations were derived in terms of outage probabilities.

In practice, there is a restriction that each node cannot transmit and receive simultaneously in the same frequency band. The "cheap" relay channel concept was introduced in [81], where the authors derived the capacity of the Gaussian degraded "cheap" relay channel. Multiple relay nodes for a transmitter-receiver pair are investigated in [82] and [83]. The authors showed that, when

compared with complex protocols that involve all relays, the simplified protocol with no more than one relay chosen can achieve the same performance. This is the reason why we consider single relay in this paper.

In [84], Ng and Yu proposed a utility maximization framework for joint optimization of node, relay strategy selection, and power, bandwidth and rate allocation in a cellular network. Cai et al. [85] presented a semi-distributed algorithm for AF relay networks. A heuristic was adopted to select relay and allocate power. Both AF and DF were considered in [86], where a polynomial time algorithm for optimal relay selection was developed and proved to be optimal. In [87], a protocol is proposed for joint routing, relay selection, and dynamic spectrum allocation for multi-hop CR networks, and its performance is evaluated through simulations.

The problem of video over CR networks has only been studied in a few recent papers [8, 9, 13, 59, 60, 88]. In [59], a dynamic channel selection scheme was proposed for CR users to transmit videos over multiple channels. In [88], a distributed joint routing and spectrum sharing algorithm for video streaming over CR ad hoc networks was described and evaluated with simulations. In our prior work, we considered video multicast in an infrastructure-based CR network [8], unicast video streaming over multihop CR networks [9] and CR femtocell networks [13]. In [60], the impact of system parameters residing in different network layers are jointly considered to achieve the best possible video quality for CR users. Unlike the heuristic approaches in [59, 88], the analytical and optimization approach taken in this paper yields algorithms with optimal or bounded performance. The cooperative relay and interference alignment techniques also distinguish this paper from prior work on this topic.

As point-to-point link capacity approaches the Shannon limit, there has been considerable interest on exploiting interference to improve wireless network capacity [74–76, 89, 90]. In addition to information theoretic work on asymptotic capacity [74, 75], practical issues have been addressed in [76, 89, 90]. In [89], the authors presented a practical design of analog network coding to exploit interference and allow concurrent transmissions, which does not make any synchronization assumptions. In [90], interference alignment and cancellation is incorporated in MIMO LANs,

and the network capacity is shown, analytically and experimentally, to be almost doubled. In [76], the authors presented a general algorithm for identifying interference alignment and cancellation opportunities in practical multi-hop mesh networks. The impact of synchronization and channel estimation was evaluated through a GNU Radio implementation. Our work was motivated by these interesting papers, and we incorporate interference alignment in cooperative CR networks and exploit the enhanced capacity for wireless video streaming.

4.3 CR and Cooperative Networking

In this section, we investigate the problem of cooperative relay in CR networks. We assume a primary network with multiple licensed bands and a CR network consisting of multiple cooperative relay links. Each cooperative relay link consists of a CR transmitter, a CR relay, and a CR receiver. The objective is to develop effective mechanisms to integrate these two wireless communication technologies, and to provide an analysis for the comparison of two representative cooperative relay strategies, i.e., *decode-and-forward* (DF) and *amplify-and-forward* (AF), in the context of CR networks. We first consider cooperative spectrum sensing by the CR nodes. We model both types of sensing errors, i.e., miss detection and false alarm, and derive the optimal value for the sensing threshold. Next, we incorporate DF and AF into the p-Persistent Carrier Sense Multiple Access (CSMA) protocol for channel access for the CR nodes. We develop closed-form expressions for the network-wide capacities achieved by DF and AF, respectively, as well as that for the case of direct link transmission for comparison purpose.

Through analytical and simulation evaluations of DF and AF-based cooperative relay strategies, we find the analysis provides upper bounds for the simulated results, which are reasonably tight. We also find cross-point with the AF and DF curves when some system parameter is varied, indicating that each of them performs better in a certain parameter range. There is no case that one completely dominates the other for the two strategies. The considerable gaps between the cooperative relay results and the direct link results exemplify the diversity gain achieved by cooperative relays in CR networks.

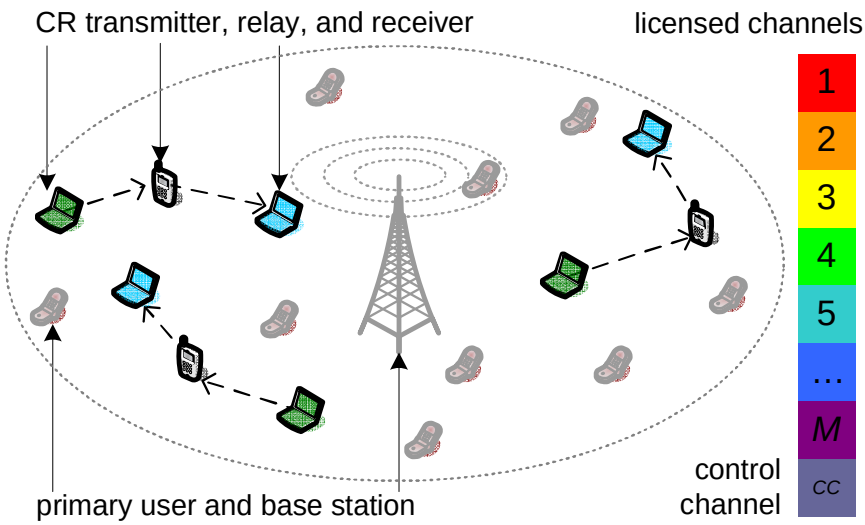


Figure 4.1: Illustration of colocated primary and CR networks. The CR network consists of a number cooperative relay links, each consisting of a CR transmitter, a CR relay and a CR receiver.

4.3.1 Network Model and Assumptions

We assume a primary network and a spectrum band that is divided into M licensed channels, each modeled as a time slotted, block-fading channel. The state of each channel evolves independently following a discrete time Markov process (see Section 2.3.1).

As illustrated in Fig. 4.1, there is a CR network colocated with the primary network. The CR network consists of N sets of cooperative relay links, each including a CR transmitter, a CR relay, and a CR receiver. Each CR node (or, secondary user) is equipped with two transceivers, each incorporating a software defined radio (SDR) that is able to tune to any of the M licensed channels and a control channel and operate from there.

As discussed in Section 2.3.2, we assume CR nodes access the licensed channels following the same time slot structure [14]. In the sensing phase, a CR node chooses one of the M channels to sense using one of its transceivers, and then exchanges sensed channel information with other CR nodes using the other transceiver over the control channel. During the transmission phase, the CR transmitter and/or relay transmit data frames on licensed channels that are believed to be idle based on sensing results, using one or both of the transceivers. We consider cooperative relay strategies AF and DF, and compare their performance in the following sections.

4.3.2 Cooperative Relay in CR Networks

In this section, we investigate how to effectively integrate the two advance wireless communication technologies, and present an analysis of the cooperative relay strategies in CR networks. We first examine cooperative spectrum sensing and derive the optimal sensing threshold. We then consider cooperative relay and spectrum access, and derive the network-wide throughput performance achievable when these two technologies are integrated.

Spectrum Sensing

As illustrated in Section 2.3.3, we assume there are N_m CR nodes sensing channel m. After the sensing phase, each CR node obtains a *sensing result vector* $\vec{\Theta}_m = [\Theta_1^m, \Theta_2^m, \cdots, \Theta_{N_m}^m]$ for channel m. The conditional probability $a_m(\vec{\Theta}_m)$ on channel m availability is

$$\begin{aligned}
& a_m(\Theta_1^m, \Theta_2^m, \cdots, \Theta_{N_m}^m) \\
\cong\ & \Pr\{H_0^m | \Theta_1^m, \Theta_2^m, \cdots, \Theta_{N_m}^m\} \\
=\ & \frac{\Pr\{\Theta_1^m, \Theta_2^m, \cdots, \Theta_{N_m}^m | H_0^m)\} \Pr\{H_0^m\}}{\sum_{j\in\{0,1\}} \Pr\{\Theta_1^m, \Theta_2^m, \cdots, \Theta_{N_m}^m | H_j^m\} \Pr\{H_j^m\}} \\
=\ & \frac{\prod_{i=1}^{N_m} \Pr\{\Theta_i^m | H_0^m\} \Pr\{H_0^m\}}{\sum_{j\in\{0,1\}} \prod_{i=1}^{N_m} \Pr\{\Theta_i^m | H_j^m\} \Pr\{H_j^m\}} \\
=\ & \left[1 + \frac{\Pr\{H_1^m\}}{\Pr\{H_0^m\}} \prod_{i=1}^{N_m} \frac{\Pr\{\Theta_i^m | H_1^m\}}{\Pr\{\Theta_i^m | H_0^m\}}\right]^{-1} \\
=\ & \left[1 + \frac{\eta_m}{1-\eta_m} \prod_{i=1}^{N_m} \frac{(\delta_i^m)^{1-\Theta_i^m}(1-\delta_i^m)^{\Theta_i^m}}{(\epsilon_i^m)^{\Theta_i^m}(1-\epsilon_i^m)^{1-\Theta_i^m}}\right]^{-1}.
\end{aligned} \tag{4.1}$$

If $a_m(\vec{\Theta}_m)$ is greater than a *sensing threshold* τ_m, channel m is believed to be idle; otherwise, channel m is believed to be busy. The decision variable D_m is defined as follows.

$$D_m = \begin{cases} 0, & \text{if } a_m(\vec{\Theta}_m) > \tau_m \\ 1, & \text{if } a_m(\vec{\Theta}_m) \le \tau_m. \end{cases} \tag{4.2}$$

CR nodes only attempt to access channel m where D_m is 0. Since function $a_m(\vec{\Theta}_m)$ in (4.1) has N_m binary variables, there can be 2^{N_m} different combinations corresponding to 2^{N_m} values for $a_m(\vec{\Theta}_m)$. We sort the 2^{N_m} combinations according to their $a_m(\vec{\Theta}_m)$ values in the non-increasing order. Let $a_m^{(j)}$ be the jth largest function value and $\vec{\theta}_m^{(j)}$ the argument that achieves the jth largest function value $a_m^{(j)}$, where

$$\vec{\theta}_m^{(j)} = [\theta_1^m(j), \theta_2^m(j), \cdots, \theta_{N_m}^m(j)].$$

In the design of CR networks, we consider two objectives: (i) how to avoid harmful interference to primary users, and (ii) how to fully exploit spectrum opportunities for the CR nodes. For primary user protection, we limit the collision probability with primary user with a threshold. Let γ_m be the *tolerance threshold*, i.e., the maximum allowable interference probability with primary users on channel m. The probability of collision with primary users on channel m is given as $\Pr\{D_m = 0 \mid H_1^m\}$; the probability of detecting an available transmission opportunity is $\Pr\{D_m = 0 \mid H_0^m\}$. Our objective is to maximize the probability of detecting available channels, while keeping the collision probability below γ_m. Therefore, the optimal spectrum sensing problem can be formulated as follows.

$$\max_{\tau_m} \quad \Pr\{D_m = 0 | H_0^m\} \tag{4.3}$$

$$\text{subect to:} \quad \Pr\{D_m = 0 | H_1^m\} \leq \gamma_m. \tag{4.4}$$

From their definitions, both $\Pr\{D_m = 0 | H_1^m\}$ and $\Pr\{D_m = 0 | H_0^m\}$ are decreasing functions of τ_m. As $\Pr\{D_m = 0 \mid H_1^m)\}$ approaches its maximum allowed value γ_m, $\Pr\{D_m = 0 \mid H_0^m\}$ also approaches its maximum. Therefore, solving the optimization problem (4.3) $\sim$ (4.4) is equivalent to solving

$$\Pr\{D_m = 0 \mid H_1^m\} = \gamma_m.$$

Table 4.1: Algorithm for Computing the Optimal Sensing Threshold

1: Compute $a_m^{(j)}$ and the corresponding $\vec{\theta}_m^{(j)}$, for all j;
2: Initialize $p_c = \Pr\{a_m(\vec{\Theta}_m) = a_m^{(1)} | H_1^m\}$ and $\tau_m = a_m^{(1)}$;
3: Set $j = 1$;
4: WHILE ($p_c \leq \gamma_m$)
5: $\quad j = j + 1$;
6: $\quad \tau_m = a_m^{(j)}$;
7: $\quad p_c = p_c + \Pr\{a_m(\vec{\Theta}_m) = a_m^{(j)} | H_1^m\}$;
8: END WHILE

If $\tau_m = a_m^{(j)}$, we have

$$\begin{aligned} & \Pr\{D_m = 0 | H_1^m\}(a_m^{(j)}) = \Pr\{a_m(\vec{\Theta}_m) > a_m^{(j)} | H_1^m\} \\ = & \sum_{l=1}^{j-1} \Pr\{a_m(\vec{\Theta}_m) = a_m^{(l)} | H_1^m\} = \sum_{l=1}^{j-1} (\delta_i^m)^{1-\theta_i^m(l)} (1-\delta_i^m)^{\theta_i^m(l)}. \end{aligned} \tag{4.5}$$

Obviously, $\Pr\{D_m = 0 \mid H_1^m\}(a_m^{(j)})$ is an increasing function of j. The optimal sensing threshold τ_m^* can be set to $a_m^{(j)}$, such that

$$\Pr\{D_m = 0 \mid H_1^m\}(a_m^{(j)}) \leq \gamma_m$$

and

$$\Pr\{D_m = 0 \mid H_1^m\}(a_m^{(j+1)}) > \gamma_m.$$

The algorithm for computing the optimal sensing threshold τ_m^* is presented in Table 4.1.

Once the optimal sensing threshold τ_m^* is determined, $\Pr\{D_m = 0 \mid H_1^m\}$ can be computed as given in (4.5) and $\Pr\{D_m = 0 \mid H_0^m\}$ can be computed as:

$$\begin{aligned} & \Pr\{D_m = 0 | H_0^m\} = \Pr\{a_m(\vec{\Theta}_m) > \tau_m^* | H_0^m\} \\ = & \sum_{l=1}^{j-1} \Pr\{a_m(\vec{\Theta}_m) = a_m^{(l)} | H_0^m\} = \sum_{l=1}^{j-1} (\epsilon_i^m)^{\theta_i^m(l)} (1-\epsilon_i^m)^{1-\theta_i^m(l)}. \end{aligned} \tag{4.6}$$

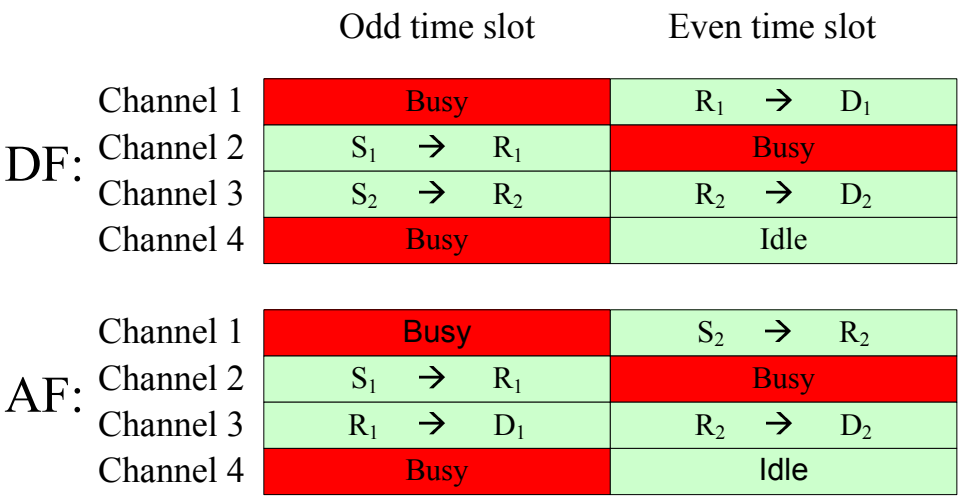


Figure 4.2: Illustration of the protocol operation of AF and DF, where $S_i \Rightarrow R_i$ represents the transmission from source to relay and $R_i \Rightarrow D_i$ represents the transmission from relay to destination, for the ith cooperative relay link.

Cooperative Relay Strategies

During the transmission phase, CR transmitters and relays attempt to send data through the channels that are believed to be idle. We assume fixed length for all the data frames. Let G_1^k and G_2^k denote the path gains from the transmitter to relay and from the relay to receiver, respectively, and let $\sigma_{r,k}^2$ and $\sigma_{d,k}^2$ denote the noise powers at the relay and receiver, respectively, for the kth cooperative relay link. We examine the two cooperation relay strategies DF and AF in the following. For comparison purpose, we also consider direct link transmission below.

Decode-and-Forward (DF) With DF, the CR transmitter and relay transmit separately on consecutive odd and event time slots: the CR transmitter sends data to the corresponding relay in an *odd* time slot; the relay node then decodes the data and forwards it to the receiver in the following *even* time slot, as shown in Fig. 4.2.

Without loss of generality, we assume a data frame can be successfully decoded if the received signal-to-noise ratio (SNR) is no less than a *decoding threshold* κ. We assume gains on different links are independent to each other. The receiver can successfully decode the frame if it is not lost or corrupted on both links. The *decoding rate* of DF at the kth receiver, denoted by P_{DF}^k, can be

computed as,

$$\begin{aligned} P_{DF}^{k} &= \Pr\left\{\left(P_s G_1^k/\sigma_{r,k}^2 \geq \kappa\right) \text{ and } \left(P_r G_2^k/\sigma_{d,k}^2 \geq \kappa\right)\right\} \\ &= \bar{F}_{G_1^k}\left(\sigma_{r,k}^2\kappa/P_s\right)\bar{F}_{G_2^k}\left(\sigma_{d,k}^2\kappa/P_r\right), \end{aligned} \tag{4.7}$$

where P_s and P_r are the transmit powers at the transmitter and relay, respectively, $\bar{F}_{G_1^k}(x)$ and $\bar{F}_{G_2^k}(x)$ are the complementary cumulative distribution functions (CCDF) of path gains G_1^k and G_2^k, respectively.

Amplify-and-Forward (AF) With AF, the CR transmitter and relay transmit simultaneously in the same time slot on different channels. A pipeline is formed connecting the CR transmitter to the relay and then to the receiver; the relay amplifies the received signal and immediately forwards it to the receiver in the same time slot, as shown in Fig. 4.2. Recall that the CR relay has two transceivers. The relay receives data from the transmitter using one transceiver operating on one or more idle channels; it forwards the data simultaneously to the receiver using the other transceiver operating on one or more *different* idle channels.

With this cooperative relay strategy, a data frame can be successfully decoded if the SNR at the receiver is no less than the decoding threshold κ. Then the decoding rate of AF at the kth receiver, denoted as P_{AF}^k, can be computed as,

$$P_{AF}^{k} = \Pr\left\{\frac{P_r}{G_1^k P_s + \sigma_{r,k}^2}\frac{P_s G_1^k G_2^k}{\sigma_{d,k}^2} \geq \kappa\right\} = \int_0^{+\infty} \bar{F}_{G_2^k}\left(\frac{(P_s x + \sigma_{r,k}^2)\sigma_{d,k}^2\kappa}{P_s P_r x}\right) dF_{G_1^k}(x). \tag{4.8}$$

Direct Link Transmission For comparison purpose, we also consider the case of direct link transmission (DL). That is, the CR transmitter transmits to the receiver via the direct link; the CR relay is not used in this case. Let the path gain be G_0^k with CCDF $\bar{F}_{G_0^k}(x)$, and recall that the noise power is $\sigma_{d,k}^2$ at the receiver, for the kth direct link transmission.

Following similar analysis, the decoding rate of DL at the kth receiver, denoted as P_{DL}^k, can be computed as

$$P_{DL}^k = \Pr\left\{P_s G_0^k/\sigma_{d,k}^2 \geq \kappa\right\} = \bar{F}_{G_0^k}\left(\sigma_{d,k}^2\kappa/P_s\right). \quad (4.9)$$

Opportunistic Channel Access

We assume greedy transmitters that always have data to send. The CR nodes use p-Persistent CSMA for channel access. At the beginning of the transmission phase of an odd time slot, CR transmitters send Request-to-Send (RTS) with probability p over the control channel. Since there are N CR transmitters, the transmission probability p is set to $1/N$ to maximize the throughput (i.e., to maximize P_1 in (4.10) given below).

The following three cases may occur:

- *Case 1*: none of the CR transmitters sends RTS for channel access. The idle licensed channels will be wasted.
- *Case 2*: only one CR transmitter sends RTS, and it successfully receives Clear-to-Send (CTS) from the receiver over the control channel. It then accesses some of or all the licensed channels that are believed to be idle for data transmission in the transmission phase.
- *Case 3*: more than one CR transmitters send RTS and collision occurs on the control channel. No CR node can access the licensed channels, and the idle licensed channels will be wasted.

Let P_0, P_1 and P_2 denote the probability corresponding to the three cases enumerated above, respectively. We then have

$$P_0 = (1-p)^N = (1-1/N)^N \quad (4.10)$$

$$P_1 = Np(1-p)^{N-1} = (1-1/N)^{N-1} \quad (4.11)$$

$$P_2 = 1 - P_0 - P_1. \quad (4.12)$$

The CR cooperative relay link that wins the channels in the odd time slot will continue to use the channels in the following even time slot. A new round of channel competition will start in the next odd time slot following these two time slots.

Since a licensed channel is accessed with probability P_1 in the odd time slot, we modify the tolerance threshold γ_m as $\gamma'_m = \gamma_m / P_1$, such that the maximum allowable collision requirement can still be satisfied. In the even time slot, the channels will continue to be used by the winning cooperative relay link, i.e., to be accessed with probability 1. Therefore, the tolerance threshold is still γ_m for the even time slots.

Capacity Analysis

Once the CR transmitter wins the competition, as indicated by a received CTS, it begins to send data over the licensed channels that are inferred to be idle (i.e., $D_m = 0$) in the transmission phase. We assume the *channel bonding and aggregation* technique is used, such that multiple channels can be used collectively by a CR node for data transmission [20, 44].

With DF, the winning CR transmitter uses all the available channels to transmit to the relay in the odd time slot. In the following even time slot, the CR transmitter stops transmission, while the relay uses the available channels in the even time slot to forward data to the receiver. If the number of available channels in the even time slot is equal to or greater than that in the odd time slot, the relay uses the same number of channels to forward all the received data. Otherwise, the relay uses all the available channels to forward part of the received data; the excess data will be dropped due to limited channel resource in the even time slot. The dropped data will be retransmitted in some future odd time slot by the transmitter.

With AF, no matter it is an odd or even time slot, the CR transmitter always uses half of the available licensed channels to transmit to the relay. The relay uses one of its transceivers to receive from the chosen half of the available channels. Simultaneously, it uses the other transceiver to forward the received data to the receiver using the remaining half of the available channels.

Let D_m^{od} and D_m^{ev} be the decision variables of channel m in the odd and even time slot, respectively (see (4.2)). Let S_m^{od} and S_m^{ev} be the status of channel m in the odd and even time slot, respectively. We have,

$$\begin{aligned} &\Pr\{D_m^{od}=i, S_m^{od}=j, D_m^{ev}=k, S_m^{ev}=l\} \\ =&\Pr\{D_m^{ev}=k|S_m^{ev}=l\}\Pr\{D_m^{od}=i|S_m^{od}=j\}\times \\ &\Pr\{S_m^{ev}=l|S_m^{od}=j\}\Pr\{S_m^{od}=j\},\ \text{for } i,j,k,l\in\{0,1\}. \end{aligned} \tag{4.13}$$

where $\Pr\{S_m^{od}=j\}$ are the probabilities that channel m is busy or idle, $\Pr\{S_m^{ev}=l \mid S_m^{od}=j\}$ are the channel m transition probabilities. $\Pr\{D_m^{ev}=k \mid S_m^{ev}=l\}$ and $\Pr\{D_m^{od}=i \mid S_m^{od}=j\}$ can be computed as in (4.5) and (4.6).

Let N_{DF}, N_{AF} and N_{DL} be the number of frames successfully delivered to the receiver in the two consecutive time slots using DF, AF and DL, respectively. Define $\bar{S}_m^{od}=1-S_m^{od}$, $\bar{S}_m^{ev}=1-S_m^{ev}$, $\bar{D}_m^{od}=1-D_m^{od}$ and $\bar{D}_m^{ev}=1-D_m^{ev}$. We have

$$N_{DF} = \left(\sum_{m=1}^{M}\bar{S}_m^{od}\bar{D}_m^{od}\right)\wedge\left(\sum_{m=1}^{M}\bar{S}_m^{ev}\bar{D}_m^{ev}\right) \tag{4.14}$$

$$N_{AF} = \left\lfloor\frac{1}{2}\sum_{m=1}^{M}\bar{S}_m^{od}\bar{D}_m^{od}\right\rfloor+\left\lfloor\frac{1}{2}\sum_{m=1}^{M}\bar{S}_m^{ev}\bar{D}_m^{ev}\right\rfloor \tag{4.15}$$

$$N_{DL} = \left(\sum_{m=1}^{M}\bar{S}_m^{od}\bar{D}_m^{od}\right)+\left(\sum_{m=1}^{M}\bar{S}_m^{ev}\bar{D}_m^{ev}\right), \tag{4.16}$$

where $x\wedge y$ represents the minimum of x and y, and $\lfloor x\rfloor$ means the maximum integer that is not larger than x.

As discussed, the probability that a frame can be successfully delivered is P_{DF}^k, P_{AF}^k, or P_{DL}^k for the three schemes, respectively. Recall that spectrum resources are allocated distributedly for every pair of two consecutive time slots. We derive the capacity for the three cooperative relay

strategies as

$$C_{DF} = \mathrm{E}\left[N_{DF}\right] \cdot \sum_{k=1}^{N} (P_{DF}^{k} P_1 L)/(2NT_s) \quad (4.17)$$

$$C_{AF} = \mathrm{E}\left[N_{AF}\right] \cdot \sum_{k=1}^{N} (P_{AF}^{k} P_1 L)/(2NT_s) \quad (4.18)$$

$$C_{DL} = \mathrm{E}\left[N_{DL}\right] \cdot \sum_{k=1}^{N} (P_{DL}^{k} P_1 L)/(2NT_s), \quad (4.19)$$

where L is the packet length and T_s is the duration of a time slot. The expectations are computed using the results derived in (4.13) $\sim$ (4.16).

4.3.3 Performance Evaluation

We evaluate the performance of the cooperative relay strategies with analysis and simulations. The analytical capacities of the schemes are obtained with the analysis presented in Section 4.3.2. The actual throughput is obtained using MATLAB simulations. The simulation parameters and their values are listed in Table 4.2, unless specified otherwise. We consider $M = 5$ licensed channels and a CR network with seven cooperative relay links. The channels have identical parameters for the Markov chain models. Each point in the simulation curves is the average of 10 simulation runs with different random seeds. We plot 95% confidence intervals for the simulation results, which are negligible in all the cases.

We first examine the impact of the number of licensed channels. To illustrate the effect of spectrum sensing, we let the decoding rate P_{AF}^{k} be equal to P_{DF}^{k}. In Fig. 4.3, we plot the throughput of AF, DF, and DL under increased number of licensed channels. The analytical curves are upper bounds for the simulation curves in all the cases, and the gap between the two is reasonably small. Furthermore, as the number of license channels is increased, the throughput of both AF and DF are increased. The slope of the AF curves is larger than that of the DF curves. There is a cross point between five and six, as predicted by both simulation and analysis curves. This indicates that AF outperforms DF when the number of channels is large. This is because AF is more flexible than DF in exploiting the idle channels in the two consecutive time slots. The DL analysis and

Table 4.2: Simulation Parameters and Values

Symbol	*Value*	*Definition*
M	5	number of licensed channels
λ	0.7	channel transition probability from idle to idle
μ	0.2	channel transition probability from busy to idle
η	0.6	channel utilization
γ	0.08	maximum allowable collision probability
N	7	number of CR cooperative relay links
P_s	10 dBm	transmit power of the CR transmitters
P_r	10 dBm	transmit power of CR relays
L	1 kb	packet length
T_s	1 ms	duration of a time slot

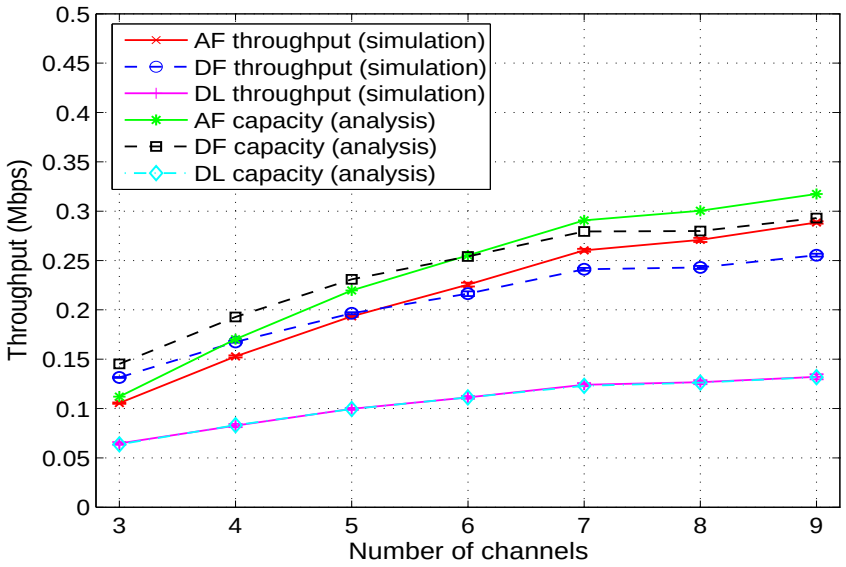


Figure 4.3: Throughput performance versus number of licensed channels.

simulation curves also increases with the number of channels, but with the lowest slope and the lowest throughput values.

In Fig. 4.4, we demonstrate the impact of channel utilization on the throughput of the schemes. The channel utilization η is increased from 0.3 to 0.9, when primary users get more active. As η is increased, the transmission opportunities for CR nodes are reduced and all the throughputs

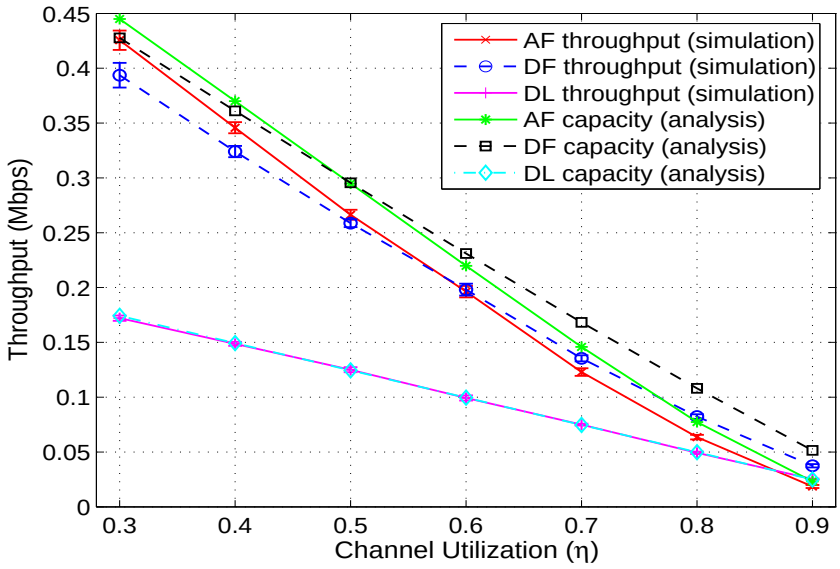


Figure 4.4: Throughput performance versus primary user channel utilization.

are degraded. We find the throughputs of AF and DF are close to each other when the channel utilization is high. AF outperforms DF in the low channel utilization region, but is inferior to DF in the high channel utilization region. There is a cross point between the AF and DF curves between $\eta = 0.5$ and $\eta = 0.6$. When the channel utilization is low, there is a big gap between the cooperative relay curves and and the DL curves.

In Fig. 4.5, we examine the channel fading factor. We consider Rayleigh block fading channels, where the received power is exponentially distributed with a distance-dependent mean. We fix the transmitter power at 10 dBm, and increase the relay power from one dBm to 18 dBm. As the relay power is increased, the throughput is also increased since the SNR at the receiver is improved. We can see the increasing speed of AF is larger than that of DF, indicating that AF has superior performance than DF when the relay transmit power is large. The capacity analysis also demonstrate the same trend. The throughput of DL does not depend on the relay node. Its throughput is better than that of AF and DF when the relay transmit power is low, since both AF and DF are limited by the relay-to-receiver link in this low power region. However, the throughputs of AF and DF quickly exceed that of DL and grow fast as the relay-to-receiver link is improved with the increased relay transmit power. The considerable gaps between the cooperative relay link curves

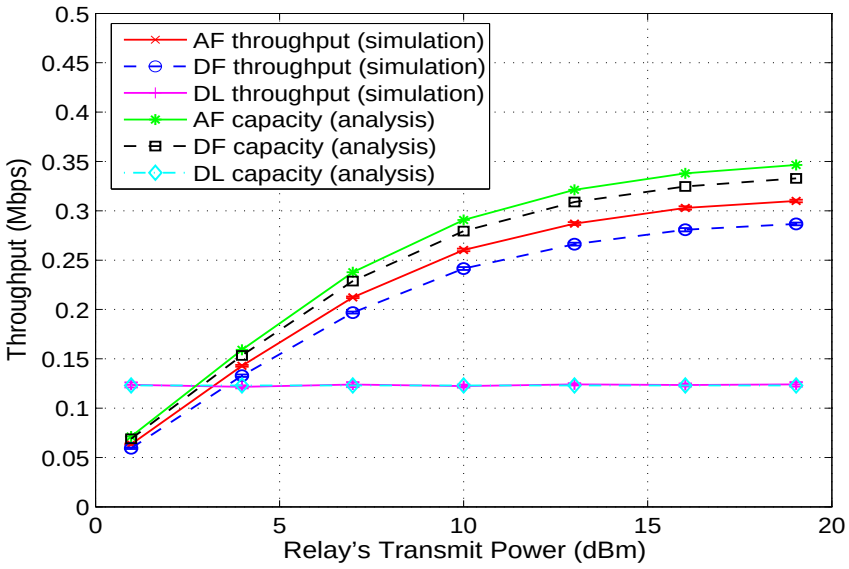


Figure 4.5: Throughput performance versus transmit power of relay nodes.

and the DL curves in Figs. 4.3, 4.4 and 4.5 exemplify the diversity gain achieved by cooperative relays in CR networks.

4.4 Cooperative CR Networks with Interference Alignment

In this section, we investigate cooperative relay in CR networks using video as a reference application. We consider a base station (BS) and multiple relay nodes (RN) that collaboratively stream multiple videos to CR users within the network. It has been shown that the performance of a cooperative relay link is mainly limited by two factors:

- the *half-duplex operation*, since the BS–RN and the RN–user transmissions cannot be scheduled simultaneously on the same channel [79]; and
- the *bottleneck channel*, which is usually the BS–user and/or the RN–user channel, usually with poor quality due to obstacles, attenuation, multipath propagation and mobility [54].

To support high quality video service in such a challenging environment, we assume a well planned relay network where the RNs are connected to the BS with high-speed wireline links, and explore interference alignment to overcome the bottleneck channel problem. Therefore the

video packets will be available at both the BS and the RNs before their scheduled transmission time, thus allowing advanced cooperative transmission techniques (e.g. interference alignment) to be adopted for streaming videos. In particular, we incorporate interference alignment to allow transmitters collaboratively send encoded signals to all CR users, such that undesired signals will be canceled and the desired signal can be decoded at each CR user.

We present a stochastic programming formulation, as well as a reformulation that greatly reduces computational complexity. In the cases of a single licensed channel and multiple licensed channels with channel bonding, we develop an optimal distributed algorithm with proven convergence and convergence speed. In the case of multiple channels without channel bonding, we develop a greedy algorithm with a proven performance bound.

4.4.1 Network Model and Assumptions

The cooperative CR network is illustrated in Fig. 4.6. There is a CR BS (indexed 1) and $(K-1)$ CR RNs (indexed from 2 to K) deployed in the area to serve N active CR users. Let $\mathcal{U} = \{1, 2, \cdots, N\}$ denote the set of active CR users. We assume that the BS and all the RNs are equipped with multiple transceivers: one is tuned to the common control channel and the others are used to sense multiple licensed channels at the beginning of each time slot, and to transmit encoded signals to CR users. We consider the case where each CR user has one software defined radio (SDR) based transceiver, which can be tuned to operate on any of the $(M+1)$ channels. If the channel bonding/aggregation techniques are used [44, 70], a transmitter can collectively use all the available channels and a CR user can receive from all the available channels simultaneously. Otherwise, only one licensed channel will be used by a transmitter and a CR user can only receive from a single chosen channel at a time.

Consider the three channels in a traditional cooperative relay link. Usually the BS and RNs are mounted on high towers, and the BS–RN channel has good quality due to line-of-sight (LOS) communications and absence of mobility. On the other hand, a CR user is typically on the ground level. The BS–user and RN–user channels usually have much poorer quality due to obstacles,

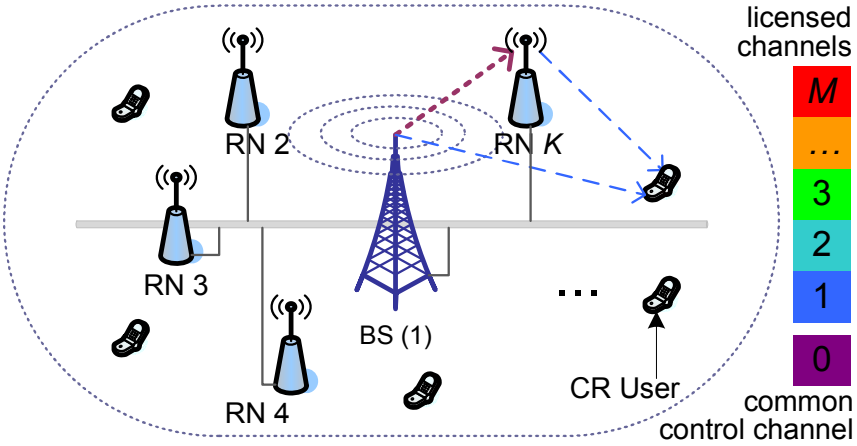


Figure 4.6: Illustration of the cooperative CR network.

attenuation, multipath propagation and mobility. To support high quality video service, we assume a well planned relay network, where the RNs are connected to the BS via broadband wireline connections (e.g., as in femtocell networks [13]). Alternatively, free space optical links can also be used to provide multi-gigabit rates between the BS and the RNs [91]. As a result, the video packets will always be available for transmission (with suitable channel coding and retransmission) at the RNs at their scheduled transmission time. To cope with the much poorer BS–user and RN–user channels, the BS and RNs adopt interference alignment to cooperatively transmit video packets to CR users, while exploiting the spectrum opportunities in the licensed channels.

Spectrum Access

The BS and the RNs sense the licensed channels and exchange their sensing results over the common control channel during the sensing phase. Recall that in Section 2.5.1 and 2.5.1. Given L sensing results obtained for channel m, the corresponding sensing result vector is $\vec{\Theta}_L^m = [\Theta_1^m, \Theta_2^m, \cdots, \Theta_L^m]$. Let $P_m^A(\vec{\Theta}_L^m) := P_m^A(\Theta_1^m, \Theta_2^m, \cdots, \Theta_L^m)$ be the conditional probability that

channel m is available, which can be computed iteratively as shown in our prior work [13]:

$$\begin{aligned} P_m^A(\Theta_1^m) &= \left[1+\frac{\eta_m}{1-\eta_m}\times\frac{(\delta_1^m)^{1-\Theta_1^m}(1-\delta_1^m)^{\Theta_1^m}}{(\epsilon_1^m)^{\Theta_1^m}(1-\epsilon_1^m)^{1-\Theta_1^m}}\right]^{-1} \\ P_m^A(\vec{\Theta}_l^m) &:= P_m^A(\Theta_1^m,\Theta_2^m,\cdots,\Theta_l^m) \\ &= \left\{1+\left[\frac{1}{P_m^A(\Theta_1^m,\Theta_2^m,\cdots,\Theta_{l-1}^m)}-1\right]\times\frac{(\delta_l^m)^{1-\Theta_l^m}(1-\delta_l^m)^{\Theta_l^m}}{(\epsilon_l^m)^{\Theta_l^m}(1-\epsilon_l^m)^{1-\Theta_l^m}}\right\}^{-1}, l\geq 2. \end{aligned}$$

For each channel m, define an index variable $D_m(t)$ for the BS or RNs to access the channel in time slot t. That is,

$$D_m(t)=\begin{cases} 0, & \text{access channel } m \text{ in time slot } t \\ 1, & \text{otherwise,} \end{cases} \quad m=1,2,\cdots,M. \tag{4.20}$$

With sensing result $P_m^A(\vec{\Theta}_L^m)$, each channel m will be opportunistically accessed. Let the probability be $P_m^D(\vec{\Theta}_L^m)$ that channel m will be accessed in time slot t (i.e., when $D_m(t)=0$). The optimal channel access probability can be computed as:

$$P_m^D(\vec{\Theta}_L^m)=\min\left\{\gamma_m/\left[1-P_m^A(\vec{\Theta}_L^m)\right],1\right\}. \tag{4.21}$$

Let $\mathcal{A}(t)$ be the set of available channels in time slot t. It follows that $\mathcal{A}(t):=\{m\mid D_m(t)=0\}$.

Interference Alignment

We next briefly describe the main idea of interference alignment considered in this paper. Interested readers are referred to [76, 90] for insightful examples, a classification of various interference alignment scenarios, and practical considerations.

Consider two transmitters (denoted as s_1 and s_2) and two receivers (denoted as d_1 and d_2). Let X_1 and X_2 be the signals corresponding to the packets to be sent to d_1 and d_2, respectively. With interference alignment, the transmitters s_1 and s_2 send compound signals $a_{1,1}X_1+a_{1,2}X_2$ and $a_{2,1}X_1+a_{2,2}X_2$, respectively, to the two receivers d_1 and d_2 simultaneously. If channel noise

is ignored, the received signals Y_1 and Y_2 can be written as:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} G_{1,1} & G_{1,2} \\ G_{2,1} & G_{2,2} \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} := \mathbf{G}^{\mathsf{T}} \times \mathbf{A} \times \vec{X}, \tag{4.22}$$

where $G_{i,j}$ is the channel gain from transmitter s_i to receiver d_j.

From (4.22), it can be seen that both receivers can perfectly decode their signals if the transformation matrix $\mathbf{A}$ is chosen to be $\left\{\mathbf{G}^{\mathsf{T}}\right\}^{-1}$, i.e., the inverse of the channel gain matrix. With this technique, the transmitters are able to send packets simultaneously and the interference between the two concurrent transmissions can be effectively canceled at both receivers [76].

4.4.2 Problem Formulation

We formulate the problem of interference alignment for scalable video streaming over cooperative CR networks in this section. As discussed in Section 4.3.1, the video packets are available at both the BS and all the RNs before their scheduled transmission time; the BS and RNs adopt interference alignment to overcome the poor BS–CR user and RN–CR user channels.

Let X_j denote the signal to be transmitted to user j, which has unit power. As illustrated in Section 4.4.1, transmitter k sends a compound signal $\sum_{j\in\mathcal{U}} a_{k,j}X_j$ to all active CR users, where $a_{k,j}$'s are the weights to be determined. Ignoring channel noise, we can compute the received signal Y_n at a user n as:

$$\begin{aligned} Y_n &= \sum_{k=1}^{K} G_{k,n} \sum_{j=1}^{N} a_{k,j}X_j = \sum_{k=1}^{K}\sum_{j=1}^{N} a_{k,j}G_{k,n}X_j \\ &= \sum_{j=1}^{N} X_j \sum_{k=1}^{K} a_{k,j}G_{k,n},\ n = 1, 2, \cdots, N, \end{aligned} \tag{4.23}$$

where $G_{k,n}$ is the channel gain from the BS (i.e., $k = 1$) or an RN k to user n. For user n, only signal X_n should be decoded and the coefficients of all other signals should be forced to zero. The

zero-forcing constraints can be written as:

$$\sum_{k=1}^{K} a_{k,j} G_{k,n} = 0, \quad \text{for all } j \neq n. \tag{4.24}$$

Usually the total transmit power of the BS and every RN is limited by a peak power P_{max}. Since X_j has unit power, for all j, the power of each transmitted signal is the square sum of all the coefficients $a_{k,j}^2$. The *peak power constraint* can be written as

$$\sum_{j=1}^{N} |a_{k,j}|^2 \leq P_{max}, \quad k = 1, \cdots, K. \tag{4.25}$$

Recall that each CR user has one SDR transceiver that can be tuned to receive from any of the $(M+1)$ channels, when channel bonding is not adopted. Let b_j^m be a binary variable indicating that user j selects licensed channel m. It is defined as

$$b_j^m = \begin{cases} 1, & \text{if user } n \text{ receives from channel } m \\ 0, & \text{otherwise,} \end{cases} \quad j = 1, \cdots, N, \ m = 1, \cdots, M. \tag{4.26}$$

Then, we have the following *transceiver constraint*:

$$\sum_{m \in \mathcal{A}(t)} b_j^m \leq 1, \quad j = 1, \cdots, N. \tag{4.27}$$

After introducing the channel selection variables b_j^m's, the overall channel gain becomes

$$G_{k,j} = \sum_{m \in \mathcal{A}(t)} b_j^m H_{k,j}^m, \tag{4.28}$$

where $H_{k,j}^m$ is the channel gain from the BS (i.e., $k = 1$) or an RN k to user j on channel m.

Let w_j^t be the PSNR of user j's reconstructed video at the beginning of time slot t and W_j^t the PSNR of user j's reconstructed video at the end of time slot t. In time slot t, w_j^t is already known,

while W_j^t is a random variable depending on the resource allocation and primary user activity during the time slot. That is, w_j^{t+1} is a realization of W_j^t.

As discussed in Section 3.3.5, the quality of reconstructed MGS video can be modeled with a linear equation [53]:

$$W(R) = \alpha + \beta \times R, \tag{4.29}$$

where $W(R)$ is the average peak signal-to-noise ratio (PSNR) of the reconstructed MGS video, R is the average data rate, and α and β are constants depending on the specific video sequence and codec.

We formulate a multistage stochastic programming problem to maximize the sum of expected logarithm of the PSNR's at the end of the GOP, i.e., $\sum_{j=1}^{N} \mathbb{E}\left[\log(W_j^T)\right]$, for proportional fairness among the video sessions [71]. It can be shown that the multistage stochastic programming problem can be decomposed into T serial sub-problems, one for each time slot t, as [9]:

$$\text{maximize:} \quad \sum_{j=1}^{N} \mathbb{E}\left[\log(W_j^t)|w_j^t\right] \tag{4.30}$$

$$\text{subject to:} \quad W_j^t = w_j^t + \Psi_j^t \tag{4.31}$$

$$b_j^m \in \{0,1\},\ a_{k,j} \geq 0,\ \text{for all } m,j,k \tag{4.32}$$

$$\text{Constraints (4.24), (4.25) and (4.27),}$$

where Ψ_j^t is a random variable that depends on spectrum sensing, power allocation, and channel selection in time slot t. This is a mixed integer nonlinear programming problem (MINLP), with binary variables b_j^m's and continuous real variables $a_{k,j}$'s.

In particular, Ψ_j^t can have two possible values: (i) zero, if the packet is not successfully received due to collision with primary users; (ii) the PSNR increase achieved in time slot t if the packet is successfully received, denoted as λ_j^t. The PSNR increase can be computed as:

$$\lambda_j^t = \frac{\beta_j B}{T} \log_2 \left(1 + \frac{1}{N_0} \left(\sum_{k=1}^{K} a_{k,j} G_{k,j}\right)^2\right), \tag{4.33}$$

where N_0 is the noise power and B is the channel bandwidth.

User j can successfully receive a video packet from channel m if it tunes to channel m (i.e., $b_j^m = 1$) and the BS and RNs transmit on channel m (i.e., with probability $P_m^D(\vec{\Theta}_L^m)$). The probability that user j successfully receives a video packet, denoted as P_j^t, is

$$P_j^t = \sum_{m \in \mathcal{A}(t)} b_j^m P_m^D(\vec{\Theta}_L^m). \tag{4.34}$$

Therefore, we can expand the expectation in (4.30) to obtain a reformulated problem:

$$\begin{aligned} \text{maximize:} \quad & \sum_{j=1}^{N} \mathbb{E}\left[P_j^t \log(w_j^t + \lambda_j^t) + (1 - P_j^t)\log(w_j^t)\right] \\ \text{subject to:} \quad & \text{constraints (4.24), (4.25), (4.27), and (4.32).} \end{aligned} \tag{4.35}$$

4.4.3 Solution Algorithms

In this section, we develop effective solution algorithms to the stochastic programming problem (4.30). In Section 4.4.3, we first consider the case of a single licensed channel, and derive a distributed, optimal algorithm with guaranteed convergence and bounded convergence speed. We then address the case of multiple licensed channels. If channel bonding/aggregation techniques are used [44, 70], the distributed algorithm in Section 4.4.3 can still be applied to achieve optimal solutions. We finally consider the case of multiple licensed channels without channel bonding, and develop a greedy algorithm with a performance lower bound in Section 4.4.3.

Case of a Single Channel

Property Consider the case when there is only one licensed channel, i.e., when $M = 1$. The K transmitters, including the BS and $(K-1)$ RNs, send video packets to active users using the licensed channel when it is sensed idle.

Definition 4.1. *A set of vectors is* linearly independent *if none of them can be written as a linear combination of the other vectors in the set [77].*

For user j, the weight and channel gain vectors are: $\vec{a}_j = [a_{1,j}, a_{2,j}, \cdots, a_{K,j}]^{\mathsf{T}}$ and $\vec{G}_j = [G_{1,j}, G_{2,j}, \cdots, G_{K,j}]^{\mathsf{T}}$, where T denotes *matrix transpose*. Due to spatial diversity, we assume that the $\vec{G}_j$ vectors are linearly independent [74].

Lemma 4.1. *To successfully decode each signal X_j, $j = 1, 2, \cdots, N$, the number of active users N should be smaller than or equal to the number of transmitters K.*

Proof. From (4.24), it can be seen that $\vec{a}_j$ is orthogonal to the $(N-1)$ vectors $\vec{G}_n$'s, for $n \neq j$. Since $\vec{a}_j$ is a K by 1 vector, there are at most $(K-1)$ vectors that are orthogonal to $\vec{a}_j$. Since the $\vec{G}_j$ vectors are linearly independent, it follows that $(N-1) \leq (K-1)$ and therefore $N \leq K$. □

According to Lemma 4.1, the following additional constraints should be enforced for the channel selection variables.

$$\sum_{j=1}^{N} b_j^m \leq K, \quad \text{for all } m \in \mathcal{A}(t). \tag{4.36}$$

That is, the number of active users receiving from any channel m cannot be more than the number of transmitters on that channel, which is K in the single channel case and less than or equal to K in the multiple channels case. We first assume that N is not greater than K, and will remove this assumption in the following subsection.

Reformulation and Complexity Reduction With a single channel, all active users receive from channel 1. Therefore $b_j^1 = 1$, and $b_j^m = 0$, for $m > 1$, $j = 1, 2, \cdots, N$. The formulated problem is now reduced to a nonlinear programming problem with constraints (4.24), (4.25), and (4.32). If the number of active users is $N = 1$, the solution is straightforward: all the transmitters send the same signal X_1 to the single user using their maximum transmit power P_{max}.

In general, the reduced problem can be solved with the dual decomposition technique [69] (i.e., a primal dual algorithm). This problem has $K \times N$ primal variables (i.e., the $a_{k,j}$'s), and we need to define $N(N-1)$ dual variables (or, Lagrangian Multipliers) for constraints (4.24) and K dual variables for constraints (4.25). These numbers could be large for even moderate-sized

systems. Before presenting the solution algorithm, we first derive a reformulation of the original problem (4.35) that can greatly reduce the number of primal and dual variables, such that the computational complexity can be reduced.

Lemma 4.2. *Each vector $\vec{a}_j = [a_{1,j}, a_{2,j}, \cdots, a_{K,j}]^\mathsf{T}$ can be represented by the linear combination of r nonzero, linearly independent vectors, where $r = K - N + 1$.*

Proof. From (4.24), each vector $\vec{a}_j$ is orthogonal to $\vec{G}_i$ where $j \neq i$. Define a reduced matrix $\mathbf{G}_{-j}$ obtained by deleting $\vec{G}_j$ from $\mathbf{G}$, i.e., $\mathbf{G}_{-j} = [\vec{G}_1, \cdots, \vec{G}_{j-1}, \vec{G}_{j+1}, \cdots, \vec{G}_N]$. Then $\vec{a}_j$ is a solution to the homogeneous linear system $\mathbf{G}_{-j}^\mathsf{T}\vec{x} = 0$. Since we assume that the $\vec{G}_i$'s are all linearly independent, the columns of $\mathbf{G}_{-j}$ are also linearly independent [77]. Thus the rank of $\mathbf{G}_{-j}$ is $(N-1)$. The solution belongs to the null space of $\mathbf{G}_{-j}$. The dimension of the null space is $r = K-(N-1)$ according to the Rank-nullity Theorem [77]. Therefore, each $\vec{a}_j$ can be presented by the linear combination of r linearly independent vectors. □

Let $\mathbf{e}_j = \{\vec{e}_{j,1}, \vec{e}_{j,2}, \cdots, \vec{e}_{j,r}\}$ be a *basis* for the null space of $\mathbf{G}_{-j}$. There are many methods to obtain the basis, such as Gaussian Elimination. However, we show that it is not necessary to solve the homogeneous linear system $\mathbf{G}_{-j}^\mathsf{T}\vec{x} = 0$ to get the basis for every different j value. Therefore the computational complexity can be further reduced.

Our algorithm for computing a basis is shown in Table 4.3. In Steps 1–6, we first solve the homogeneous linear system $\mathbf{G}^\mathsf{T}\vec{x} = 0$ to get a basis $[\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_{K-N}]$. Note that if K is equal to N, the basis is the empty set $\emptyset$. We then set the $K - N$ basis vectors to be the first $K - N$ vectors in all the basses $\mathbf{e}_j$, $j = 1, 2, \cdots, N$. In Step 8, we orthogonalize each $\mathbf{G}_{-j}$ and obtain $(N-1)$ orthogonal vectors $\vec{\omega}_{j,i}$, $i = 1, 2, \cdots, N-1$. Finally in Step 9, we let the rth vector $\vec{e}_{j,r}$ be orthogonal to all the $\vec{\omega}_{j,i}$'s by subtracting all the projections on each $\vec{\omega}_{j,i}$ from $\vec{G}_j$ (recall that $r = K - N + 1$). The operation is:

$$\vec{e}_{j,r} = \vec{e}_{j,N-K+1} = \vec{G}_j - \sum_{i=1}^{N-1} \frac{\vec{G}_j^\mathsf{T}\vec{\omega}_{j,i}}{\vec{\omega}_{j,i}^\mathsf{T}\vec{\omega}_{j,i}}\vec{\omega}_{j,i}. \tag{4.37}$$

Table 4.3: Basis Computation Algorithm

1:	IF ($K > N$)
2:	Solve homogeneous linear system $\mathbf{G}^{\mathsf{T}}\vec{x} = 0$ and get basis $[\vec{v}_1, \cdots, \vec{v}_{K-N}]$;
3:	FOR $i = 1$ to $K - N$
4:	$\vec{e}_{j,i} = \vec{v}_i$, for all j;
5:	END FOR
6:	END IF
7:	FOR $j = 1$ to N
8:	Orthogonalize $\mathbf{G}_{-j}$ and get $(N-1)$ orthogonal vectors $\vec{w}_{j,i}$'s;
9:	Calculate $\vec{e}_{j,r}$ as in (4.37);
10:	END FOR

Lemma 4.3. *The solution space constructed by the basis $[\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_{K-N}]$ is a sub-space of the solution space of $\mathbf{G}_{-j}^{\mathsf{T}}\vec{x} = 0$ for all j.*

Proof. It is easy to see that each vector $\vec{v}_i$ is a solution of $\mathbf{G}_{-j}^{\mathsf{T}}\vec{x} = 0$ by substituting $\vec{x}$ with $\vec{v}_i$, for $i = 1, 2, \cdots, K - N$. □

Lemma 4.4. *The vectors $[\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_{K-N}, \vec{e}_{j,r}]$ computed in Table 4.3 is a basis of the null space of $\mathbf{G}_{-j}$.*

Proof. Obviously, the $\vec{v}_i$'s are linearly independent. From (4.37), it is easy to verify that $\vec{e}_{j,r}$ is orthogonal to all the $\vec{\omega}_{j,i}$'s. Therefore, $\vec{e}_{j,r}$ is also a solution to system $\mathbf{G}_{-j}^{\mathsf{T}}\vec{x} = 0$. Since $\vec{G}_j$ and $\vec{\omega}_{j,i}$ are orthogonal to all the $\vec{v}_i$'s, and $\vec{e}_{j,r}$ is a linear combination of $\vec{G}_j$ and $\vec{\omega}_{j,i}$, $\vec{e}_{j,r}$ is also orthogonal and linearly independent to all the $\vec{v}_i$'s. The conclusion follows. □

Define coefficients $\vec{c}_j = [c_{j,1}, c_{j,2}, \cdots, c_{j,r}]^{\mathsf{T}}$. Then we can represent $\vec{a}_j$ as a linear combination of the basis vectors, i.e., $\vec{a}_j = \sum_{l=1}^{r} c_{j,l}\vec{e}_{j,l} = \mathbf{e}_j\vec{c}_j$. Eq. (4.33) can be rewritten as

$$\lambda_j^t = \frac{\beta_j B}{T}\log_2\left(1 + \frac{1}{N_0}\left(\vec{c}_j^{\mathsf{T}}\mathbf{e}_j^{\mathsf{T}}\vec{G}_j\right)^2\right) = \frac{\beta_j B}{T}\log_2\left(1 + \frac{1}{N_0}\left(c_{j,r}\vec{e}_{j,r}^{\mathsf{T}}\vec{G}_j\right)^2\right). \tag{4.38}$$

The second equality is because the first $K - N$ column vectors in $\mathbf{e}_j$ are orthogonal to G_j. The random variable W_j^t in the objective function now only depends on $c_{j,r}$. The peak power constraint

Table 4.4: Comparison of Computational Complexity

	Original Problem	Reformulated Problem
Primal Variables	KN	$(K-N+1)N$
Dual Variables	$N(N-1)+K$	K

can be revised as:

$$\sum_{j=1}^{N}[\mathbf{e}_j(k)\vec{c}_j]^2 \leq P_{max}, \quad k=1,\cdots,K, \tag{4.39}$$

where $\mathbf{e}_j(k)$ is the kth row of matrix $\mathbf{e}_j$.

With such a reformulation, the number of primal and dual variables can be greatly reduced. In Table 4.4, we show the numbers of variables in the original problem and in the reformulated problem. The number of primary variables is reduced from KN to $(K-N+1)N$, and the number of dual variables is reduced from $N(N-1)+K$ to K. Such reductions result in greatly reduced computational complexity.

Distributed Algorithm To solve the reformulated problem, we define non-negative dual variables $\vec{\mu}=[\mu_1,\cdots,\mu_K]^{\mathsf{T}}$ for the inequality constraints. The Lagrangian function is

$$\begin{aligned}\mathcal{L}(\mathbf{c},\vec{\mu}) &= \sum_{j=1}^{N}\mathbb{E}\left[\log(W_j^t(c_{j,r}))|w_j^t\right]+\sum_{k=1}^{K}\mu_k(P_{max}-\sum_{j=1}^{N}[\mathbf{e}_j(k)\vec{c}_j]^2)\\ &= \sum_{j=1}^{N}\mathcal{L}_j(\vec{c}_j,\vec{\mu})+P_{max}\sum_{k=1}^{K}\mu_k,\end{aligned} \tag{4.40}$$

where $\mathbf{c}$ is a matrix consisting of all column vector $\vec{c}_j$'s and

$$\mathcal{L}_j(\vec{c}_j,\vec{\mu})=\mathbb{E}\left[\log(W_j^t(c_{j,r}))|w_j^t\right]-\sum_{k=1}^{K}\mu_k[\mathbf{e}_j(k)\vec{c}_j]^2.$$

The corresponding problem can be decomposed into N sub-problems and solved iteratively [69]. In Step $\tau \geq 1$, for given vector $\vec{\mu}(\tau)$, each CR user solves the following sub-problem using

local information

$$\vec{c}_j(\tau) = \arg\max \mathcal{L}_j(\vec{c}_j, \vec{\mu}(\tau)). \tag{4.41}$$

Obviously, the objective function in (4.41) is concave. Therefore, there is a unique optimal solution. The CR users then exchange their solutions over the common control channel. To solve the primal problem, we adopt the gradient method [69].

$$\vec{c}_j(\tau+1) = \vec{c}_j(\tau) + \phi \nabla \mathcal{L}_j(\vec{c}_j(\tau), \vec{\mu}(\tau)), \tag{4.42}$$

where $\nabla \mathcal{L}_j(\vec{c}_j(\tau), \vec{\mu}(\tau))$ is the gradient of the primal problem and ϕ is a small positive step size.

The master dual problem for a given $\mathbf{c}(\tau)$ is:

$$\min_{\mu_i \geq 0, i=1,\cdots,K} q(\vec{\mu}) = \sum_{j=1}^{N} \mathcal{L}_j(\vec{c}_j(\tau), \vec{\mu}) + P_{max} \sum_{k=1}^{K} \mu_k. \tag{4.43}$$

Since the Lagrangian function is differentiable, the subgradient iteration method can be adopted.

$$\vec{\mu}(\tau+1) = [\vec{\mu}(\tau) - \rho(\tau)\vec{g}(\tau)]^+, \tag{4.44}$$

where $\rho(\tau) = \frac{q(\vec{\mu}(\tau)) - q(\vec{\mu}^*)}{||\vec{g}(\tau)||^2}$ is a positive step size, $\vec{\mu}^*$ is the optimal solution, $\vec{g}(\tau) = \nabla q(\vec{\mu}(\tau))$ is the gradient of the dual problem, and $[\cdot]^+$ denotes the projection onto the nonnegative axis. Since the optimal solution $\vec{\mu}^*$ is unknown a priori, we choose the mean of the objective values of the primal and dual problems as an estimate for $\vec{\mu}^*$ in the algorithm. The updated $\mu_k(\tau+1)$ will again be used to solve the sub-problems (4.41). Since the problem is convex, we have strong duality; the duality gap between the primal and dual problems will be zero. The distributed algorithm is shown in Table 4.5, where $0 \leq \kappa \ll 1$ is a threshold for convergence.

Table 4.5: Algorithm for the Case of a Single Channel

1:	IF ($N = 1$)
2:	Set $a_{k,j}$ to P_{max} for all k;
3:	ELSE
4:	Set $\tau = 0$; $\vec{\mu}(0)$ to positive values; $\mathbf{c}(0)$ to random values;
5:	Compute bases $\mathbf{e}_j$'s as in Table 4.3;
6:	DO
7:	$\tau = \tau + 1$;
8:	Compute $\vec{c}_j(\tau)$ as in (4.42);
9:	Broadcast $\vec{c}_j(\tau)$ on the common control channel;
10:	Update $\vec{\mu}(\tau)$ as in (4.44);
11:	WHILE ($\|\|\vec{\mu}(\tau) - \vec{\mu}(\tau-1)\|\| > \kappa$);
12:	Compute $a_{k,j}$'s;
13:	END IF

Performance Analysis We analyze the performance of the distributed algorithm in this section. In particular, we prove that it converges to the optimal solution at a speed faster than $\sqrt{1/\tau}$ as τ goes to infinity.

Theorem 4.1. *The series $q(\vec{\mu}(\tau))$ converges to $q(\vec{\mu}^*)$ as τ goes to infinity and the square error sum $\sum_{\tau=1}^{\infty}(q(\vec{\mu}(\tau)) - q(\vec{\mu}^*))^2$ is bounded.*

Proof. For the optimality gap, we have:

$$\begin{aligned}
& ||\vec{\mu}(\tau+1) - \vec{\mu}^*||^2 = ||[\vec{\mu}(\tau) - \rho(\tau)\vec{g}(\tau)]^+ - \vec{\mu}^*||^2 \\
\leq\ & ||\vec{\mu}(\tau) - \rho(\tau)\vec{g}(\tau) - \vec{\mu}^*||^2 \\
=\ & ||\vec{\mu}(\tau) - \vec{\mu}^*||^2 - 2\rho(\tau)(\vec{\mu}(\tau) - \vec{\mu}^*)^{\mathsf{T}}\vec{g}(\tau) + (\rho(\tau))^2||\vec{g}(\tau)||^2 \\
=\ & ||\vec{\mu}(\tau) - \vec{\mu}^*||^2 - 2\rho(\tau)(q(\vec{\mu}(\tau)) - q(\vec{\mu}^*)) + (\rho(\tau))^2||\vec{g}(\tau)||^2.
\end{aligned}$$

Since the step size is $\rho(\tau) = \frac{q(\vec{\mu}(\tau)) - q(\vec{\mu}^*)}{||\vec{g}(\tau)||^2}$, it follows that

$$\begin{aligned}
||\vec{\mu}(\tau+1) - \vec{\mu}^*||^2 &\leq ||\vec{\mu}(\tau) - \vec{\mu}^*||^2 - \frac{(q(\vec{\mu}(\tau)) - q(\vec{\mu}^*))^2}{||\vec{g}(\tau)||^2} \\
&\leq ||\vec{\mu}(\tau) - \vec{\mu}^*||^2 - \frac{(q(\vec{\mu}(\tau)) - q(\vec{\mu}^*))^2}{\hat{g}^2}, \qquad (4.45)
\end{aligned}$$

where $\hat{g}^2$ is an upper bound of $||\vec{g}(\tau)||^2$. Since the second term on the right-hand-side of (4.45) is non-negative, it follows that $\lim_{\tau\to\infty} q(\vec{\mu}(\tau)) = q(\vec{\mu}^*)$.

Summing Inequality (4.45) over τ, we have

$$\sum_{\tau=1}^{\infty}(q(\vec{\mu}(\tau)) - q(\vec{\mu}^*))^2 \leq \hat{g}^2||\vec{\mu}(1) - \vec{\mu}^*||^2.$$

That is, the square error sum is upper bounded. □

Theorem 4.2. *The sequence $\{q(\vec{\mu}(\tau))\}$ converges faster than $\{1/\sqrt{\tau}\}$ as τ goes to infinity.*

Proof. Assume $\lim_{\tau\to\infty} \sqrt{\tau}(q(\vec{\mu}(\tau)) - q(\vec{\mu}^*)) > 0$. Then there is a sufficiently large τ' and a positive number ξ such that $\sqrt{\tau}(q(\vec{\mu}(\tau)) - q(\vec{\mu}^*)) \geq \xi$, for all $\tau \geq \tau'$. Taking the square sum from τ' to ∞, we have:

$$\sum_{\tau=\tau'}^{\infty}(q(\vec{\mu}(\tau)) - q(\vec{\mu}^*))^2 \geq \xi^2 \sum_{\tau=\tau'}^{\infty}\frac{1}{\tau} = \infty. \tag{4.46}$$

Eq. (4.46) contradicts with Theorem 4.1, which states that $\sum_{\tau=1}^{\infty}(q(\vec{\mu}(\tau)) - q(\vec{\mu}^*))^2$ is bounded. Therefore, we have

$$\lim_{\tau\to\infty} \frac{q(\vec{\mu}(\tau)) - q(\vec{\mu}^*)}{1/\sqrt{\tau}} = 0, \tag{4.47}$$

indicating that the convergence speed of $q(\vec{\mu}(\tau))$ is faster than that of $1/\sqrt{\tau}$. □

Case of Multiple Channels with Channel Bonding

When there are multiple licensed channels, we first consider the case where the channel bonding/aggregation techniques are used by the transmitters and CR users [44, 70]. With channel bonding, a transmitter can utilize all the available channels in $\mathcal{A}(t)$ collectively to transmit the mixed signal. We assume that at the end of the sensing phase in each time slot, CR users tune their SDR transceiver to the common control channel to receive the set of available channels $\mathcal{A}(t)$ from the

BS. Then each CR user can receive from all the channels in $\mathcal{A}(t)$ and decode its desired signal from the compound signal it receives.

This case is similar to the case of a single licensed channel. Now all the active CR users receive from the set of available channels $\mathcal{A}(t)$. We thus have $b_j^m = 1$, for $m \in \mathcal{A}(t)$, and $b_j^m = 0$, for $m \notin \mathcal{A}(t)$, $j = 1, 2, \cdots, N$. When all the b_j^m's are determined this way, problem (4.30) is reduced to a nonlinear programming problem with constraints (4.24) and (4.25). The distributed algorithm described in Section 4.4.3 can be applied to solve this reduced problem to get optimal solutions.

Case of Multiple Channels without Channel Bonding

We finally consider the case of multiple channels without channel bonding, where each CR user has a narrow band SDR transceiver and can only receive from one of the channels. We first present a greedy algorithm that leverages the optimal algorithm in Table 4.5 for near-optimal solutions, and then derive a lower bound for its performance.

Greedy Algorithm When $M > 1$, the optimal solution to problem (4.30) depends also on the binary variables b_j^m's, which determines whether user j receives from channel m. Recall that there are two constraints for the b_j^m's: (i) each user can use at most one channel (see (4.27)); (ii) the number of users on the same channel cannot exceed the number of transmitters K (see (4.36)). Let $\vec{b}$ be the channel allocation vector with elements b_j^m's, and $\Phi(\vec{b})$ the corresponding objective value for a given user channel allocation $\vec{b}$.

We take a two-step approach to solve problem (4.30). First, we apply the greedy algorithm in Table 4.6 to choose one available channel in $\mathcal{A}(t)$ for each CR user (i.e., to determine $\vec{b}$). Second, we apply the algorithm in Table 4.5 to obtain a near-optimal solution for the given channel allocation $\vec{b}$.

In Table 4.6, $\vec{v}_j^m$ is a unit vector with 1 for the $[(j-1) \times M + m]$-th element and 0 for all other elements, and $\vec{b} = \vec{b} + \vec{v}_{j'}^{m'}$ indicates choosing channel m' for user j'. In each iteration, the

Table 4.6: Channel Selection Algorithm for the Case of Multiple Channels without Channel Bonding

1:	Initialize $\vec{b}$ to a zero vector, user set $\mathcal{U} = \{1, \cdots, N\}$ and user-channel set $\mathcal{C} = \mathcal{U} \times \mathcal{A}(t)$;
2:	WHILE ($\mathcal{C} \neq \emptyset$)
3:	Find the user-channel pair $\{j', m'\}$, such that $\{j', m'\} = \arg\max_{\{(j,m)\in\mathcal{C}\}}\{\Phi(\vec{b} + \vec{v}_j^m) - \Phi(\vec{b})\}$;
4:	Set $\vec{b} = \vec{b} + \vec{v}_{j'}^{m'}$ and remove j' from $\mathcal{U}$;
5:	IF ($\sum_{j=1}^{N} b_j^{m'} = K$)
6:	Remove m' from $\mathcal{A}(t)$;
7:	END IF
8:	Update user-channel set $\mathcal{C} = \mathcal{U} \times \mathcal{A}(t)$;
9:	END WHILE

user-channel pair (j', m') that can achieve the largest increase in the objective value is chosen, as in Step 3. The complexity of the greedy algorithm in the worst case is $O(K^2M^2)$.

Performance Bound We next analyze the greedy algorithm and derive a lower bound for its performance. Let ν_l be the sequence from the first to the lth user-channel pair selected by the greedy algorithm. The increase in objective value is denoted as:

$$F_l := F(\nu_l, \nu_{l-1}) = \Phi(\nu_l) - \Phi(\nu_{l-1}). \tag{4.48}$$

Sum up (4.48) from 1 to L. We have $\sum_{l=1}^{L} F_l = \Phi(\nu_L)$ since $\Phi(\nu_0) = 0$. Let Ω be the global optimal solution for user-channel allocation. Define π_l as a subset of Ω. For given ν_l, π_l is the subset of user-channel pairs that cannot be allocated due to the conflict with the l-th user channel allocation ν_l (but not conflict with the user-channel allocations in ν_{l-1}).

Lemma 4.5. *Assume the greedy algorithm stops in L steps, we have*

$$\Phi(\Omega) \leq \Phi(\nu_L) + \sum_{l=1}^{L} \sum_{\sigma\in\pi_l} F(\sigma \cup \nu_{l-1}, \nu_{l-1}).$$

Proof. The proof is similar to the proof of Lemma 7 in [13] and is omitted for brevity. □

Theorem 4.3. *The greedy algorithm for channel selection in Table 4.6 can achieve an objective value that is at least* $\frac{1}{|\mathcal{A}(t)|}$ *of the global optimum in each time slot.*

Proof. According to Lemma 4.5, it follows that:

$$\Phi(\Omega) \leq \Phi(\nu_L) + \sum_{l=1}^{L} |\pi_l| F_l \leq \Phi(\nu_L) + (|\mathcal{A}(t)| - 1) \sum_{l=1}^{L} F_l = |\mathcal{A}(t)| \Phi(\nu_L). \tag{4.49}$$

The second inequality is due to the fact that each user can choose at most one channel and there are at most $(|\mathcal{A}(t)| - 1)$ pairs in π_l according to the definition. The equality in (4.49) is because $\sum_{l=1}^{L} F_l = \Phi(\nu_L)$. Then we have:

$$\frac{1}{|\mathcal{A}(t)|} \Phi(\Omega) \leq \Phi(\nu_L) \leq \Phi(\Omega). \tag{4.50}$$

The greedy heuristic solution is lower bounded by $1/|\mathcal{A}(t)|$ of the global optimum. □

Define competitive ratio $\chi = \Phi(\nu_L)/\Phi(\Omega) = 1/|\mathcal{A}(t)|$. Assume all the licensed channels have identical utilization η. Since $|\mathcal{A}(t)|$ is a random variable, we take the expectation of χ and obtain:

$$\mathbb{E}[\chi] = \eta^M + \sum_{n=1}^{M} \left(\frac{1}{n} \right) \eta^{M-n} (1 - \eta)^n. \tag{4.51}$$

In Fig. 4.7, we evaluate the impact of channel utilization η and the number of licensed channels M on the competitive ratio. We increase η from 0.05 to 0.95 in steps of 0.05 and increase M from 6 to 12 in steps of 2. The lower bound (4.50) becomes tighter when η is larger or when M is smaller. For example, when $\eta = 0.6$ and $M = 6$, the greedy algorithm solution is guaranteed to be no less than 52.7% of the global optimal. when η is increased to 0.95, the greedy algorithm solution is guaranteed to be no less than 98.3% of the global optimal.

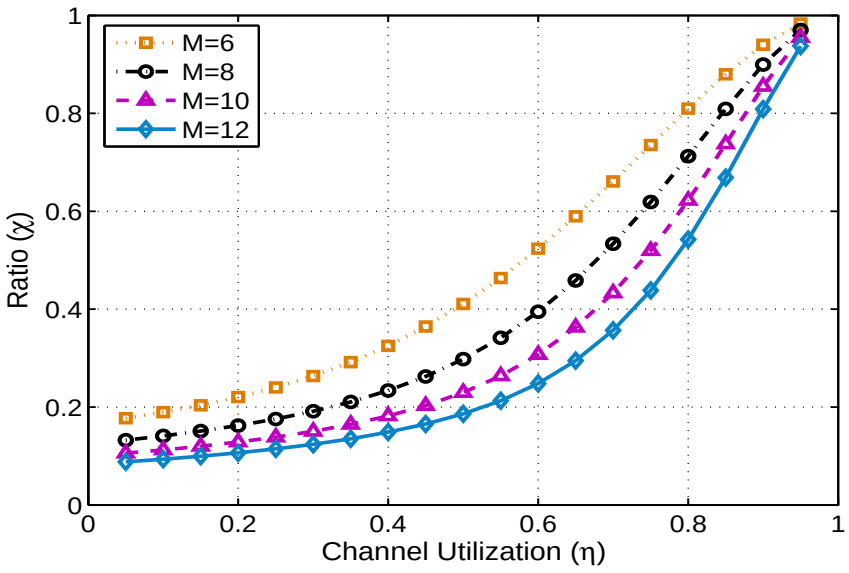


Figure 4.7: Competitive ratio $\mathbb{E}[\chi]$ defined in (4.51) versus channel utilization η.

4.4.4 Performance Evaluation

We evaluate the performance of the proposed algorithms with a MATLAB implementation and the JVSM 9.13 Video Codec. We present simulation results for the following two scenarios: (i) a single licensed channel and (ii) multiple licensed channels without channel bonding, since we observe similar performance for the case of multiple licensed channels with channel bonding. For comparison purpose, we also developed two simpler heuristic schemes that do not incorporate interference alignment.

- *Heuristic 1*: each CR user selects the best channel in $\mathcal{A}(t)$ based on channel condition. The time slot is equally divided among the active users receiving from the same channel, to send their signals separately in each time slice.

- *Heuristic 2*: in each time slot, the active user with the best channel is selected for each available channel. The entire time slot is used to transmit this user's signal.

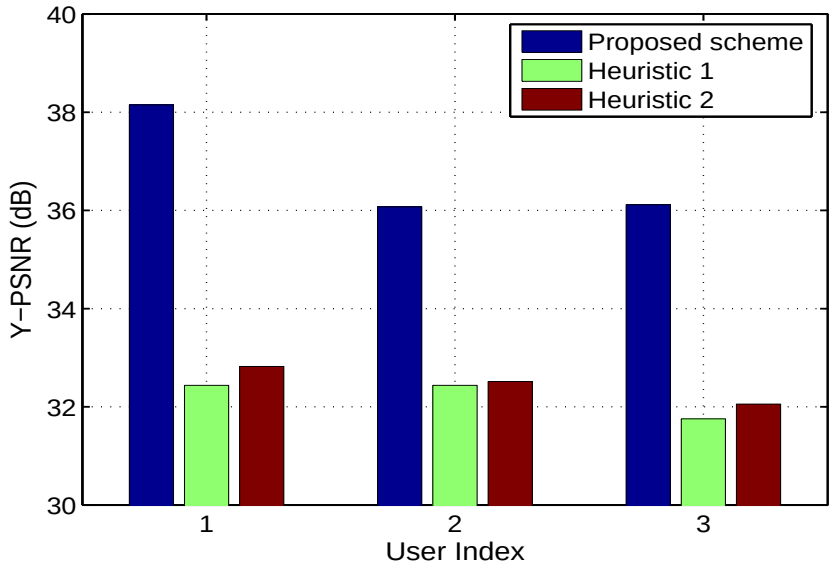


Figure 4.8: Received video quality for each CR user with a single channel.

Case of a Single Licensed Channel

In the first scenario, there are $K = 4$ transmitters, i.e., one BS and three RNs. The channel utilization η is set to 0.6 and the maximum allowable collision probability γ is set to 0.2. There are three active CR users, each receives an MGS video stream from the BS: *Bus* to CR user 1, *Mobile* to CR user 2, and *Harbor* to CR user 3. The video sequences are in the Common Intermediate Format (CIF, 252×288). The GOP size of the videos is 16 and the delivery deadline T is 10. The false alarm probability is $\epsilon_l^m = 0.3$ and the miss detection probability is $\delta_l^m = 0.3$ for all spectrum sensors. The channel bandwidth B is 1 MHz. The peak power limit is 10 W for all the transmitters, unless otherwise specified.

We first plot the average Y-PSNRs of the three reconstructed MGS videos in Fig. 4.8, i.e., only the Y (Luminance) component of the original and reconstructed videos are used. Among three schemes, the proposed algorithm achieves the highest PSNR value, while the two heuristic algorithms have similar performance. Note that the proposed algorithm is optimal in the single channel case. It achieves significant improvements ranging from 3.1 dB to 5.25 dB over the two heuristic algorithms. Such PSNR gains are considerable, since in video coding and communications, a half dB gain is distinguishable and worth pursing.

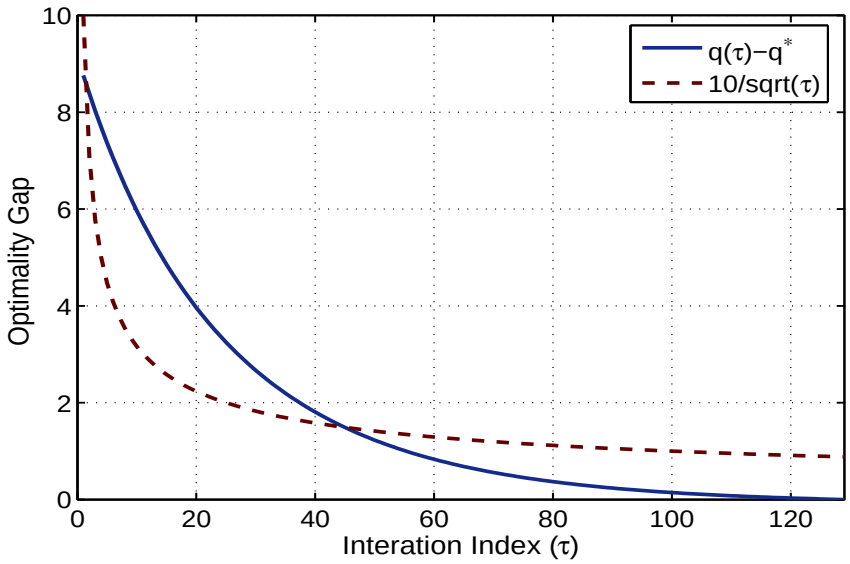


Figure 4.9: Convergence rate of the distributed algorithm with a single channel.

We next examine the convergence rate of the distributed algorithm. According to Theorem 4.2, the distributed algorithm converges at a speed faster than $1/\sqrt{\tau}$ asymptotically. We compare the optimality gap of the proposed algorithm, i.e., $|q(\tau) - q^*|$, with series $10/\sqrt{\tau}$ in Fig. 4.9. Both curves converge to 0 as τ goes to infinity. It can be seen that the convergence speed, i.e., the slope of the curve, of the proposed scheme is larger than that of $10/\sqrt{\tau}$ after about 10 iterations. The convergence of the optimality gap is much faster than $10/\sqrt{\tau}$, which exhibits a heavy tail.

In the case of multiple channels with channel bonding, the performance of the proposed algorithm is similar to that in the single channel case. We omit the results for lack of space.

Case of Multiple Channels without Channel Bonding

We next investigate the second scenario with six licensed channels and four transmitters. There are 12 CR users, each streaming one of the three different videos *Bus*, *Mobile*, and *Harbor*. The rest of the parameters are the same as those in the single channel case, unless otherwise specified. Eq. (4.49) can also be interpreted as an upper bound on the global optimal, i.e., $\Phi(\Omega) \leq |\mathcal{A}(t)|\Phi(\nu_L)$, which is also plotted in the figures. Each point in the following figures is the average

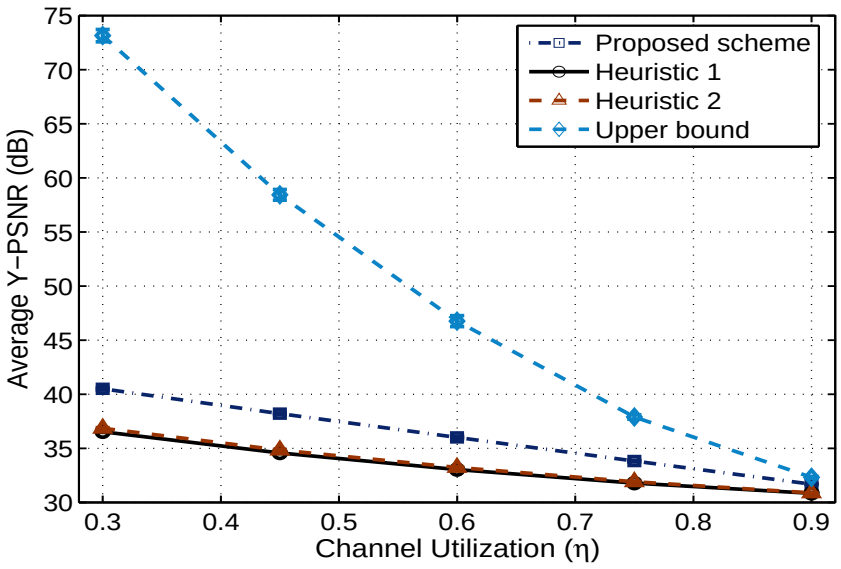


Figure 4.10: Reconstructed video quality vs. channel utilization η in the multi-channel without channel bonding case.

of 10 simulation runs with different random seeds. The 95% confidence intervals are plotted as error bars, which are generally negligible.

The impact of channel utilization η on received video quality is presented in Fig. 4.10. We increase η from 0.3 to 0.9 in steps of 0.15, and plot the Y-PSNRs of reconstructed videos averaged over all the 12 CR users. Intuitively, a smaller η allows more transmission opportunities for CR users, thus allowing the CR users to achieve higher video rates and better video quality. This is shown in the figure, in which all four curves decrease as η is increased. We also observe that the gap between the upper bound and proposed schemes becomes smaller as η gets larger, from 32.65 dB when $\eta = 0.3$ to 0.63 dB when $\eta = 0.9$. This trend is also demonstrated in Fig. 4.7. The proposed scheme outperforms the two heuristic schemes with considerable gains, ranging from 0.8 dB to 3.65 dB.

Finally, we investigate the impact of the number of transmitters K on the video quality. In this simulation we increase K from 2 to 6 with step size 1. The average Y-PSNRs of all the 12 CR users are plotted in Fig. 4.11. As expected, the more transmitters, the more effective the interference alignment technique, and thus the better the video quality. The proposed algorithm

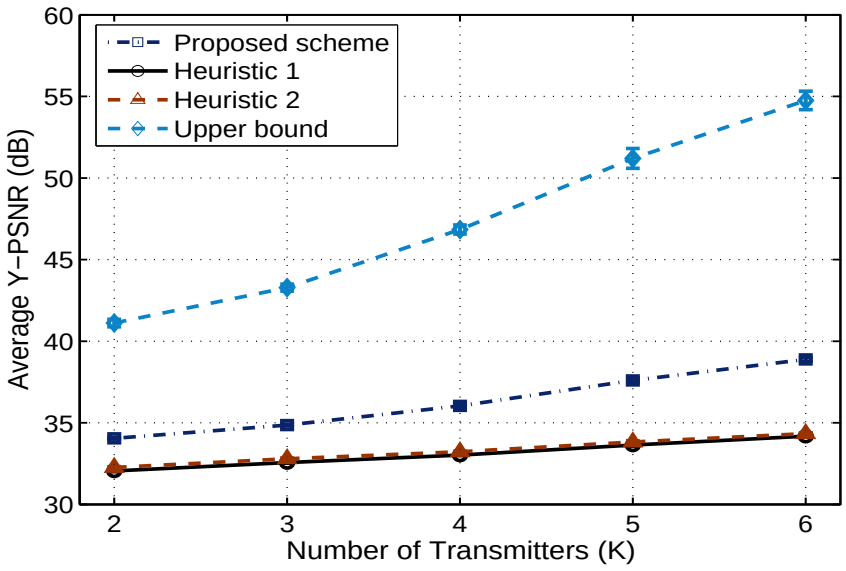


Figure 4.11: Reconstructed video quality vs. number of transmitters K in the multi-channel without channel bonding case.

achieves gains ranging from 1.78 dB (when $K = 2$) to 4.55 dB (when $K = 6$) over the two heuristic schemes.

4.5 Conclusions

In this chapter, we first studied the problem of cooperative relay in CR networks. We modeled the two cooperative relay strategies, i.e., DF and AF, which are integrated with p-Persistent CSMA. We analyzed their throughput performance and compared them under various parameter ranges. Cross-point with the AF and DF curves are found when some parameter is varied, indicating that each of them performs better in a certain parameter range; there is no case of dominance for the two strategies. Considerable gains were observed over conventional DL transmissions, as achieved by exploiting cooperative diversity with the cooperative relays in CR networks.

Then, we investigated the problem of interference alignment for MGS video streaming in a cooperative relay enhanced CR network. We presented a stochastic programming formation, and derived a reformulation that leads to considerable reduction in computational complexity. A

distributed optimal algorithm was developed for the case of a single channel and the case of multi-channel with channel bonding, with proven convergence and convergence speed. We also presented a greedy algorithm for the multi-channel without channel bonding case, with a proven performance bound. The proposed algorithms are evaluated with simulations and are shown to outperform two heuristic schemes without interference alignment with considerable gains.

Chapter 5

CR Femtocell Networks

5.1 Introduction

Due to the use of open space as transmission medium, capacity of wireless networks are usually limited by interference. When a mobile user moves away from the base station, a considerably larger transmit power is needed to overcome attenuation, while causing interference to other users and deteriorating network capacity. To this end, femtocells provide an effective solution that brings network infrastructure closer to mobile users. A femtocell is a small (e.g., residential) cellular network, with a *femto base station* (FBS) connected to the owner's broadband wireline network [5, 92, 93]. The FBS serves approved users when they are within the coverage. Among the many benefits, femtocells are shown effective on improving network coverage and capacity [5]. Due to reduced distance, transmit power can be greatly reduced, leading to prolonged battery life, improved signal-to-interference-plus-noise ratio (SINR), and better spatial reuse of spectrum.

Femtocells have received significant interest from the wireless industry. Although highly promising, many important problems should be addressed to fully harvest their potential, such as interference mitigation, resource allocation, synchronization, and QoS provisioning [5, 92]. It is also critical for the success of this technology to support important applications such as real-time video streaming in femtocell networks.

In this chapter, we first investigate the problem of data multicast in femtocell networks. It is not atypical that many users may request for the same content, as often observed in wireline networks. By allowing multiple users to share the same downlink multicast transmission, significant spectrum and power savings can be achieved.

In particular, we adopt *superposition coding* (SC) and *successive interference cancellation* (SIC), two well-known PHY techniques, for data multicast in femtocell networks [94]. With SC,

a compound signal is transmitted, consisting of multiple signals (or, layers) from different senders or from the same sender. With SIC, a strong signal can be first decoded, by treating all other signals as noise. Then the decoder will reconstruct the signal from the decoded bits, and subtract the reconstructed signal from the compound signal. The next signal will be decoded from the residual, by treating the remaining signals as noise. And so forth. A special strength of the SC with SIC approach is that it enables simultaneous unicast transmissions (e.g., many-to-one or one-to-many). It has been shown that SC with SIC is more efficient than PHY techniques with orthogonal channels [94, 95].

We adopt SC and SIC for the unique femtocell network environment, and investigate how to enable efficient data multicast from the femtocells to multiple users. We formulate a Mixed Integer Nonlinear Programming (MINLP) problem, which is NP-hard in general. The objective is to minimize the total BS power consumption. Then we reformulate the MINLP problem into a simpler form, and derive upper and lower performance bounds. We also derive a simple heuristic scheme that assigns users to the BS's with a greedy approach. Finally, we consider three typical connection scenarios in the femtocell network, and develop optimal and near-optimal algorithms for the three scenarios. The proposed algorithms have low computational complexity, and are shown to outperform the heuristic scheme with considerable gains.

Then, we investigate the problem of video streaming in femtocell cognitive radio (CR) networks. We consider a femtocell network consisting of a *macro base station* (MBS) and multiple FBS's. The femtocell network is co-located with a primary network with multiple licensed channels. This is a challenging problem due to the stringent QoS requirements of real-time videos and, on the other hand, the new dimensions of network dynamics (i.e., channel availability) and uncertainties (i.e., spectrum sensing and errors) found in CR networks.

We adopt Scalable Video Coding (SVC) in our system. SVC encodes a video into multiple substreams, subsets of which can be decoded to provide different quality levels for the reconstructed video [53]. Such scalability is very useful for video streaming systems, especially in

CR networks, to accommodate heterogeneous channel availabilities and dynamic network conditions. We consider H.264/SVC medium grain scalable (MGS) videos, since MGS can achieve better rate-distortion performance over Fine-Granularity-Scalability (FGS), although it only has Network Abstraction Layer (NAL) unit-based granularity [53].

The unique femtocell network architecture and the scalable video allow us to develop a framework that captures the key design issues and trade-offs, and to formulate a *stochastic programming* problem. It has been shown that the deployment of femtocells has a significant impact on the network performance [5]. In this paper, we examine three deployment scenarios. In the case of a single FBS, we apply *dual decomposition* to develop a distributed algorithm that can compute the optimal solution. In the case of multiple non-interfering FBS's, we show that the same distributed algorithm can be used to compute optimal solutions. In the case of multiple interfering FBS's, we develop a greedy algorithm that can compute near-optimal solutions, and prove a closed-form lower bound for its performance based on an *interference graph* model. The proposed algorithms are evaluated with simulations, and are shown to outperform three alternative schemes with considerable gains.

The remainder of this chapter is organized as follows. The related work is discussed in Section 5.2. We investigate the problem of data multicast over fenmtocell networks in Section 5.3. The problem of streaming multiple MGS videos in a femtocell CR network is discussed in Section 5.4. Section 5.5 concludes this paper.

5.2 Background and Related Work

Femtocells have attracted considerable interest from both industry and academia. Technical and business challenges, requirements and some preliminary solutions to femtocell networks are discussed in [5]. Since FBS's are distributedly located and are able to spatially reuse the same channel, considerable research efforts were made on interference analysis and mitigation [35, 96].

A distributed utility based SINR adaptation scheme was presented in [96] to alleviate cross-tire interference at the macrocell from co-channel femtocells. Lee, Oh and Lee [35] proposed a fractional frequency reuse scheme to mitigate inter-femtocell interference.

Deploying femtocells by underlaying the macrocell has been proved to significantly improve indoor coverage and system capacity. However, interference mitigation in a two-tier heterogeneous network is a challenging problem. In [97], the interference from macrocell and femtocells was mitigated by a spatial channel separation scheme with codeword-to-channel mapping. In [98], the rate distribution in the macrocell was improved by subband partitioning and modest gains were achieved by interference cancellation. In [99], the interference was controlled by denying the access of femtocell base stations to protect the transmission of nearby macro base station. A novel algorithmic framework was presented in [100] for dynamic interference management to deliver QoS, fairness and high system efficiency in LTE-A femtocell networks. Requiring no modification of existing macrocells, CR was shown to achieve considerable performance improvement when applied to interference mitigation [101]. In [102], the orthogonal time-frequency blocks and transmission opportunities were allocated based on a safe/victim classification.

SIC has high potential of sending or receiving multiple signals concurrently, which improves the transmission efficiency. In [95], the authors developed MAC and routing protocols that exploit SC and SIC to enable simultaneous unicast transmissions. Sen, et al. investigated the possible throughput gains with SIC from a MAC layer perspective [103]. Power control for SIC was comprehensively investigated and widely applied to code division multiple access (CDMA) systems [104–108]. Applying game theory, Jean and Jabbari proposed an uplink power control under SIC in direct sequence-CDMA networks [104]. In [105], the authors introduced an iterative two-stage SIC detection scheme for a multicode MIMO system and showed the proposed scheme significantly outperformed the equal power allocation scheme. A scheme on joint power control and receiver optimization of CDMA transceivers was presented in [106]. In [107, 108], the impact of imperfect channel estimation and imperfect interference cancellation on the capacity of CDMA systems was examined.

5.3 Multicast in Femtocell Networks with Superposition Coding and Successive Interference Cancellation

In this section, we formulate a Mixed Integer Nonlinear Programming (MINLP) problem of data multicast in femotcell networks, which is NP-hard in general. Then we reformulate the MINLP problem into a simpler form, and derive upper and lower performance bounds. We also derive a simple heuristic scheme that assigns users to the BS's with a greedy approach. Finally, we consider three typical connection scenarios in the femtocell network, and develop optimal and near-optimal algorithms for the three scenarios. The proposed algorithms have low computational complexity, and are shown to outperform the heuristic scheme with considerable gains.

5.3.1 System Model and Problem Statement

System Model

Consider a femtocell network with an MBS (indexed 0) and M FBS's (indexed from 1 to M) deployed in the area. The M FBS's are connected to the MBS and the Internet via broadband wireline connections. Furthermore, we assume a spectrum band that is divided into two parts, one is allocated to the MBS with bandwidth B_0 and the other is allocated to the M FBS's. The bandwidth allocated to FBS m is denoted by B_m. When there is no overlap between the coverages of two FBS's, they can spatially reuse the same spectrum. Otherwise, the MBS allocates disjoint spectrum to the FBS's with overlapping coverages. We assumed the spectrum allocation is known a priori.

There are K mobile users in the femtocell network. Each user is equipped with one transceiver that can be tuned to one of the two available channels, i.e., connecting to a nearby FBS or to the MBS. The network is time slotted. We assume block-fading channels, where the channel condition is constant in each time slot [94]. We focus on a multicast scenario, where the MBS and FBS's multicast a data file to the K users. The data file is divided into multiple packets with equal length

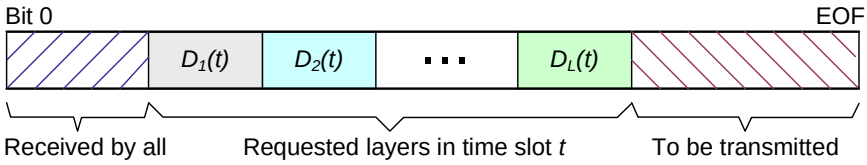


Figure 5.1: Superposition coding and successive interference cancellation.

and transmitted in sequence with the same modulation scheme. Once packet l is successfully received and decoded at the user, it requests packet $(l+1)$ in the next time slot.

We adopt SC and SIC to transmit these packets [94], as illustrated in Fig. 5.1. In each time slot t, the compound signal has L *layers* (or, levels), denoted as $D_1(t)$, $\cdots$, $D_L(t)$. Each level $D_i(t)$, $i = 1, \cdots, L$, is a packet requested by some of the users in time slot t. A user that has successfully decoded $D_i(t)$, for all $i = 1, \cdots, l-1$, is able to subtract these signals from the received compound signal and then decodes $D_l(t)$, while the signals from $D_{l+1}(t)$ to $D_L(t)$ are treated as noise.

Problem Statement

For the SC and SIC scheme to work, the transmit powers for the levels should be carefully determined, such that there is a sufficiently high SNR for the levels to be decodable. It is also important to control the transmit powers of the BS's to reduce interference and leverage frequency reuse. The annual power bill is a large part of a mobile operator's costs [109]. Minimizing BS power consumption is important to reduce not only the operator's OPEX, but also the global CO_2 emission; an important step towards "green" communications.

Therefore, we focus on BS power allocation in this paper. The objective is to minimize the total power of all the BS's, while guaranteeing a target rate R_{tar} for each user. Recall that the data file is partitioned into equal-length packets. The target rate R_{tar} ensures that a packet can be transmitted within a time slot, for given modulation and channel coding schemes.

Define binary indicator I_m^k, for all m and k, as:

$$I_m^k = \begin{cases} 1, & \text{if user } k \text{ connects to BS } m \\ 0, & \text{otherwise.} \end{cases} \tag{5.1}$$

Consider a general time slot t when L data packets, or levels, are requested. We formulate the optimal power allocation problem (termed OPT-Power) as follows.

$$\text{minimize: } \sum_{m=0}^{M}\sum_{l=1}^{L} P_l^m \tag{5.2}$$

$$\text{subject to: } B_m \log_2(1+\gamma_m^k I_m^k) \geq R_{tar} I_m^k, \text{ for all } k \tag{5.3}$$

$$\sum_{m=0}^{M} I_m^k = 1, \text{for all } k \tag{5.4}$$

$$P_l^m \geq 0, \text{ for all } l, m, \tag{5.5}$$

where P_l^m is the power of BS m for transmitting the level l packet; γ_m^k is the SNR at user k if it connects to BS m. Constraint (5.3) guarantees the minimum rate at each user. Constraint (5.4) is due to the fact that each user is equipped with one transceiver, so it can only connect to one BS.

Let $\mathcal{U}_l$ denote the set of users requesting the level l packet. A user $k \in \mathcal{U}_l$ has decoded all the packets up to D_{l-1}. It subtracts the decoded signals from the received signal and treats signals $D_{l+1}, \cdots, D_L$ as noise. The SNR at user $k \in \mathcal{U}_l$, for $l = 1, \cdots, L-1$, can be written as:

$$\gamma_m^k = H_m^k P_l^m / \left(N_0 + H_m^k \sum_{i=l+1}^{L} P_i^m \right), \tag{5.6}$$

where H_m^k is the random channel gain from BS m to user k and N_0 is the noise power. For user $k \in \mathcal{U}_L$ that requests the last packet D_L, the SNR is

$$\gamma_m^k = H_m^k P_L^m / N_0. \tag{5.7}$$

The optimization variables in Problem OPT-Power consist of the binary variables I_m^k's and the continuous variables P_l^m's. It is an MINLP problem, which is NP-hard in general. In Section 5.3.2, we first reformulate the problem to a obtain a simpler form, and then develop effective algorithms for optimal and suboptimal solutions.

5.3.2 Reformulation and Power Allocation

In this section, we reformulate Problem OPT-Power to obtain a simpler form, and derive an upper bound and a lower bound for the total BS power. The reformulation also leads to a simple heuristic algorithm. Finally, we introduce power allocation algorithms for three connection scenarios.

Problem Reformulation

Due to the monotonic logarithm functions and the binary indicators I_m^k, constraint (5.3) can be rewritten as:

$$\gamma_m^k I_m^k \geq \Gamma_m^k I_m^k, \quad m = 0, 1, \cdots, M, \tag{5.8}$$

where $\Gamma_m^k = \Gamma_m =: 2^{R_{tar}/B_m} - 1$ is the minimum SNR requirement at user k that connects to BS m. To further simplify the problem, define $Q_l^m = \sum_{i=l}^{L} P_i^m$, with $Q_{L+1}^m = 0$. Then power P_l^m is the difference

$$P_l^m = Q_l^m - Q_{l+1}^m. \tag{5.9}$$

Problem OPT-Power can be reformulated as:

$$\text{minimize} \sum_{m=0}^{M} Q_1^m \tag{5.10}$$

$$\text{subject to: } H_m^k (Q_l^m - Q_{l+1}^m) / \left(N_0 + H_m^k Q_{l+1}^m\right) I_m^k \geq \Gamma_m I_m^k,$$

$$\text{for all } k \in \mathcal{U}_l, l = 1, \cdots, L \tag{5.11}$$

$$Q_l^m \geq Q_{l+1}^m, l = 1, \cdots, L \tag{5.12}$$

$$\sum_{m=0}^{M} I_m^k = 1, \text{ for all } k. \tag{5.13}$$

For $l \leq L$, constraint (5.11) can be rewritten as:

$$Q_l^m I_m^k \geq \left[N_0 \Gamma_m / H_m^k + (1+\Gamma_m) Q_{l+1}^m\right] I_m^k. \tag{5.14}$$

Let $\mathcal{U}_l^m$ be the subset of users connecting to BS m in $\mathcal{U}_l$. Since $Q_l^m \geq Q_{l+1}^m$, (5.14) can be rewritten as,

$$Q_l^m = \max\left\{Q_{l+1}^m, \max_{k\in\mathcal{U}_l^m}\left[N_0\Gamma_m/H_m^k + (1+\Gamma_m)Q_{l+1}^m\right]\right\}. \tag{5.15}$$

From (5.15), we define a function $Q_l^m = F_m(Q_{l+1}^m, \mathcal{U}_l^m)$ as:

$$F_m(Q_{l+1}^m, \mathcal{U}_l^m) = \begin{cases} Q_{l+1}^m, & \mathcal{U}_l^m = \emptyset \\ \max_{k\in\mathcal{U}_l^m}\left\{\frac{N_0\Gamma_m}{H_m^k} + (1+\Gamma_m)Q_{l+1}^m\right\}, & \mathcal{U}_l^m \neq \emptyset. \end{cases} \tag{5.16}$$

Obviously, $F_m(Q_{l+1}^m, \mathcal{U}_l^m)$ is non-decreasing with respect to Q_{l+1}^m. It follows that

$$\begin{aligned} Q_1^m &= F_m(Q_2^m, \mathcal{U}_1^m) = F_m(F_m(Q_3^m, \mathcal{U}_2^m), \mathcal{U}_1^m) \\ &= F_m(\cdots(F_m(Q_{L+1}^m, \mathcal{U}_L^m), \mathcal{U}_{L-1}^m), \cdots, \mathcal{U}_1^m) \\ &= F_m(\cdots(F_m(0, \mathcal{U}_L^m), \mathcal{U}_{L-1}^m), \cdots, \mathcal{U}_1^m). \end{aligned} \tag{5.17}$$

If none of the subsets $\mathcal{U}_l^m$ ($l = 1, \cdots, L$) is empty, we can expand the above recursive term using (5.16). It follows that

$$Q_1^m = N_0\Gamma_m \sum_{l=1}^{L}(1+\Gamma_m)^{c_l^m} \max_{k\in\mathcal{U}_l^m}\left\{1/H_m^k\right\}, \tag{5.18}$$

where the exponent c_l^m is defined as $c_1^m = 0$ and $c_{l+1}^m = c_l^m + 1$. Otherwise, if a subset $\mathcal{U}_l^m = \emptyset$ for some m, we have that $Q_l^m = Q_{l+1}^m$, $\max_{k\in\mathcal{U}_l^m}\left\{1/H_m^k\right\} = \max_{k\in\emptyset}\left\{1/H_m^k\right\} = 0$, and $c_l^m = c_{l-1}^m$. Eq. (5.18) still holds true.

Finally, the objective function (5.10) can be rewritten as

$$\sum_{m=0}^{M} N_0 \Gamma_m \sum_{l=1}^{L} (1+\Gamma_m)^{c_l^m} \max_{k \in \mathcal{U}_l^m} \left\{ 1/H_m^k \right\}. \tag{5.19}$$

Since $(1+\Gamma_m) > 0$, it can be seen that to minimize the total BS power, we need to keep the c_l^m's as low as possible.

Performance Bounds

The reformulation and simplification allow us to derive performance bounds for the total BS power consumption. First, we derive the upper bound for the objective function (5.10). Define a variable

$$\overline{G}_m = \max_{l \in \{1,\cdots,L\}} \max_{k \in \mathcal{U}_l^m} \left\{ \Gamma_m / H_m^k \right\}, \tag{5.20}$$

which corresponds to the user with the worst channel condition among all users that connect to BS m. It follows that:

$$\begin{aligned}
\sum_{m=0}^{M} Q_1^m &= N_0 \sum_{m=0}^{M} \sum_{l=1}^{L} (1+\Gamma_m)^{c_l^m} \max_{k \in \mathcal{U}_l^m} \left\{ \Gamma_m / H_m^k \right\} \\
&\leq N_0 \sum_{m=0}^{M} \sum_{l=1}^{L} (1+\Gamma_m)^{c_l^m} \overline{G}_m \\
&\leq N_0 \sum_{m=0}^{M} \overline{G}_m \sum_{l=1}^{L} (1+\Gamma_m)^{l-1} \\
&= N_0 \sum_{m=0}^{M} \overline{G}_m \left[(1+\Gamma_m)^L - 1 \right] / \Gamma_m.
\end{aligned} \tag{5.21}$$

In (5.21), the first inequality is from the definition of $\overline{G}_m$. The second inequality is from the definition of c_{l+1}^m. Specifically, $c_1^m = 0$; when $\mathcal{U}_l^m \neq \emptyset$, we have $c_l^m = c_{l-1}^m + 1$; when $\mathcal{U}_l^m = \emptyset$, we have $c_l^m = c_{l-1}^m$. It follows that $c_l^m \leq l-1$. Therefore, (5.21) is an upper bound on the objective function (5.10).

Furthermore, by defining $\overline{G} = \max_{m\in\{0,\cdots,M\}}\left\{\overline{G}_m\right\}$, and $\overline{\Gamma} = \max_{m\in\{0,\cdots,M\}}\{\Gamma_m\}$, we can get a looser upper bound from 5.21 as

$$\sum_{m=0}^{M} Q_1^m \leq N_0\overline{G}(M+1)\left[(1+\overline{\Gamma})^L - 1\right]/\overline{\Gamma}. \tag{5.22}$$

Next, we derive a lower bound for (5.10). Define

$$\begin{cases} \underline{G}^l = \min_{m\in\{0,\cdots,M\}} \max_{k\in\mathcal{U}_l^m}\left\{\Gamma_m/H_m^k\right\} \\ \underline{\Gamma} = \min_{m\in\{0,\cdots,M\}}\{\Gamma_m\}. \end{cases} \tag{5.23}$$

We have that

$$\begin{aligned} \sum_{m=0}^{M} Q_1^m &= N_0\sum_{m=0}^{M}\sum_{l=1}^{L}(1+\Gamma_m)^{c_l^m}\max_{k\in\mathcal{U}_l^m}\left\{\Gamma_m/H_m^k\right\} \\ &\geq N_0\sum_{m=0}^{M}\sum_{l=1}^{L}(1+\Gamma_m)^{c_l^m}\underline{G}^l \\ &\geq N_0\sum_{l=1}^{L}\underline{G}^l\sum_{m=0}^{M}(1+\underline{\Gamma})^{c_l^m} \\ &\geq N_0(M+1)\sum_{l=1}^{L}\underline{G}^l(1+\underline{\Gamma})^{\frac{\sum_{m=0}^{M}c_l^m}{M+1}} \\ &\geq N_0(M+1)\sum_{l=1}^{L}\underline{G}^l(1+\underline{\Gamma})^{\frac{l-1}{M+1}}. \end{aligned} \tag{5.24}$$

In (5.24), the first inequality is from the definition of $\underline{G}^l$. The second inequality is due to the definition of $\underline{\Gamma}$. The third inequality is due to the fact that $(1+\Gamma)^{c_l^m}$ is a convex function. The fourth inequality is because that each level must be transmitted by at least one BS. Thus for each level l, there is at least one $c_l^m = c_{l-1}^m + 1$ for some m. It follows that the sum $\sum_{m=0}^{M} c_l^m$ should be greater than $l-1$. Therefore, (5.24) provides a lower bound for (5.10).

Furthermore, by defining $\underline{G} = \min_{l\in\{1,\cdots,L\}} \left\{\underline{G}^l\right\}$, we can obtain a looser lower bound from (5.24) as

$$\sum_{m=0}^{M} Q_1^m \geq N_0 \underline{G}(M+1)\frac{(1+\underline{\Gamma})^{\frac{L}{M+1}}-1}{(1+\underline{\Gamma})^{\frac{1}{M+1}}-1}. \tag{5.25}$$

A Simple Heuristic Scheme

We first describe a greedy heuristic algorithm that solves OPT-Power with suboptimal solutions. With this heuristic, each user compares the channel gains from the MBS and the FBS's. It chooses the BS with the best channel condition to connect to, thus the values of the binary variables I_m^k are determined. Once the binary variables are fixed, all the subsets $\mathcal{U}_l^m$'s are determined. Starting with $Q_{L+1}^m = 0$, we can apply (5.15) iteratively to find the Q_l^m's. Finally, the transmit powers P_l^m can be computed using (5.9).

With this approach, among the users requesting the level l packet, it is more likely that some of them connect to the MBS and the rest connect to some FBS's, due to the random channel gains in each time slot. In this situation, both MBS and FBS will have to transmit all the requested data packets. Such situation is not optimal for minimizing the total power, as will be discussed in Section 5.3.2.

Power Allocation Algorithms

In the following, we develop three power allocation algorithms for three different connection scenarios with a more structured approach.

Case I–One Base Station We first consider the simplest connection scenario where all the K users connect to the same BS (i.e., either the MBS or an FBS). Assume all the users connect to BS m. Then we have $I_m^k = 1$ for all k, and all the subsets $\mathcal{U}_l^m$ are non-empty; $I_{m'}^k = 0$ for all k and all $m' \neq m$, and all the subsets $\mathcal{U}_l^{m'}$ are empty for $m' \neq m$.

From (5.16), we can derive the optimal solution as:

$$\begin{aligned} Q_l^{m*} &= (1+\Gamma_m)Q_{l+1}^{m*} + \max_{k\in\mathcal{U}_l^m}\left\{N_0\Gamma_m/H_m^k\right\}, \\ &= N_0\Gamma_m\sum_{i=l}^{L}(1+\Gamma_m)^{i-l}\max_{k\in\mathcal{U}_l^m}\left\{1/H_m^k\right\}, \quad l=1,2,\cdots,L. \end{aligned} \tag{5.26}$$

Recall that $Q_{L+1}^{m*} = Q_{L+1}^m = 0$, the optimal power allocation for Problem OPT-Power in this case is:

$$P_l^{m'*} = \begin{cases} Q_l^{m*} - Q_{l+1}^{m*}, & m' = m, \text{ for all } l \\ 0, & m' \neq m, \text{ for all } l. \end{cases} \tag{5.27}$$

Case II–MBS and One FBS We next consider the case with one MBS and one FBS (i.e., $M = 1$), where each user has two choices: connecting to either the FBS or the MBS.

Recall that $\mathcal{U}_l^0$ and $\mathcal{U}_l^1$ are the subset of users who connected to the MBS and the FBS, respectively, and who request the level l packet. Examining (5.18), we find that the total power of BS m can be significantly reduced if one or more levels are not transmitted, since the exponent c_l^m will not be increased in this case. Furthermore, consider the two choices: (i) not transmitting level l, and (ii) not transmitting level $l' > l$ from BS m. The first choice will yield larger power savings, since more exponents (i.e., $c_l^m, c_{l+1}^m, \cdots, c_{l'-1}^m$) will assume smaller values. Therefore, we should let these two subsets be empty whenever possible, i.e., either $\mathcal{U}_l^0 = \emptyset$ or $\mathcal{U}_l^1 = \emptyset$. According to this policy, all the users requesting the level l packet will connect to the same BS. We only need to make the optimal connection decision for each subset of users requesting the same level of packet, rather than for each individual user.

Since not transmitting a lower level packet yields more power savings for a BS, we calculate the power from the lowest to the highest level, and decide whether connecting to the MBS or the FBS for users in each level. Define $G_l^0 = \max_{k\in\mathcal{U}_l}\left\{1/H_0^k\right\}$ and $G_l^1 = \max_{k\in\mathcal{U}_l}\left\{1/H_1^k\right\}$. The algorithm for solving Problem OPT-Power in this case is given in Table 5.1. In Steps 2–10, the decision on whether connecting to the MBS or the FBS is made by comparing the expected increments in the total power. The user subsets $\mathcal{U}_l^0$ and $\mathcal{U}_l^1$ are determined in Steps 4 and 7. In

Table 5.1: Power Allocation Algorithm For Case II

1: Initialize all c_l^0, c_l^1, Q_{L+1}^0 and Q_{L+1}^1 to zero;
2: FOR $l = 1$ TO L
3: IF ($\Gamma_0(1+\Gamma_0)^{c_l^0} G_l^0 \leq \Gamma_1(1+\Gamma_1)^{c_l^1} G_l^0$)
4: Set $\mathcal{U}_l^0 = \mathcal{U}_l$ and $\mathcal{U}_l^1 = \emptyset$;
5: $c_l^0 = c_l^0 + 1$;
6: ELSE
7: Set $\mathcal{U}_l^0 = \emptyset$ and $\mathcal{U}_l^1 = \mathcal{U}_l$;
8: $c_l^1 = c_l^1 + 1$;
9: END IF
10: END FOR
11: FOR $l = L$ TO 1
12: $Q_l^0 = F_0(Q_{l+1}^0, \mathcal{U}_l^0)$ and $P_l^0 = Q_l^0 - Q_{l+1}^0$;
13: $Q_l^1 = F_1(Q_{l+1}^1, \mathcal{U}_l^1)$ and $P_l^1 = Q_l^1 - Q_{l+1}^1$;
14: END FOR

Steps 11–14, Q_l^m's and the corresponding P_l^m's are computed in the reverse order, based on the determined subsets $\mathcal{U}_l^0$ and $\mathcal{U}_l^1$.

The computational complexity of this algorithm is $\mathcal{O}(L)$.

Case III–MBS and Multiple FBS's Finally, we consider the general case with one MBS and multiple FBS's in the network. Each user is able to connect to the MBS or a nearby FBS. Recall that we define $\mathcal{U}_l$ as the set of users requesting the level l packet, and $\mathcal{U}_l^m$ as the subset of users in $\mathcal{U}_l$ that *connect* to BS m. These sets have the following properties.

$$\begin{cases} \bigcup_{m=0}^{M} \mathcal{U}_l^m = \mathcal{U}_l \\ \mathcal{U}_l^m \bigcap \mathcal{U}_l^{m'} = \emptyset, \quad \text{for all } m' \neq m. \end{cases}$$

The first property is due to the fact that each user must connect to the MBS or an FBS. The second property is because each user can connect to only one BS. The user subsets connecting to different BS's do not overlap. Therefore, $\mathcal{U}_l^m$'s is a *partition* of $\mathcal{U}_l$ with respect to m.

In addition, we define $\mathcal{S}_l^m$ as the set of possible users that are *covered* by BS m and request the level l packet. These sets have the following properties.

$$\begin{cases} \bigcup_{m=1}^{M} \mathcal{S}_l^m = \mathcal{S}_l^0 = \mathcal{U}_l \\ \mathcal{S}_l^m \bigcap \mathcal{S}_l^0 = \mathcal{S}_l^m, \text{ for all } m \neq 0 \\ \mathcal{S}_l^m \bigcap \mathcal{S}_l^{m'} = \emptyset, \text{ for all } m' \neq m \text{ and } m, m' \neq 0. \end{cases}$$

The first property is because all users in each femtocell are covered by the MBS. The second property indicates that the users covered by FBS m are a subset of the users covered by the MBS. The third property shows that the user subsets in different femtocells do not overlap. We can see that the $\mathcal{S}_l^m$'s, for $m = 1, \cdots, M$, are also a partition of $\mathcal{U}_l$.

Define $W_m(\mathcal{U}) = \max_{k \in \mathcal{U}} \left\{1/H_m^k\right\}$, where $\mathcal{U}$ is the set of users and $m = 0, \cdots, M$. If the set $\mathcal{U}$ is empty, we define $W_m(\emptyset) = 0$. For example, consider Case II where $M = 1$. We have $\mathcal{S}_l^0 = \mathcal{S}_l^1 = \mathcal{U}_l$, $W_0(\mathcal{U}_l) = G_l^0$, and $W_1(\mathcal{U}_l) = G_l^1$.

The power allocation algorithm for Case III is presented in Table 5.2. The algorithm iteratively picks users from the *eligible* subset $\mathcal{S}_l^m$ and assigns them to the *allocated* subset $\mathcal{U}_l^m$. In each step l, Ψ is the subset of FBS's that will transmit the level l packet; the complementary set $\overline{\Psi}$ is the subset of FBS's that will not transmit the level l packet. The expected increment in total power for each partition is computed, and the partition with the smallest expected increment will be chosen. Δ_l^m is the power of BS m for transmitting the level l data packet. In Steps 6–15, the MBS and FBS combination Ψ is determined for transmitting the level l packet, with the lowest power Δ_0. In Steps 16–30, elements in $\mathcal{S}_l^m$ are assigned to $\mathcal{U}_l^m$ according to Ψ. In Steps 31–35, power sums Q_l^m and the corresponding power allocations P_l^m are calculated in the reverse order from the known $\mathcal{U}_l^m$'s.

The complexity of the algorithm is $\mathcal{O}(ML)$.

Table 5.2: Power Allocation Algorithm For Case III

Initialize: $c_l^m = 0$ and $Q_{L+1}^m = 0$, for all l, m;
FOR $l = 1$ TO L
  FOR $m = 0$ TO M
    $\Delta_l^m = \Gamma_m(1+\Gamma_m)^{c_l^m} W_m(\mathcal{S}_l^m)$;
  END FOR
  Set $\Omega = \{1, \cdots, M\}$ and $\Psi = \emptyset$;
  WHILE ($\Omega \neq \emptyset$)
    $m' = \arg\min_{m\in\Omega} \Delta_l^m$;
    Compute $\Delta' = \Gamma_0(1+\Gamma_0)^{c_l^0} W_0(\bigcup_{m\in\overline{\Psi\cup m'}} \mathcal{S}_l^m)$;
    IF $((\sum_{m\in\Psi\cup m'} \Delta_l^m + \Delta') < \Delta_0)$
      Add m' to Ψ;
      $\Delta_0 = \sum_{m\in\Psi} \Delta_l^m + \Delta'$;
    END IF
    Remove m' from Ω;
  END WHILE
  IF ($\Psi = \emptyset$)
    $\mathcal{U}_l^0 = \mathcal{S}_l^0$;
    $c_l^0 = c_l^0 + 1$;
    Set $\mathcal{U}_l^m = \emptyset$, for all $m \neq 0$;
  ELSE
    $\mathcal{U}_l^0 = \bigcup_{m\in\overline{\Psi}} \mathcal{S}_l^m$;
    IF ($|\Psi| < M$)
      $c_l^0 = c_l^0 + 1$;
    END IF
    FOR $m \in \Psi$
      $c_l^m = c_l^m + 1$;
      $\mathcal{U}_l^m = \mathcal{S}_l^m$;
    END FOR
  END IF
END FOR
FOR $l = L$ TO 1
  FOR $m = 0$ TO M
    $Q_l^m = F_m(Q_{l+1}^m, \mathcal{U}_l^m)$ and $P_l^m = Q_l^m - Q_{l+1}^m$;
  END FOR
END FOR

5.3.3 Performance Evaluation

We evaluate the performance of the proposed power allocation algorithms using MATLAB™. Three scenarios corresponding to the three cases in Section 5.3.2 are simulated: (i) Case I: a single MBS; (ii) Case II: one MBS and one FBS; and (iii) Case III: one MBS and three FBS's.

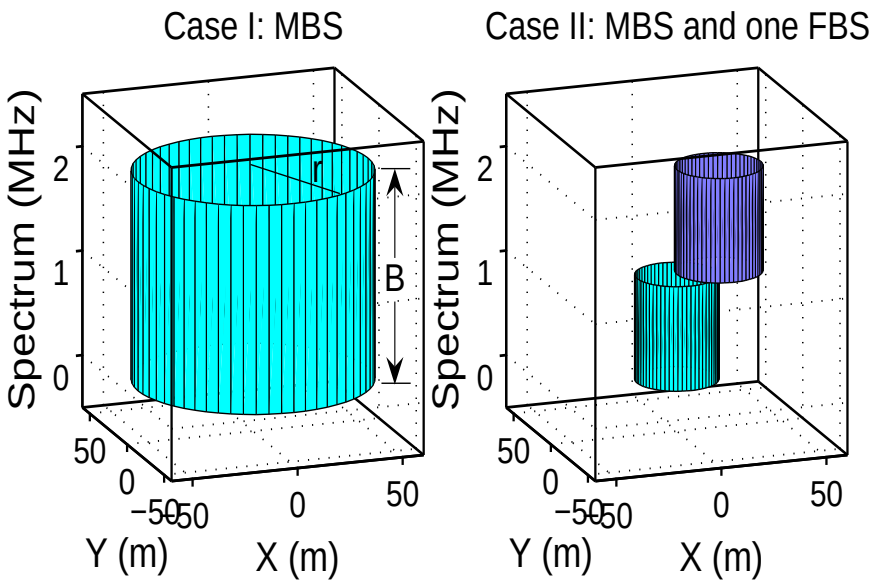


Figure 5.2: Case I vs. Case II: interference footprints.

Since we do not find any similar schemes in the literature, we made the following comparisons. First, we compare Cases I and II with respect to BS power consumption and interference footprint. In both cases, there are $K = 8$ users and $L = 4$ levels. In Case I, the MBS bandwidth is $B_0 = 2$ MHz. In Case II, the MBS and the FBS share the 2 MHz total bandwidth; the MBS bandwidth is $B_0 = 1$ MHz and the FBS bandwidth is $B_1 = 1$ MHz. The target data rate R_{tar} is set to 2 Mbps. The channel gain from a base station to each user is exponentially distributed in each time slot.

The interference footprints in the three dimensional space are plotted in Fig. 5.2. The height B of the cylinders indicates the spectrum used by a BS, while the radius r is proportional to the BS transmit power. In Case I when only the MBS is used, the total BS power is 45.71 dBm and the volume of the cylinder is $\pi r^2 B = 18,841$ MHz m^2. In Case II when both the MBS and FBS are used, the total BS power is 34.58 dBm and the total volume of the two cylinders is $2,378$ MHz m^2. Using an additional FBS achieves a 11.13 dB power saving and the interference footprint is reduced to 12.62% of that in Case I. This simple comparison clearly demonstrate the advantages of femtocells achieved by bringing BS's closer to users.

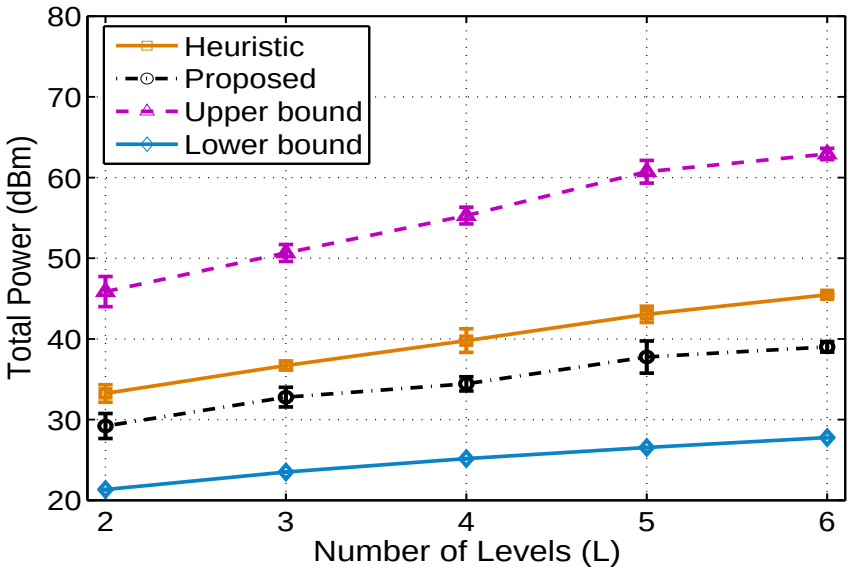


Figure 5.3: Case III: impact of number of levels L.

We next consider the more general Case III, using a femtocell network of one MBS and three FBS's. The MBS bandwidth is $B_0 = 1$ MHz and each FBS is assigned with bandwidth $B_m = 1$ MHz, $m = 1, 2, 3$. The target data rate is still 2 Mbps. In Figs. 5.3 and 5.4, we plot four curves, each obtained with: (i) the heuristic scheme described in Section 5.3.2; (ii) The proposed algorithm presented in Section 5.3.2; (iii) The upper bound; and (iv) the lower bound derived in Section 5.3.2. Each point in the figures is the average of 10 simulation runs. The 95% confidence intervals are plotted as error bars, which are all negligible.

In Fig. 5.3, we examine the impact of the number of packet levels L on the total BS transmit power. We increase L from 2 to 6, and plot the total power of base stations. As expected, the more packet levels, the larger the BS power consumption. Both the proposed and heuristic curves lie in between the upper and lower bound curves. When L is increased from 2 to 6, the power consumption of the heuristic scheme is increased by 12.22 dB, while the power consumption of the proposed algorithm is increased by 9.94 dB. The power savings achieved by the proposed algorithm over the heuristic scheme range from 3.92 dB to 6.45 dB.

In Fig. 5.4, we show the impact of the BS bandwidths. The number of levels is $L = 4$. We fix the total bandwidth at 2 MHz, which is shared by the MBS and FBS's. We increase the MBS

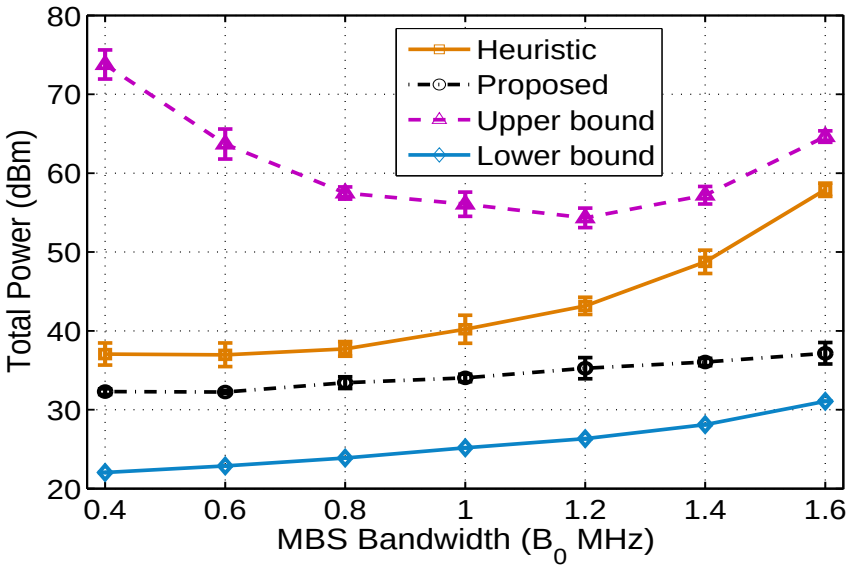


Figure 5.4: Case III: impact of MBS bandwidth B_0.

bandwidth from 0.4 MHz to 1.6 MHz in steps of 0.2 MHz, while decrease the bandwidth of FBS's from 1.6 MHz to 0.4 MHz. We find that the total power consumption is increased as B_0 gets large. This is due to the fact that as the FBS bandwidth gets smaller, the FBS's have to spend more power to meet the minimum data rate requirement. The curve produced by the proposed algorithm has a smaller slop than that of the heuristic scheme: the overall increase in the total power of the proposed algorithm is 4.86 dB, while that of the heuristic scheme is 20.84 dB. This implies that the proposed scheme is not very sensitive to the bandwidth allocation between the MBS and FBS's. The proposed algorithm also achieves consider power savings over the heuristic scheme. When $B_0 = 1.6$ MHz, the total power of the proposed algorithm is 20.75 dB lower than that of the heuristic scheme.

5.4 Video over CR Femtocell Networks

In this section, we investigate the problem of video streaming in femtocell cognitive radio (CR) networks and formulate a *stochastic programming* problem to examine three deployment scenarios. In the case of a single FBS, we apply *dual decomposition* to develop an optimum-achieving distributed algorithm, which is shown also optimal for the case of multiple non-interfering FBS's.

In the case of multiple interfering FBS's, we develop a greedy algorithm that can compute near-optimal solutions, and prove a closed-form lower bound for its performance based on an *interference graph* model. The proposed algorithms are evaluated with simulations, and are shown to outperform three alternative schemes with considerable gains.

5.4.1 System Model and Preliminaries

Spectrum and Network Model

We consider a spectrum consisting of $(M+1)$ channels, including one common, unlicensed channel (indexed as channel 0) and M licensed channels (indexed as channels 1 to M). The M licensed channels are allocated to a primary network, and the common channel is exclusively used by all CR users. We assume all the channels follow a synchronized time slot structure [14]. The capacity of each licensed channel is B_1 Mbps, while the capacity of the common channel is B_0 Mbps. The channel states evolve independently, while the occupancy of each licensed channel follows a two-state discrete-time Markov process as described in Section 2.3.1.

The femtocell CR network is illustrated in Fig. 5.5. There is an MBS and N FBS's deployed in the area to serve CR users. The N FBS's are connected to the MBS (and the Internet) via broadband wireline connections. Due to advances in antenna technology, it is possible to equip multiple antennas at the base stations. The MBS has one antenna that is always tuned to the common channel. Each FBS is equipped with multiple antennas (e.g., M) and is able to sense multiple licensed channels at the beginning of each time slot. There are K_i CR users in femtocell i, $i = 1, 2, \cdots, N$, and $\sum_{i=1}^{N} K_i = K$. Each CR user has a software radio transceiver, which can be tuned to any of the M+1 channels. A CR user will either connect to a nearby FBS using one or more of the licensed channels or to the MBS via the common channel.

Although the CR users are mobile, we assume constant topology during a time slot. If the topology is changed during a time slot, the video transmission will only be interrupted for the time slot, since the proposed algorithms are executed in every time slot for new channel assignment and schedule.

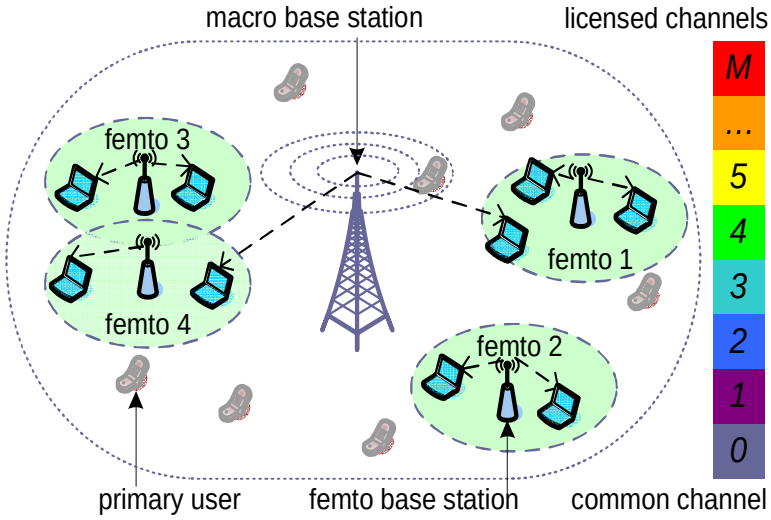


Figure 5.5: A femtocell CR network with one MBS and four FBS's.

Spectrum Sensing and Access

The femtocell CR network is within the coverage of the infrastructure-based primary network. Both FBS's and CR users sense the channels to identify spectrum opportunities in each time slot. Each time slot consists of (i) a *sensing phase*, when CR users and FBS's sense licensed channels, (ii) a *transmission phase*, when CR users and FBS's attempt to access licensed channels, and (iii) an *acknowledgment phase*, when acknowledgments (ACK) are returned to the source.

As described in Section 2.5.1, cooperative sensing policy is also adopted here. We also adopt a *hypothesis test* to detect channel availability. We assume that each CR user chooses one channel to sense in each time slot, since it only has one transceiver. The sensing results will be shared among CR users and FBS's via the common channel in the sensing phase. Given L sensing results on channel m, the availability of channel m, i.e., $P_m^A(\Theta_1^m, \cdots, \Theta_L^m)$, can be computed iteratively

as follows.

$$P_m^A(\Theta_1^m) = \left[1 + \frac{\eta_m}{1-\eta_m} \times \frac{(\delta_1^m)^{1-\Theta_1^m}(1-\delta_1^m)^{\Theta_1^m}}{(\epsilon_1^m)^{\Theta_1^m}(1-\epsilon_1^m)^{1-\Theta_1^m}}\right]^{-1} \tag{5.28}$$

$$\begin{aligned} P_m^A(\vec{\Theta}_l^m) &= P_m^A(\Theta_1^m, \Theta_2^m, \cdots, \Theta_l^m) \\ &= \left\{1 + \left[\frac{1}{P_m^A(\Theta_1^m, \Theta_2^m, \cdots, \Theta_{l-1}^m)} - 1\right] \times \right. \\ &\quad \left. \frac{(\delta_l^m)^{1-\Theta_l^m}(1-\delta_l^m)^{\Theta_l^m}}{(\epsilon_l^m)^{\Theta_l^m}(1-\epsilon_l^m)^{1-\Theta_l^m}}\right\}^{-1}, l = 2, \cdots, L. \end{aligned} \tag{5.29}$$

As in Section 2.5.1, we adopt a probabilistic approach: based on sensing results $\vec{\Theta}_m$, we have $D_m(t) = 0$ with probability $P_m^D(\vec{\Theta}_m)$ and $D_m(t) = 1$ with probability $1 - P_m^D(\vec{\Theta}_m)$. For primary user protection, the collision probability with primary users caused by CR users should be bounded. The probability $P_m^D(\vec{\Theta}_m)$ is determined as follows

$$P_m^D(\vec{\Theta}_m) = \min\left\{\gamma_m / \left[1 - P_m^A(\vec{\Theta}_m)\right], 1\right\}. \tag{5.30}$$

Let $\mathcal{A}(t) := \{m | D_m(t) = 0\}$ be the set of available channels in time slot t. Then $G^t = \sum_{m \in \mathcal{A}(t)} P_m^A(\Theta_1^m)$ is the expected number of available channels. These channels will be accessed in the transmission phase of time slot t.

Channel Model

Without loss of generality, we consider independent block fading channels that is widely used in prior work [110]. The channel fading-gain process is piecewise constant on blocks of one time slot, and fading in different time slots are independent. Let $f_X^{i,j}(x)$ denote the *probability density function* of the received SINR X from a base station i at CR user j. We assume the packet can be successfully decoded if the received SINR exceeds a threshold H. The packet loss probability from base station i to CR user j is

$$P_{i,j} = \Pr\{X \leq H\} = \int_0^H f_X^{i,j}(x)dx = F_X^{i,j}(H), \tag{5.31}$$

where $F_X^{i,j}(H)$ is the cumulative density function of X.

In the case of correlated fading channels, which can be modeled as finite state Markov Process [111], the packet loss probability in the next time slot can be estimated from the known state of the previous time slot and the transition probabilities. If the packet is successfully decoded, the CR user returns an ACK to the base station in the ACK phase. We assume ACKs are always successfully delivered.

Video Performance Measure

We assume each active CR user receives a real-time video stream from either the MSB or an FSB. Without loss of generality, we adopt the MGS option of H.264/SVC, for scalability to accommodate the high variability of network bandwidth in CR networks.

Due to real-time constraint, each Group of Pictures (GOP) of a video stream must be delivered in the next T time slots. With MGS, enhancement layer NAL units can be discarded from a quality scalable bit stream, and thus packet-based quality scalable coding is provided. Our approach is to encode the video according to the maximum rate the channels can support. During transmission, only part of the MGS video gets transmitted as allowed by the current available channel bandwidth. The video packets are transmitted in decreasing order of their significance in decoding. When a truncated MGS video is received and decoded, the PSNR is computed by substituting the effective rate of the received MGS video into (5.32) given below, thus the original video is not required.

Without loss of generality, we assume that the last wireless hop is the bottleneck; video data is available at the MBS and FBS's when they are scheduled to be transmitted. The quality of reconstructed MGS video can be modeled as [53]:

$$W(R) = \alpha + \beta \times R, \tag{5.32}$$

where $W(R)$ is the average peak signal-to-noise ratio (PSNR) of the reconstructed video, R is the received data rate, α and β are constants depending on the video sequence and codec.

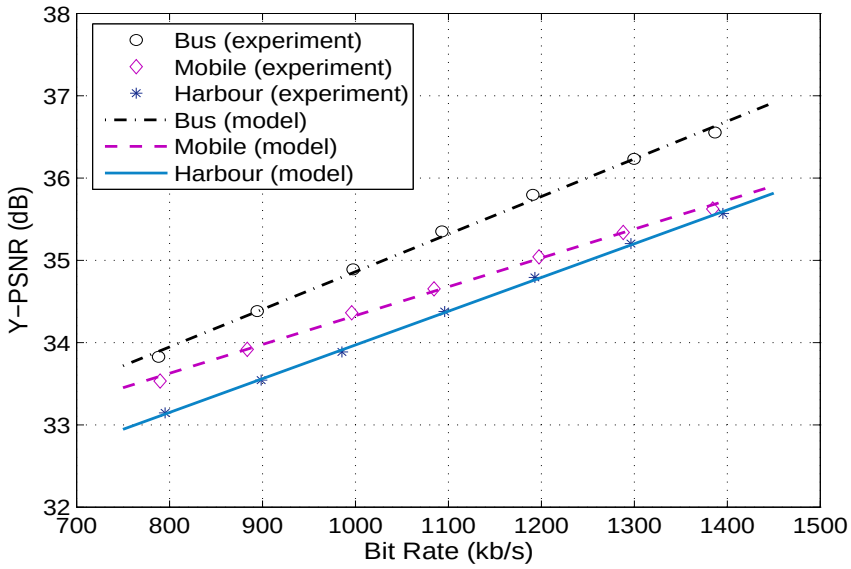


Figure 5.6: Rate-distortion curves of three H.264/SVC MGS videos.

We verified (5.32) using an H.264/SVC codec and the *Bus*, *Mobile*, and *Harbour* test sequences. In Fig. 5.6, the markers are obtained by truncating the encoded video's enhancement layer at different positions to obtain different effective rates, while the curves are computed using (5.32). The curves fit well with measurements for the three sequences. It is worth noting that PSNR may not be a good measure of video quality as compared with alternative metrics such as MS-SSIM [112]. The main reason for choosing PSNR is that there is a closed-form model relating it to network level metrics–video rate. With the closed-form model, we can have a mathematical formulation of the scheduling/resource allocation problem, and derive effective algorithms. Should such closed-form models be available for MS-SSIM, it is possible to incorporate it into the optimization framework as well.

5.4.2 MGS Video over Femtocell CR Networks

In this section, we address the problem of resource allocation for MGS videos over femtocell CR networks. We first examine the case of a single FBS, and then the more general case of multiple non-interfering or interfering FBS's. The algorithms for the single and non-interfering FBS cases are distributed ones and optimal. The algorithm for the interfering FBS case is a centralized one

that can be executed at the MBS. To simplify notation, we omit the time slot index t for most of the variables in this Section. For example, x represents a variable for time slot t, x^- represents the variable in time slot $(t-1)$, and x^+ represents the variable in time slot $(t+1)$.

Case of Single FBS

Formulation We first consider the case of a single FBS in the CR network, where the FBS can use all the G available channels to stream videos to K active CR users. Let w_j be the PSNR of CR user j at the beginning of time slot t and W_j the PSNR of CR user j at the end of time slot t. In time slot t, w_j is already known; W_j is a random variable that depends on channel condition and primary user activity; and w_j^+ is a *realization* of W_j. Let $\xi_{0,j}$ and $\xi_{1,j}$ indicate the random packet losses from the MBS and FBS, respectively, to CR user j in time slot t. That is, $\xi_{i,j}$ is 1 with probability $\bar{P}_{i,j} = 1 - P_{i,j}$ and 0 with probability $P_{i,j}$. Due to block fading channels, $P_{i,j}$'s do not change within the time slot.

Let $\rho_{0,j}$ and $\rho_{1,j}$ be the portions of time slot t when CR user j receives video data from the MBS and FBS, respectively. The average PSNR is computed every T time slots. We first have $W_j(0) = \alpha_j$, when $t = 0$. In each time slot t, the CR user receives $\xi_{0,j}\rho_{0,j}B_0$ bits through the MBS, and $\xi_{1,j}\rho_{1,j}GB_1$ bits through the FBS (assuming that OFDM is used), which contribute an increase of $\beta(\xi_{0,j}\rho_{0,j}B_0 + \xi_{1,j}\rho_{1,j}GB_1)/T$ to the total PSNR in this T time slot interval, according to (5.32). Therefore we have the following recursive relationship: $W_j = W_j^- + \beta(\xi_{0,j}\rho_{0,j}B_0 + \xi_{1,j}\rho_{1,j}GB_1)/T$ = $W_j^- + \xi_{0,j}\rho_{0,j}R_{0,j} + \xi_{1,j}\rho_{1,j}GR_{1,j}$, where $R_{0,j} = \beta B_0/T$ and $R_{1,j} = \beta B_1/T$.

For proportional fairness, we aim to maximize the sum of the logarithms of the PSNRs of all CR users [71]. We formulate a *multistage stochastic programming problem* by maximizing the

expectation of the logarithm-sum at time T.

$$\begin{aligned}
\text{maximize: } & \sum_{j=1}^{K} \mathbb{E}[\log(W_j(T))] \\
\text{subject to: } & W_j = W_j^- + \xi_{0,j}\rho_{0,j}R_{0,j} + \xi_{1,j}\rho_{1,j}GR_{1,j}, \;\; j = 1, \cdots, K, \; t = 1, \cdots, T \\
& \sum_{j=1}^{K} \rho_{i,j} \leq 1, \;\; i = 0, 1, \; t = 1, \cdots, T \\
& \rho_{i,j} \geq 0, \; i = 0, 1, \; j = 1, \cdots, K, \; t = 1, \cdots, T.
\end{aligned} \tag{5.33}$$

$R_{0,j} = \beta_j B_0/T$ and $R_{1,j} = \beta_j B_1/T$ are constants for the j-th MGS video.

At the beginning of the last time slot T, a realization $\boldsymbol{\xi}_{[T-1]} = [\vec{\xi}_1, \vec{\xi}_2, \cdots, \vec{\xi}_{T-1}]$ is known, where $\vec{\xi}_t = [\xi_{0,1}^t, \xi_{0,2}^t, \cdots, \xi_{0,K}^t, \xi_{1,1}^t, \cdots, \xi_{1,K}^t]$, $t = 1, 2, \cdots, T-1$. It can be shown that the multistage stochastic programming problem (5.33) can be decomposed into T serial sub-problems, each to be solved in a time slot t [9].

$$\begin{aligned}
\text{maximize: } & \sum_{j=1}^{K} \mathbb{E}\{\log(W_j)|\boldsymbol{\xi}_{[t-1]}\} \\
\text{subject to: } & W_j = W_j^- + \xi_{0,j}\rho_{0,j}R_{0,j} + \xi_{1,j}\rho_{1,j}GR_{1,j}, j = 1, \cdots, K \\
& \sum_{j=1}^{K} \rho_{i,j} \leq 1, \;\; i = 0, 1 \\
& \rho_{i,j} \geq 0, \; i = 0, 1, \; j = 1, \cdots, K,
\end{aligned} \tag{5.34}$$

where $\mathbb{E}\{\log(W_j)|\boldsymbol{\xi}_{[t-1]}\}$ denotes the *conditional expectation* of $\log(W_j)$ given realization $\boldsymbol{\xi}_{[t-1]}$. W_j^- is known given the realization. When $t = 1$, the conditional expectation becomes an unconditional expectation.

Since a CR user has only one transceiver, it can operate on either one or more licensed channels (i.e., connecting to the FBS) or the common channel (i.e., connecting to the MBS), but not both simultaneously. Assume CR user j operates on the common channel with probability p_j and

one or more licensed channels with probability q_j. We then rewrite problem (5.34) as

$$\begin{aligned}
\text{maximize: } & \sum_{j=1}^{K} \left[p_j \bar{P}_{0,j} \log(W_j^- + \rho_{0,j} R_{0,j}) \ + q_j \bar{P}_{1,j} \log(W_j^- + \rho_{1,j} G R_{1,j}) \right] \\
\text{subject to: } & \sum_{j=1}^{K} \rho_{i,j} \leq 1, \ \ i = 0, 1 \\
& p_j + q_j = 1, \ \ j = 1, \cdots, K \\
& \rho_{i,j},\ p_j,\ q_j \geq 0, \ \ i = 0, 1,\ j = 1, \cdots, K.
\end{aligned}$$

Properties In this section, we analyze the formulated problem (5.35) and derive its properties. We have Lemmas 1, 2, and 3 and Theorem 1 and provide the proofs in the following.

Lemma 5.1. *Problem (5.35) is a convex optimization problem.*

Proof. First, it can be shown that the single term $p_j \bar{P}_{0,j} \log(W_j^- + \rho_{0,j} R_{0,j}) + q_j \bar{P}_{1,j} \log(W_j^- + \rho_{1,j} G R_{1,j})$ is a concave function, because its *Hessian matrix* is negative semi-definite. Then, the objective function is concave since the sum of concave functions is also concave. Finally, all the constraints are linear. We conclude that problem (5.35) is convex with a unique optimal solution. □

Lemma 5.2. *If $[\rho, p, q]$ is a feasible solution to problem (5.35), then $[\rho, q, p]$ is also feasible.*

Proof. Since $[\rho, p, q]$ is feasible, we have $p + q = 1$. Switching the two probabilities, we still have $q + p = 1$. Therefore, the derived new solution is also feasible. □

Lemma 5.3. *Let the optimal solution be $[\rho^*, p^*, q^*]$. If $p_j^* \geq q_j^*$, then $\bar{P}_{0,j} \log(W_j^- + \rho_{0,j}^* R_{0,j})$ is greater than or equal to $\bar{P}_{1,j} \log(W_j^- + \rho_{1,j}^* G R_{1,j})$. And vice versa.*

Proof. Assume $\bar{P}_{0,j} \log(W_j^- + \rho_{0,j}^* R_{0,j})$ is less than $\bar{P}_{1,j} \log(W_j^- + \rho_{1,j}^* G R_{1,j})$. Since $p_j^* \geq q_j^*$, the sum of the product $p_j^* \bar{P}_{0,j} \log(W_j^- + \rho_{0,j}^* R_{0,j}) + q_j^* \bar{P}_{1,j} \log(W_j^- + \rho_{1,j}^* G R_{1,j})$ is smaller than the sum of the product $q_j^* \bar{P}_{0,j} \log(W_j^- + \rho_{0,j}^* R_{0,j}) + p_j^* \bar{P}_{1,j} \log(W_j^- + \rho_{1,j}^* G R_{1,j})$. Thus we can obtain an objective value larger than the optimum by switching the values of p_j^* and q_j^*, which is

still feasible according to Lemma 2. This conflicts with the assumption that $[\rho^*, p^*, q^*]$ is optimal. The reverse statement can be proved similarly. □

Theorem 5.1. *Let the optimal solution be* $[\rho^*, p^*, q^*]$. *If* $p_j^* > q_j^*$, *then we have* $p_j^* = 1$ *and* $q_j^* = 0$. *Otherwise, we have* $p_j^* = 0$ *and* $q_j^* = 1$.

Proof. If $p_j^* > q_j^*$, we have $\bar{P}_{0,j} \log(W_j^- + \rho_{0,j}^* R_{0,j}) \geq \bar{P}_{1,j} \log(W_j^- + \rho_{1,j}^* G R_{1,j})$ according to Lemma 3. Since the objective function is linear with respect to p_j and q_j, the optimal value can be achieved by setting p_j to its maximum value 1 and q_j to its minimum value 0. The reverse statement can be proved similarly. □

According to Theorem 1, a CR user is connected to either the MBS or the FBS for the *entire* duration of a time slot in the optimal solution. That is, it does not switch between base stations during a time slot under optimal scheduling.

Distributed Solution Algorithm To solve problem (5.35), we define non-negative *dual variables* $\lambda = [\lambda_0, \lambda_1]$ for the two inequality constraints. The *Lagrangian function* is

$$\begin{aligned} \mathcal{L}(p, \rho, \lambda) &= \sum_{j=1}^{K} \left[p_j \bar{P}_{0,j} \log(W_j^- + \rho_{0,j} R_{0,j}) + (1 - p_j) \bar{P}_{1,j} \log(W_j^- + \rho_{1,j} G R_{1,j}) \right] + \\ &\quad \lambda_0 (1 - \sum_{j=1}^{K} \rho_{0,j}) + \lambda_1 (1 - \sum_{j=1}^{K} \rho_{1,j}) \\ &= \sum_{j=1}^{K} \mathcal{L}_j(p_j, \rho_{0,j}, \rho_{1,j}, \lambda_0, \lambda_1) + \lambda_0 + \lambda_1, \end{aligned} \tag{5.35}$$

where

$$\begin{aligned} \mathcal{L}_j(p_j, \rho_{0,j}, \rho_{1,j}, \lambda_0, \lambda_1) = p_j \bar{P}_{0,j} \log(W_j^- + \rho_{0,j} R_{0,j}) + \\ (1 - p_j) \bar{P}_{1,j} \log(W_j^- + \rho_{1,j} G R_{1,j}) - \lambda_0 \rho_{0,j} - \lambda_1 \rho_{1,j}. \end{aligned}$$

The corresponding problem can be decomposed into K sub-problems and solved iteratively. In Step $\tau \geq 1$, for given $\lambda_0(\tau)$ and $\lambda_1(\tau)$ values, each CR user j solves the following sub-problem

using local information.

$$[p_j^*(\tau), \rho_{0,j}^*(\tau), \rho_{1,j}^*(\tau)] = \underset{p_j, \rho_{0,j}, \rho_{1,j} \geq 0}{\arg\max} \mathcal{L}_j(p_j, \rho_{0,j}, \rho_{1,j}, \lambda_0(\tau), \lambda_1(\tau)). \tag{5.36}$$

There is a unique optimal solution since the objective function in (5.36) is concave. The CR users then exchange their solutions. The *master dual problem*, for given $p(\tau)$ and $\rho(\tau)$, is:

$$\min_{\lambda \geq 0} \mathcal{L}(p(\tau), \rho(\tau), \lambda) = \sum_{j=1}^{K} \mathcal{L}_j(p_j(\tau), \rho_{0,j}(\tau), \rho_{1,j}(\tau), \lambda_0, \lambda_1) + \lambda_0 + \lambda_1. \tag{5.37}$$

Since the Lagrangian function is differentiable, the *gradient iteration* approach can be used.

$$\lambda_i(\tau + 1) = \left[\lambda_i(\tau) - s \times \left(1 - \sum_{j=1}^{K} \rho_{i,j}^*(\tau) \right) \right]^+, \ i = 0, 1, \tag{5.38}$$

where s is a sufficiently small positive *step size* and $[\cdot]^+$ denotes the projection onto the nonnegative axis. The updated $\lambda_i(\tau + 1)$ will again be used to solve the sub-problems, and so forth. Since the problem is convex, we have *strong duality*; the *duality gap* between the primal and dual problems is zero. The dual variables $\lambda(\tau)$ will converge to the optimal values as τ goes to infinity. Since the optimal solution to (5.36) is unique, the primal variables $p(\tau)$ and $\rho_{i,j}(\tau)$ will also converge to their optimal values when τ is sufficiently large.

The distributed solution procedure is presented in Table 5.3. In the table, Steps 3–8 solve the sub-problem in (5.36); Step 9 updates the dual variables. The threshold ϕ is a prescribed small value with $0 \leq \phi \ll 1$. The algorithm terminates when the dual variables are sufficiently close to the optimal values.

Case of Multiple Non-interfering FBS's

We next consider the case of $N > 1$ non-interfering FBS's. The coverages of the FBS's do not overlap with each other, as FBS 1 and 2 in Fig. 5.5. Consequently, each FBS can use all the available licensed channels without interfering other FBS's. Assume each CR user knows the

Table 5.3: Algorithm for the Case of Single FBS

1:	Set $\tau = 0$, $\lambda_0(0)$ and $\lambda_1(0)$ to some nonnegative value;
2:	DO % (each CR user j executes Steps 3–8)
3:	$\rho_{0,j}(\tau) = \left[\frac{\bar{P}_{0,j}}{\lambda_0(\tau)} - \frac{W_j^-}{R_{0,j}}\right]^+$, $\rho_{1,j}(\tau) = \left[\frac{\bar{P}_{1,j}}{\lambda_1(\tau)} - \frac{W_j^-}{R_{1,j}G}\right]^+$;
4:	IF $\left[\bar{P}_{0,j}\log(W_j^- + \rho_{0,j}(\tau)R_{0,j}) - \lambda_0(\tau)\rho_{0,j}(\tau)\right] >$ $\left[\bar{P}_{1,j}\log(W_j^- + \rho_{1,j}(\tau)GR_{1,j}) - \lambda_1(\tau)\rho_{1,j}(\tau)\right]$
5:	Set $p_j(\tau) = 1$ and $\rho_{1,j}(\tau) = 0$;
6:	ELSE
7:	Set $p_j(\tau) = 0$ and $\rho_{0,j}(\tau) = 0$;
8:	END IF
9:	MBS updates $\lambda_i(\tau+1)$ as in (5.38);
10:	$\tau = \tau + 1$;
11:	WHILE $\left(\sum_{i=0}^{1}(\lambda_i(\tau+1) - \lambda_i(\tau))^2 > \phi\right)$

nearest FBS and is associate with it. Let $\mathcal{U}_i$ denote the set of CR users associated with FBS i. The resource allocation problem becomes:

$$\text{maximize: } \sum_{j=1}^{K} p_j \bar{P}_{0,j}\log(W_j^- + \rho_{0,j}R_{0,j}) + \sum_{i=1}^{N}\sum_{j\in\mathcal{U}_i} q_j \bar{P}_{i,j}\log(W_j^- + \rho_{i,j}GR_{i,j}) \quad (5.39)$$

$$\text{subject to: } \sum_{j=1}^{K}\rho_{0,j} \le 1$$

$$\sum_{j\in\mathcal{U}_i}\rho_{i,j} \le 1, \;\; i = 1,\cdots,N$$

$$p_j + q_j = 1, \;\; j = 1,\cdots,K$$

$$\rho_{i,j},\; p_j,\; q_j \ge 0, \;\; i = 1,\cdots,N,\; j = 1,\cdots,K.$$

Since all the available channels can be allocated to each FBS with spatial reuse, problem (5.39) can be solved using the algorithm in Table 5.3 with some modified notation: $\rho_{1,j}(\tau)$ now becomes $\rho_{i,j}(\tau)$ and $\lambda_1(\tau)$ becomes $\lambda_i(\tau)$, $i = 1,\cdots,N$. The dual variables are iteratively updated as

$$\lambda_0(\tau+1) = \left[\lambda_0(\tau) - s\times\left(1 - \sum_{j=1}^{K}\rho_{0,j}^*(\tau)\right)\right]^+ \quad (5.40)$$

$$\lambda_i(\tau+1) = \left[\lambda_i(\tau) - s\times\left(1 - \sum_{j\in\mathcal{U}_i}\rho_{i,j}^*(\tau)\right)\right]^+, \;\; i = 1,\cdots,N. \quad (5.41)$$

Table 5.4: Algorithm for the Case of Multiple Non-Interfering FBS's

1: Set $\tau = 0$, and $\lambda_0(0)$ and $\lambda_i(0)$ to some nonnegative values, for all i;
2: DO % (each CR user j executes Steps 3–8)
3: $\rho_{0,j}(\tau) = \left[\frac{\bar{P}_{0,j}}{\lambda_0(\tau)} - \frac{W_j^-}{R_{0,j}}\right]^+$, $\rho_{i,j}(\tau) = \left[\frac{\bar{P}_{i,j}}{\lambda_i(\tau)} - \frac{W_j^-}{R_{i,j}G}\right]^+$;
4: IF $\left[\bar{P}_{0,j}\log(W_j^- + \rho_{0,j}(\tau)R_{0,j}) - \lambda_0(\tau)\rho_{0,j}(\tau)\right] > \left[\bar{P}_{i,j}\log(W_j^- + \rho_{i,j}(\tau)GR_{i,j}) - \lambda_i(\tau)\rho_{i,j}(\tau)\right]$
5: Set $p_j(\tau) = 1$ and $\rho_{i,j}(\tau) = 0$;
6: ELSE
7: Set $p_j(\tau) = 0$ and $\rho_{0,j}(\tau) = 0$;
8: END IF
9: MBS updates $\lambda_i(\tau + 1)$ as in (5.40) and (5.41);
10: $\tau = \tau + 1$;
11: WHILE $\left(\sum_{i=0}^{N}(\lambda_i(\tau + 1) - \lambda_i(\tau))^2 > \phi\right)$

The modified solution algorithm is presented in Table 5.4. As in the case of single FBS, the algorithm is jointly executed by the CR users and MBS, by iteratively updating the dual variables $\lambda_0(\tau)$ and $\lambda_i(\tau)$'s, and the resource allocations $\rho_{0,j}^*(\tau)$ and $\rho_{i,j}^*(\tau)$'s. It can be shown that the distributed algorithm can produce the optimal solution for problem (5.39).

Case of Multiple Interfering FBS's

Formulation Finally, we consider the case of multiple interfering FBS's. Assume that the coverages of some FBS's overlap with each other, as FBS 3 and 4 in Fig. 5.5. They cannot use the same channel simultaneously, but have to compete for the available channels in the transmission phase. Define *channel allocation variables* $c_{i,m}$ for time slot t as:

$$c_{i,m} = \begin{cases} 1, & \text{if channel } m \text{ is allocated to FBS } i \\ 0, & \text{otherwise.} \end{cases} \tag{5.42}$$

Given an allocation, the expected number of available channels for FBS i is $G_i = \sum_{m \in \mathcal{A}(t)} c_{i,m} P_m^A$.

We use *interference graph* to model the case of overlapping coverages, which is defined below.

Figure 5.7: Interference graph for the femtocell CR network shown in Fig. 5.5.

Definition 5.1. *An* interference graph $G_I = (V_I, E_I)$ *is an undirected graph where each vertex represents an FBS and each edge indicates interference between the two end FBS's.*

For the example given in Fig. 5.5, we can derive an interference graph as shown in Fig. 5.7. FBS 3 and 4 cannot use the same channel simultaneously, as summarized in the following lemma.

Lemma 5.4. *If channel m is allocated to FBS i, the neighboring vertices of FBS i in the interference graph G_I, denoted as $\mathcal{R}(i)$, cannot use the same channel m simultaneously.*

Further define index variables d_i^k as

$$d_i^k = \begin{cases} 1, & \text{if FBS } i \text{ is an endpoint of link } k \in G_I \\ 0, & \text{otherwise.} \end{cases} \tag{5.43}$$

The interference constraint can be described as $\sum_{i=1}^{N} d_i^k c_{i,m} \leq 1$, for $m = 0, \cdots, M$, and for all link $k \in G_I$. We then have the following problem formulation.

$$\begin{aligned} \text{maximize: } & \sum_{j=1}^{K} p_j \bar{P}_{0,j} \log(W_j^- + \rho_{0,j} R_{0,j}) + \sum_{i=1}^{N} \sum_{j \in \mathcal{U}_i} q_j \bar{P}_{i,j} \log(W_j^- + \rho_{i,j} G_i R_{i,j}) \qquad (5.44) \\ \text{subject to: } & \sum_{j=1}^{K} \rho_{0,j} \leq 1 \\ & \sum_{j \in \mathcal{U}_i} \rho_{i,j} \leq 1, \quad i = 1, \cdots, N \\ & p_j + q_j = 1, \quad j = 1, \cdots, K \\ & G_i = \sum_{m \in \mathcal{A}(t)} c_{i,m} P_m^A, \quad i = 1, \cdots, N \\ & \sum_{i=1}^{N} d_i^k c_{i,m} \leq 1, m = 0, \cdots, M, \text{for link } k \in G_I, \\ & \rho_{i,j}, p_j, q_j, c_{i,m} \geq 0, \quad i = 1, \cdots, N, \ j = 1, \cdots, K, \ m = 0, \cdots, M. \end{aligned}$$

Table 5.5: Channel Allocation Algorithm for Case of Interfering FBS's

1: Initialize $\boldsymbol{c}$ to a zero matrix, FBS set $\mathcal{N} = \{1, \cdots, N\}$, and FBS-channel set $\mathcal{C} = \mathcal{N} \times \mathcal{A}(t)$;
2: WHILE ($\mathcal{C}$ is not empty)
3: Find FBS-channel pair $\{i', m'\}$, such that $\{i', m'\} = \arg\max_{\{i,m\} \in \mathcal{C}} \{Q(\boldsymbol{c} + \boldsymbol{e}_{i,m}) - Q(\boldsymbol{c})\}$;
4: Set $\boldsymbol{c} = \boldsymbol{c} + \boldsymbol{e}_{i',m'}$;
5: Remove $\{i', m'\}$ from $\mathcal{C}$;
6: Remove $\mathcal{R}(i') \times m'$ from $\mathcal{C}$;
7: END WHILE

Solution Algorithm The optimal solution to problem (5.44) depends on the channel allocation variables $c_{i,m}$. Problem (5.44) can be solved with the algorithm in Table 5.4 if the $c_{i,m}$'s are known. Let $Q(\boldsymbol{c})$ be the suboptimal objective value for a given channel allocation $\boldsymbol{c}$, where $\boldsymbol{c} = [\vec{c}_1, \vec{c}_2, \cdots, \vec{c}_N]$ and $\vec{c}_i$ is a vector of elements $c_{i,m}$, for FBS i and channels $m \in \mathcal{A}(t)$. If all the FBS's are disjointedly distributed with no overlap, each FBS can use all the available channels. We have $c_{i,m} = 1$ for all i and $m \in \mathcal{A}(t)$, i.e., it is reduced to the case in Section 5.4.2.

To solve problem (5.44), we first apply a *greedy algorithm* to allocate the available channels in $\mathcal{A}(t)$ to the FBS's (i.e., to determine $\boldsymbol{c}$). We then apply the algorithm in Table 5.4 with the computed $\boldsymbol{c}$ to obtain a near-optimal solution. Let $\boldsymbol{e}_{i,m}$ be a matrix with 1 at position $\{i, m\}$ and 0 at all other positions, representing the allocation of channel $m \in \mathcal{A}(t)$ to FBS i. The greedy channel allocation algorithm is given in Table 5.5, where the FBS-channel pair that can achieve the largest increase in $Q(\cdot)$ is chosen in each iteration. The worst case complexity of the greedy algorithm is $O(N^2M^2)$.

Performance Lower Bound We next present a lower bound for the greedy algorithm. Let $e(l)$ be the l-th FBS-channel pair chosen in the greedy algorithm, and π_l denote the sequence $\{e(1), e(2), \cdots, e(l)\}$. The increase in object value (5.44) due to the l-th allocated FBS-channel pair is denoted as

$$\Delta_l := \Delta(\pi_l, \pi_{l-1}) = Q(\pi_l) - Q(\pi_{l-1}). \tag{5.45}$$

Since $Q(\pi_0) = Q(\emptyset) = 0$, we have

$$\begin{aligned}\sum_{l=1}^{L} \Delta_l &= Q(\pi_L) - Q(\pi_{L-1}) + \cdots + Q(\pi_1) - Q(\pi_0) \\ &= Q(\pi_L) - Q(\pi_0) = Q(\pi_L).\end{aligned}$$

For two FBS-channel pairs $e(l)$ and $e(l')$, we say $e(l)$ *conflicts with* $e(l')$ when there is an edge connecting the FBS in $e(l)$ and the FBS in $e(l')$ in the interference graph G_I, and the two FBS's choose the same channel. Let Ω be the global optimal solution. We define ω_l as the subset of Ω that conflicts with allocation $e(l)$ but not with the previous allocations $\{e(1), e(2), \cdots, e(l-1)\}$.

Lemma 5.5. *Assume the greedy algorithm in Table 5.5 stops in L steps. The global optimal solution Ω can be partitioned into L non-overlapping subsets ω_l, $l = 1, 2, \cdots, L$.*

Proof. According to the definition of ω_l, the L subsets of the optimal solution Ω do not intersect with each other. Assume the statement is false, then the union of these L subsets is not equal to the optimal set Ω. Let the *set difference* be $\omega_{L+1} = \Omega \setminus (\cup_{l=1}^{L} \omega_l)$. By definition, ω_{L+1} does not conflict with the existing L allocations $\{e(1), \cdots, e(L)\}$, meaning that the greedy algorithm can continue to at least the $(L+1)$-th step. This conflicts with the assumption that the greedy algorithm stops in L steps. It follows that $\Omega = \cup_{l=1}^{L} \omega_l$. □

Let $\Delta(\pi_2, \pi_1) = Q(\pi_2) - Q(\pi_1)$ denote the difference between two feasible allocations π_1 and π_2. We next derive a lower bound on the performance of the greedy algorithm. We assume two properties for function $\Delta(\pi_2, \pi_1)$ in the following.

Property 5.1. *Consider FBS-channel pair sets π_1, π_2, and σ, satisfying $\pi_1 \subseteq \pi_2$ and $\sigma \cap \pi_2 = \emptyset$. We have $\Delta(\pi_2 \cup \sigma, \pi_1 \cup \sigma) \leq \Delta(\pi_2, \pi_1)$.*

Property 5.2. *Consider FBS-channel pair sets π, σ_1, and σ_2 satisfying $\sigma_1 \cap \pi = \emptyset$, $\sigma_2 \cap \pi = \emptyset$, and $\sigma_1 \cap \sigma_2 = \emptyset$. We have $\Delta(\sigma_1 \cup \sigma_2 \cup \pi, \pi) \leq \Delta(\sigma_1 \cup \pi, \pi) + \Delta(\sigma_2 \cup \pi, \pi)$.*

In Property 1, we have $\sigma \cap \pi_1 = \emptyset$ since $\pi_1 \subseteq \pi_2$ and $\sigma \cap \pi_2 = \emptyset$. This property states that the incremental objective value does not get larger as more channels are allocated and as

the objective value gets larger. Property 2 states that the incremental objective value achieved by allocating multiple FBS-channel pair sets does not exceed the sum of the incremental objective values achieved by allocating each individual FBS-channel pair set. These are generally true for many resource allocation problems [71].

Since we choose the maximum incremental allocation in each step of the greedy algorithm, we have Lemma 5.6 that directly follows Step 3 in Table 5.5.

Lemma 5.6. *For any FBS-channel pair $\sigma \in \omega_l$, we have $Q(\pi_{l-1} \cup \sigma) - Q(\pi_{l-1}) = \Delta(\pi_{l-1} \cup \sigma, \pi_{l-1}) \leq \Delta_l$.*

Lemma 5.7. *Assume the greedy algorithm stops in L steps, we have*

$$Q(\Omega) \leq Q(\pi_L) + \sum_{l=1}^{L} \sum_{\sigma \in \omega_l} \Delta(\sigma \cup \pi_{l-1}, \pi_{l-1}).$$

Proof. The following inequalities hold true according to the properties of the $\Delta(\cdot, \cdot)$ function:

$$\begin{aligned} Q((\cup_{i=l+1}^{L} \omega_i) \cup \pi_l) &= Q((\cup_{i=l+2}^{L} \omega_i) \cup \pi_l) + \Delta((\cup_{i=l+1}^{L} \omega_i) \cup \pi_l, (\cup_{i=l+2}^{L} \omega_i) \cup \pi_l) \\ &\leq Q((\cup_{i=l+2}^{L} \omega_i) \cup \pi_l) + \Delta(\omega_{l+1} \cup \pi_l, \pi_l) \\ &\leq Q((\cup_{i=l+2}^{L} \omega_i) \cup \pi_{l+1}) + \Delta(\omega_{l+1} \cup \pi_l, \pi_l) \\ &\leq Q((\cup_{i=l+2}^{L} \omega_i) \cup \pi_{l+1}) + \sum_{\sigma \in \omega_{l+1}} \Delta(\sigma \cup \pi_l, \pi_l). \end{aligned}$$

We have $\pi_0 = \emptyset$ and $\omega_{L+1} = \emptyset$ (see Lemma 5.5). With induction from $l = 0$ to $l = L - 1$, we have $Q((\cup_{i=1}^{L} \omega_i) \cup \emptyset) = Q(\Omega)$ and $Q(\Omega) \leq Q(\pi_L) + \sum_{l=1}^{L} \sum_{\sigma \in \omega_l} \Delta(\sigma \cup \pi_{l-1}, \pi_{l-1})$. □

Lemma 5.8. *The maximum size of ω_l is equal to the degree, in the interference graph G_I, of the FBS selected in the l-th step of the greedy algorithm, which is denoted as $D(l)$.*

Proof. Once FBS i is allocated with channel m, the neighboring FBS's in G_I, $\mathcal{R}(i)$, cannot use the same channel m anymore due to the interference constraint. The maximum number of FBS-channel pairs that conflict with the selected FBS-channel pair $\{i, m\}$, i.e., the maximum size of ω_l, is equal to the degree of FBS i in G_I. □

Then we have Theorem 5.2 that provides a lower bound on the objective value achieved by the greedy algorithm given in Table 5.5.

Theorem 5.2. *The greedy algorithm can achieve an objective value that is at least* $\frac{1}{1+D_{max}}$ *of the global optimum, where* D_{max} *is the maximum node degree in the interference graph* G_I *of the femtocell CR network.*

Proof. According to Lemmas 5.7 and 5.8, we have:

$$\begin{aligned} Q(\Omega) &\leq Q(\pi_L) + \sum_{l=1}^{L} D(l)\Delta_l = Q(\pi_L) + \bar{D}\sum_{l=1}^{L}\Delta_l \\ &= (1+\bar{D})Q(\pi_L), \end{aligned} \tag{5.46}$$

where $\bar{D} = \sum_{l=1}^{L} D(l)\Delta_l / \sum_{l=1}^{L}\Delta_l$. The second equality is due to the facts that $\sum_{l=1}^{L}\Delta_l = Q(\pi_L)$.

To further simplify the bound, we replace $D(l)$ with the maximum node degree D_{max}. We then have $\bar{D} \leq \sum_{l=1}^{L} D_{max}\Delta_l / \sum_{l=1}^{L}\Delta_l = D_{max}$ and

$$\frac{1}{1+D_{max}} Q(\Omega) \leq Q(\pi_L) \leq Q(\Omega), \tag{5.47}$$

which provides a lower bound on the performance of the greedy algorithm. □

When there is a single FBS in the CR network, we have $D_{max} = 0$ and $Q(\pi_L) = Q(\Omega)$ according to Theorem 5.2. The proposed algorithm produces the optimal solution. In the case of multiple non-interfering FBS's, we still have $D_{max} = 0$ and can obtain the optimal solution using the proposed algorithm. For the femtocell CR network given in Fig. 5.5 (with interference graph

shown in Fig. 5.7), we have $D_{max} = 1$ and the low bound is a half of the global optimal. Note that (5.46) provides a tighter bound for the optimum than (5.47), but with higher complexity. These are interesting performance bounds since they bound the achievable video quality, an application layer performance measure, rather than lower layer metrics (e.g., bandwidth or time share).

5.4.3 Simulation Results

We evaluate the performance of the proposed algorithms using MATLAB and JSVM 9.13 Video codec. Two scenarios are used in the simulations: a single FBS CR network and a CR network with interfering FBS's. In every simulation, we compare the proposed algorithms with the following three more straightforward heuristic schemes:

- Heuristic 1 based on *equal allocation*: each CR user chooses the better channel (i.e., the common channel or a licensed channel) based on the channel conditions; time slots are equally allocated among active CR users;
- Heuristic 2 exploiting *multiuser diversity*: the MBS and each FBS chooses one active CR user with the best channel condition; the entire time slot is allocated to the selected CR user.
- *SCA-MAC* proposed in [23]: with this scheme, the successful transmission rate is evaluated based on channel packet loss rate and collision probability with primary users; the channel-user pair with the highest transmission probability is selected.

We choose SCA-MAC because it adopts similar models and assumptions as in this paper. Once the channels are selected, the same distributed algorithm is used for scheduling video data for all the three schemes.

We adopt the Raleigh block fading model and the packet loss probability is between [0.004, 0.028]. The frame rate is set to 30 fps and the GoP size is 16. The base layer mode is set to be AVC compatible. The motion search mode is set to Fast Search with search range 32. Each point in the figures presented in this section is the average of 10 simulation runs with different random seeds. We plot 95% confidence intervals in the figures, which are generally negligible.

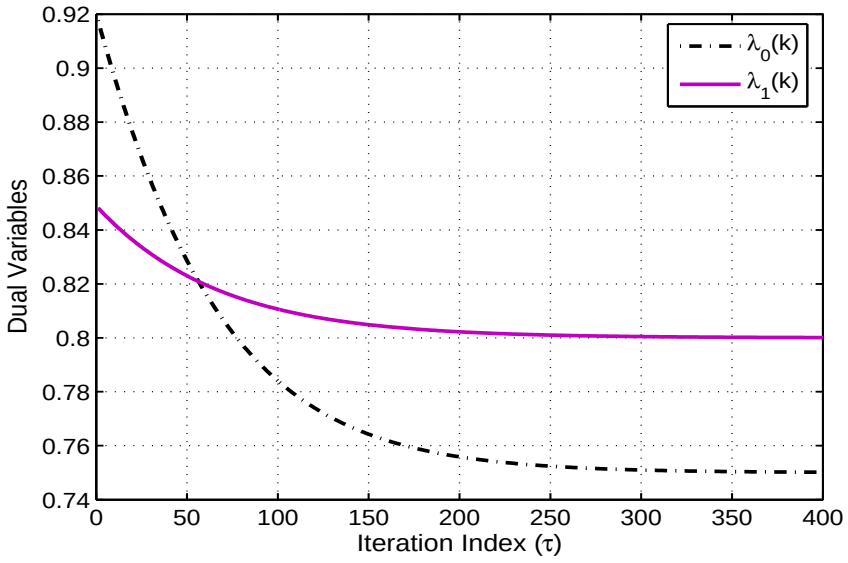


Figure 5.8: Convergence of the two dual variables in the single FBS case.

Case of Single FBS

In the first scenario, there are $M = 8$ channels and the channel parameters P_{01}^m and P_{10}^m are set to 0.4 and 0.3, respectively, for all m. The maximum allowable collision probability γ_m is set to 0.2 for all m. There is one FBS and three active CR users. Three Common Intermediate Format (CIF, 352×288) video sequences are streamed to the CR users, i.e., *Bus* to CR user 1, *Mobile* to CR user 2, and *Harbor* to CR user 3. We have $T = 10$ as the delivery deadline. Both probabilities of false alarm ϵ and miss detection δ are set to 0.3 for all the FBS's and CR users, unless otherwise specified.

First we investigate the convergence of the distributed algorithm. The traces of the two dual variables are plotted in Fig. 5.8. To improve the convergence speed, the correlation in adjacent time slots can be exploited. In particular, we set the optimal values for the optimization variables in the previous time slot as the initialization values for the variables in the current time slot. By doing so, the convergence speed can be improved. It can be seen that both dual variables converge to their optimal values after 300 iterations. After convergence, the optimal solution for the primary problem can be obtained.

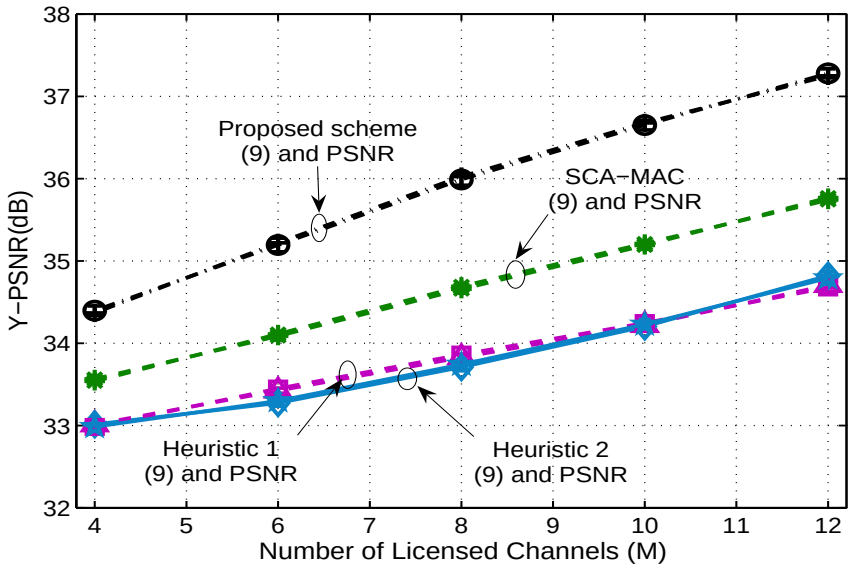


Figure 5.9: Single FBS: received video quality vs. number of channels (computed with (9) and measured by PSNR).

Our proposed scheme achieves the best performance among the three algorithms, with up to 4.3 dB improvement over the two heuristic schemes and up to 2.5 dB over SCA-MAC. Such gains are significant with regard to video quality, since a 0.5 dB difference is distinguishable by human eyes. Compared to the two heuristic schemes and SCA-MAC, the video quality of our proposed scheme is well balanced among the three users, indicating better fairness performance.

In Fig. 5.9, we examine the impact of the number of channels M on received video quality. First, we validate the video quality measure used in our formulation by comparing the PSNR value computed using (5.32) with that computed from real decoded video frames. The average PSNR for three received videos are plotted in the figure. It can be seen that the real PSNRs are very close to those predicted by (5.32), with overlapping confidence intervals. This is also consistent with the results shown in Fig. 5.6. Second, as expected, the more licensed channels, the more spectrum opportunities for CR users and the higher PSNR for received videos. SCA-MAC performs better than two heuristics, but is inferior to the proposed scheme.

We also plot the MS-SSIM of the received videos at the three CR users in Fig. 5.10 [112]. Similar observations can be made from the MS-SSIM plot. All MS-SSIMs for the four curves are more than 0.97 and very close to 1. The proposed scheme still outperforms the other three

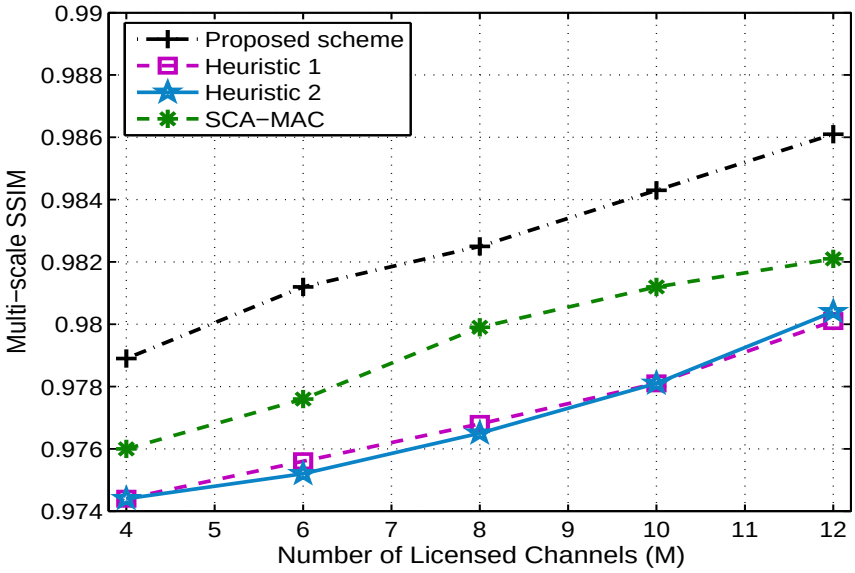


Figure 5.10: Single FBS: received video quality vs. number of channels (measured by MS-SSIM).

schemes. In the remaining figures, we will use model predicted PSNR values, since the model (5.32) is sufficient to predict the real video quality.

In Fig. 5.11, we demonstrate the impact of channel utilization η on received video quality. The average PSNRs achieved by the four schemes are plotted when η is increased from 0.3 to 0.7. Intuitively, a smaller η allows more spectrum opportunities for video transmission. This is illustrated in the figure where all the three curves decrease as η gets larger. The performance of both heuristics are close and the proposed scheme achieves a gain about 3 dB over the heuristics and 2 dB over SCA-MAC.

We also compare the MGS and FGS videos while keeping other parameters identical. We find that MGS video achieves over 0.5 dB gain in video quality over FGS video. The results are omitted for brevity.

Case of Interfering FBS's

We next investigate the second scenario with three FBS's, and each FBS has three active CR users. Each FBS streams three different videos to the corresponding CR users. The coverages of FBS 1 and 2 overlap with each other, and the coverages of FBS 2 and 3 overlap with each other.

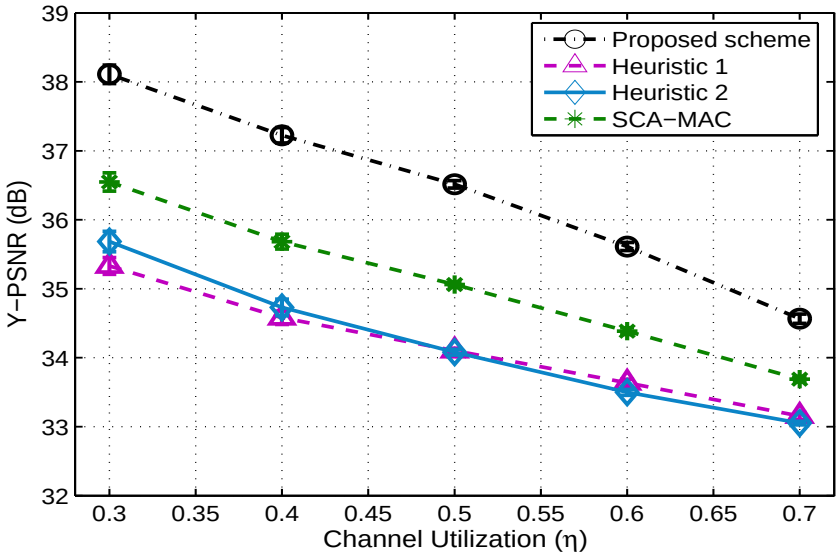


Figure 5.11: Single FBS: received video quality vs. channel utilization.

In Fig. 5.12, we examine the impact of the number of channels M on the received video quality. The average PSNRs of all the active CR users are plotted in the figure when we increase M from 12 to 20 with step size 2. As mentioned before, more channels imply more transmission opportunities for video transmission. In this scenario, heuristic 2 (with a multiuser diversity approach) outperforms heuristic 1 (with an equal allocation approach). But its PSNRs are still about $0.3 \sim 0.5$ dB lower that those of the proposed algorithm. The proposed scheme has up to 0.4 dB improvement over SCA-MAC. In Fig. 5.12, we also plot an upper bound on the optimal objective value, which is obtained as in (5.46). It can be seen that the performance of our proposed scheme is close to optimal solution since the gap between the upper bound and our scheme is generally small (about 0.5 dB).

Next, we examine the impact of sensing errors on the received video quality. In Fig. 5.13, we test five pairs of $\{\epsilon, \delta\}$ values: $\{0.2,0.48\}$, $\{0.24,0.38\}$, $\{0.3,0.3\}$, $\{0.38,0.24\}$, and $\{0.48,0.2\}$. It is interesting to see that the performance of all the four schemes get worse when the probability of one of the two sensing errors gets large. We can trade-off between false alarm and miss detection probabilities to find the optimal operating point for the spectrum sensors. Moreover, the dynamic range of video quality is not big for the range of sensing errors simulated, compared to that in

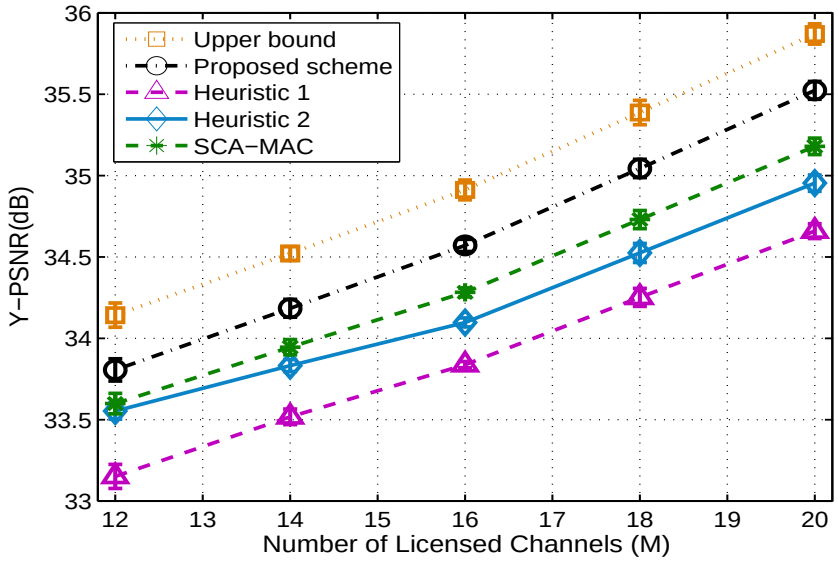


Figure 5.12: Interfering FBS's: received video quality vs. number of channels.

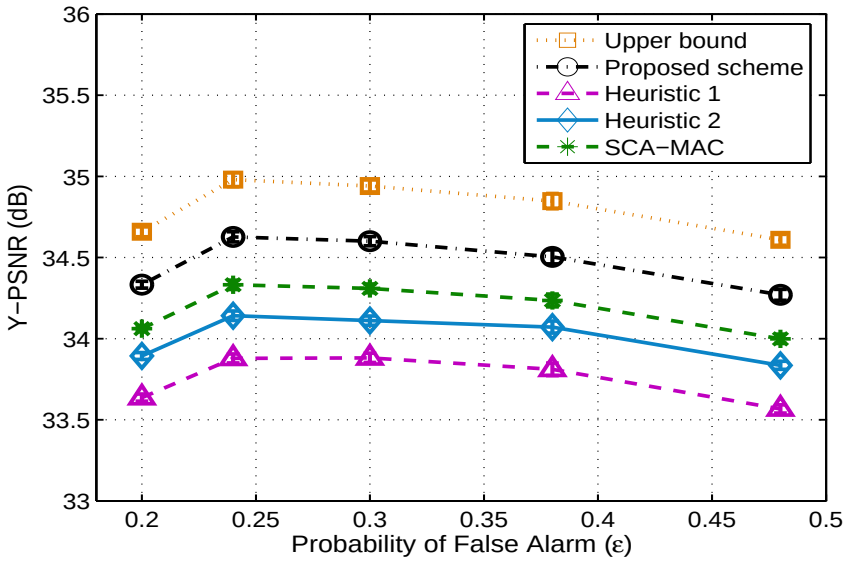


Figure 5.13: Interfering FBS's: received video quality vs. sensing error probability.

Fig. 5.12. This is because both sensing errors are modeled and treated in the algorithms. Again, our proposed scheme outperforms the two heuristic schemes and SCA-MAC with considerable margins for the entire range.

We also investigate the impact of the bandwidth of the common channel B_0. In this simulation, we fix B_1 at 0.3 Mbps and increase B_0 from 0.1 Mbps to 0.5 Mbps with step size 0.1 Mbps. The

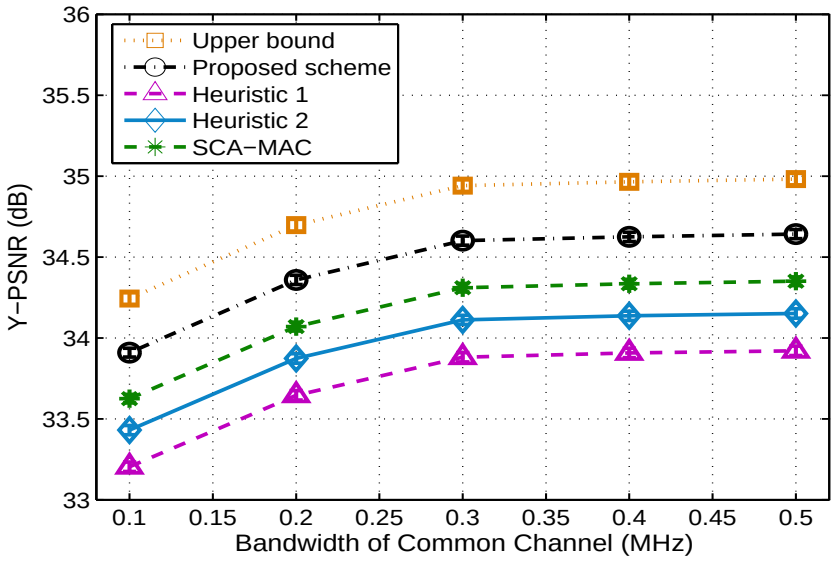


Figure 5.14: Interfering FBS's: received video quality vs. bandwidth of the common channel.

results are presented in Fig. 5.14. We notice that the average video quality increases rapidly as the common channel bandwidth is increased from 0.1 Mbps to 0.3 Mbps. Beyond 0.3 Mbps, the increases of the PSNR curves slow down and the curves get flat. This implies that a very large bandwidth for the common channel is not necessary, since the gain for additional bandwidth diminishes as B_0 gets large. Again, the proposed scheme outperforms the other three schemes and the gap between our scheme and the upper bound is small.

Next, we stop the distributed algorithm after a fixed amount of time, and evaluate the suboptimal solutions. In particular, we vary the duration of time slots, and let the distributed algorithm run for 5% of the time slot duration at the beginning of the time slot. Then the solution obtained this way will be used for the video data transmissions. The results are presented in Fig. 5.15. It can be seen that when the time slot is 5 ms, the algorithm does not converge after 5%$\times$5 = 0.25 ms and the PSNR produced by the distributed algorithm is close to that of Heuristic 1, and lower than those of Heuristic 2 and SCA-MAC. When the time slot is sufficiently large, the algorithm can get closer to the optimal and the proposed algorithm produces better video quality as compared to the two heuristic algorithms and SCA-MAC. Beyond 20 ms, the increase in PSNR is small since all

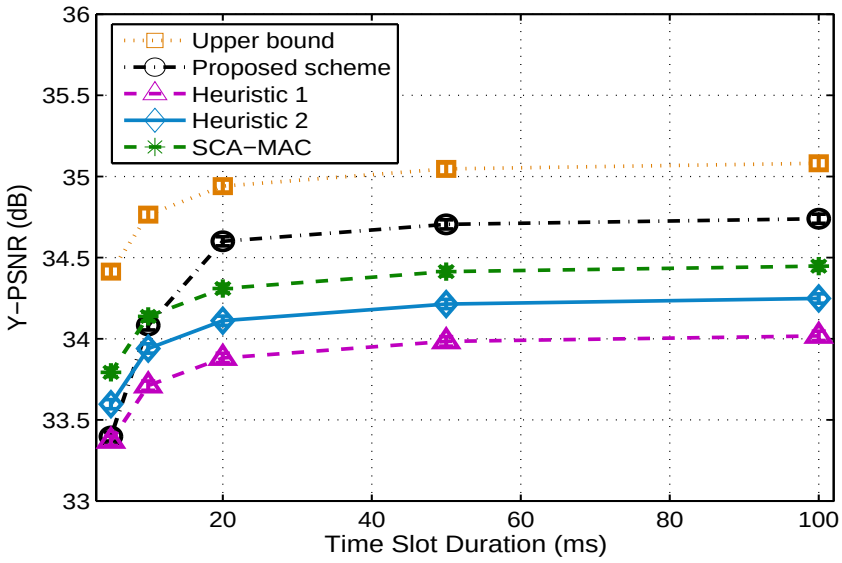


Figure 5.15: Video quality achieved by the algorithms when they are only executed for 5% of the time slot duration.

the curves gets flat. Therefore the proposed algorithm could be useful even when there is no time for it to fully converge to the optimal.

During the simulations, we find the collision rate with primary users are strictly kept below the prescribed collision tolerance γ. These results are omitted for brevity.

5.5 Conclusions

In this chapter, we first addressed the problem of multicasting FGS video in CR networks. The problem formulation took video quality and proportional fairness as objectives, while considering cross-layer design factors such as FGS coding, spectrum sensing, opportunistic spectrum access, primary user protection, scheduling, error control and modulation. We proposed efficient optimization and scheduling algorithms for highly competitive solutions, and proved the complexity and optimality bound of the proposed greedy algorithm. Our simulation results demonstrate not only the viability of video over CR networks, but also the efficacy of the proposed approach.

Then, we studied the challenging problem of streaming multiple scalable videos in a multi-hop CR network. The problem formulation considered spectrum sensing and sensing errors, spectrum

access and primary user protection, video quality and fairness, and channel/path selection for concurrent video sessions. We first solved the formulated MINLP problem using a sequential fixing scheme that produces lower and upper bounds on the achievable video quality. We then applied dual decomposition to derive a distributed algorithm, and analyzed its optimality and convergence performance. Our simulations validated the efficacy of the proposed scheme.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In the previous chapters, we investigated several challenging problems of effective CR networking and scalable video streaming over CR networks. Our research consisted of network modeling, cross-layer design and optimization, performance analysis, algorithm development, and simulation validation.

In Chapter 2, we first studied the problem of design and analysis of MAC protocols for CR networks. We explicitly considered sensing errors in the design of MAC protocols and developed analytical models to evaluate the performance of the proposed protocols. In the second part, we considered the problem of interference mitigation via channel assignment and power allocation for CR users. We proposed a distributed greedy algorithm that only needs local channel gain information. It was shown to outperform other two alternatives via simulations.

In Chapter 3, a more challenging problem, scalable video streaming in CR networks, was investigated. We first studied the problem of scalable video multicast over infrastructure based CR networks. We proposed an efficient greedy algorithm with proved complexity and optimality bound. Then, we considered the problem of streaming multiple videos over multi-hop CR networks. We developed a distributed algorithm by applying dual decomposition and proved its optimality and convergence conditions.

In Chapter 4, we first investigated the problem of cooperative relay in CR networks. We compared two typical cooperative relay strategies and developed an analysis for the comparison. We found each of the strategies performed better in a certain parameter range and diversity gain was achieved by cooperative relays. Then, we investigated the problem of interference alignment for MGS video streaming in a cooperative relay enhanced CR network. We developed a distributed

optimal algorithm for the case of a single channel and the case of multi-channel bonding, with proven convergence and convergence speed. We also proposed a greedy algorithm for the multi-channel without channel bonding case, with a proven performance bound.

In Chapter 5, we first investigated the problem of data multicast in femtocell networks that incorporates SC and SIC. We developed optimal and near-optimal algorithms with low computational complexity, as well as performance bounds. Then, we tackled the problem of streaming multiple MGS videos in a femtocell CR networks. A distributed optimal algorithm was developed in the case of non-interfering FBS's and a greedy algorithm for near-optimal solutions was proposed in the case of interfering FBS's with proved lower bound.

6.2 Summary of Contributions

Wireless video has been a challenging area with considerable research efforts. However, video over CR networks has not been well studied, since the main stream CR research has focused on spectrum sensing and access. It was not clear if video can be offered in such highly dynamic networks even a few years ago. There is a compelling need for innovative research in this area given the Cisco prediction of exploding wireless video traffic in the next few years.

In this dissertation, we investigated the problem of effective CR networking with application to multi-user video communications over four emerging CR networking paradigms, including infrastructure-based CR networks, multi-hop ad hoc CR networks, CR femtocell networks, and relay-assisted CR networks. This research provides in-depth treatment of the problems with both theory and algorithm ingredients. The findings not only successfully demonstrate the feasibility of video CR networks, but also shed useful insights on developing practical CR video systems.

6.3 Open Problems and Future Work

Although considerable progresses have been made through this dissertation work in the area of video over CR networks, there are many interesting open problems to be explored in this important problem area. Some of such open problems are briefly described below, which we plan to investigate in our future research.

In most of our prior work, we assumed that the occupancy of each licensed channel evolves over time following a two state discrete-time Markov process and the primary user activities on different channels are independent. Although this assumption makes the problem manageable, it may not hold true in certain CR networks. The primary user transmission may be modeled as a more general process, while the primary transmissions on different channels may be statistically correlated. Thus, a more sophisticate spectrum sensing and access scheme is required to be integrated into the cross-layer optimization framework. The accuracy of the sensing process could be improved by exploiting the sensing results from adjacent channels and historic sensing results.

Similarly, another assumption is that the lowest video quality requirement for CR users can always be guaranteed. However, the network capacity for CR users strongly depends on both the primary user transmissions and randomly fading and shadowing channels. Therefore, an admission control mechanism is required that can estimate the level of QoS that a new video session will have and whether there is enough bandwidth available to serve that session. A simple yet efficient admission control mechanism that considers both primary user activities and channel conditions is essential for QoS provisioning for video over CR networks.

We investigated several challenging problems in CR networks using video as a reference application. In an operating CR network, there will be multiple applications that generate different types of traffic flows, all sharing the extra bandwidth provided by CRs. It is thus interesting to investigate how to provide quality of service (QoS) guarantees for different traffic flows each with different characteristics and different QoS requirements. This is a general problem for both wireline and wireless networks. The Internet adopts the Integrated Services (intserv) and Differentiated Services (diffserv) approaches to address this problem. We conjecture a certain classification

scheme should be adopted to intelligently identify and classify application traffic according to their characteristics and QoS requirements, and a resource allocation scheme will be used to treat the different classes of traffic flows differently. These are interesting problems that worth further investigation.

We presented a theoretical framework for video streaming in CR networks and demonstrated the performance with extensive simulations. In the future, we are interested in building a CR video streaming testbed network, such that the system performance can be demonstrated under a realistic wireless environment. Our research will focus on the combination of hardware components (e.g. USRP models) and software techniques (e.g. network optimization algorithms). Such a CR video testbed can not only validate the theoretical results, but also reveal new practical constraints that should be considered in the modeling and analysis, as well identifying new research problems.

Bibliography

[1] I. Akyildiz, W. Lee, M. Vuran, and S. Mohanty, "NeXt Generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Elsevier Computer Networks*, vol. 50, no. 13, pp. 2127–2159, Sept. 2006.

[2] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2009-2014," Feb. 2010, [online] Available: www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html.

[3] Q. Zhang, J. Jia, and J. Zhang, "Cooperative relay to improve diversity in cognitive radio networks," *IEEE Commun. Mag.*, vol. 47, no. 2, pp. 111–117, Feb. 2009.

[4] J. Jia, J. Zhang, and Q. Zhang, "Cooperative relay for cognitive radio networks," in *Proc. IEEE INFOCOM'09*, Apr. 2009, pp. 2304–2312.

[5] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: a survey," *IEEE Commun. Mag.*, vol. 46, no. 9, pp. 59–67, Sept. 2008.

[6] D. Hu and S. Mao, "Design and analysis of a sensing error-aware MAC protocol for cognitive radio networks," in *Proc. IEEE GLOBECOM'09*, Honolulu, HI, Nov./Dec. 2009.

[7] ——, "Co-channel and adjacent channel interference mitigation in cognitive radio networks," in *Proc. IEEE MILCOM'11*, Baltimore, MD, Nov. 2011, pp. 1–6.

[8] D. Hu, S. Mao, Y. Hou, and J. Reed, "Fine grained scalability video multicast in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 3, pp. 334–344, Apr. 2010.

[9] D. Hu and S. Mao, "Streaming scalable videos over multi-hop cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3501–3511, Nov. 2010.

[10] ——, "Cooperative relay in cognitive radio networks: Decode-and-forward or amplify-and-forward?" in *IEEE GLOBECOM'10*, Miami, FL, Dec. 2010, pp. 1–5.

[11] ——, "Cooperative relay with interference alignment for video over cognitive radio networks," in *Proc. IEEE INFOCOM'12*, Orlando, FL, Mar. 2012.

[12] ——, "Multicast in femtocell networks: a successive interference cancellation approach," in *Proc. IEEE GLOBECOM'11*, Huston, TX, Dec. 2011, pp. 1–6.

[13] ——, "On medium grain scalable video streaming over cognitive radio femtocell networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, Apr. 2012.

[14] Q. Zhao and B. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.

[15] Y. Zhao, S. Mao, J. Neel, and J. H. Reed, "Performance evaluation of cognitive radios: Metrics, utility functions, and methodologies," *Proc. IEEE, Special Issue on Cognitive Radio*, vol. 97, no. 4, pp. 642–659, Apr. 2009.

[16] A. Motamedi and A. Bahai, "MAC protocol design for spectrum-agile wireless networks: Stochastic control approach," in *Proc. IEEE DySPAN'07*, Dublin, Ireland, Apr. 2007, pp. 448–451.

[17] S. Geirhofer, L. Tong, and B. Sadler, "Cognitive medium access: constraining interference based on experimental models," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 95–105, Jan. 2008.

[18] L. Le and E. Hossain, "OSA-MAC: A MAC protocol for opportunistic spectrum access in cognitive radio networks," in *Proc. IEEE WCNC'08*, Las Vegas, NV, Mar./Apr. 2008, pp. 1426–1430.

[19] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 589–600, Apr. 2007.

[20] H. Su and X. Zhang, "Cross-layer based opportunistic MAC protocols for QoS provisionings over cognitive radio wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 118–129, Jan. 2008.

[21] J. Jia, Q. Zhang, and X. Shen, "HC-MAC: A hardware-constrained cognitive MAC for efficient spectrum management," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 106–117, Jan. 2008.

[22] B. Hamdaoui and K. Shin, "OS-MAC: An efficient MAC protocol for spectrum-agile wireless networks," *IEEE Trans. Mobile Comput.*, vol. 7, no. 8, pp. 915–930, Aug. 2008.

[23] A. Hsu, D. Wei, and C. Kuo, "A cognitive MAC protocol using statistical channel allocation for wireless ad-hoc networks," in *Proc. IEEE WCNC'07*, Hong Kong, P.R. China, Mar. 2007, pp. 105–110.

[24] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2053–2071, May 2008.

[25] Y. R. Kondareddy and P. Agrawal, "Synchronized mac protocol for multi-hop cognitive radio networks," in *Proc. IEEE Int. Conf. Communications ICC '08*, Beijing, China, May 2008, pp. 3198–3202.

[26] Y. R. Kondareddy, P. Agrawal, and K. Sivalingam, "Cognitive radio network setup without a common control channel," in *Proc. IEEE Military Communications Conf. MILCOM 2008*, San Diego, CA, Nov. 2008, pp. 1–6.

[27] L. Yang and M.-S. Alouini, "Performance comparison of different selection combining algorithms in presence of co-channel interference," *IEEE Trans. Veh. Technol.*, vol. 55, no. 2, pp. 559–571, Mar. 2006.

[28] J. Nachtigall, A. Zubow, and J.-P. Redlich, "The impact of adjacent channel interference in multi-radio systems using IEEE 802.11," in *Proc. IWCMC'08*, Crete Island, Greece, Aug. 2008, pp. 874–881.

[29] V. Angelakis, S. Papadakis, V. A. Siris, and A. Traganitis, "Adjacent channel interference in 802.11a is harmful: Testbed validation of a simple quantification model," *IEEE Commun. Mag.*, vol. 49, no. 3, pp. 160–166, Mar. 2011.

[30] M. G. Sanchez, M. A. Acuna, I. Cuinas, R. M. Rodriguez-Osorio, L. de Haro, and A. Garcia-Pino, "Cochannel and adjacent channel interference in actual terrestrial TV scenarios - part I: Field measurements," *IEEE Trans. Broadcast.*, vol. 48, no. 2, pp. 111–115, June 2002.

[31] Z. A. Shamsan, L. F. Abdulrazak, and T. A. Rahman, "Co-channel and adjacent channel interference evaluation for IMT-Advanced coexistence with existing fixed systems," in *Proc. IEEE International RF and Microwave Conference 2008*, Kuala Lumpur, Malaysia, Dec. 2008, pp. 65–69.

[32] C.-S. Sum, R. Funada, J. Wang, T. Baykas, M. A. Rahman, and H. Harada, "Error performance and throughput evaluation of a multi-Gbps millimeter-wave WPAN system in the presence of adjacent and co-channel interference," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1433–1442, Oct. 2009.

[33] K. Gulati, B. Evans, J. Andrews, and K. Tinsley, "Statistics of co-channel interference in a field of Poisson and Poisson-Poisson clustered interferers," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6207–6222, Dec. 2010.

[34] E. Obregon, L. Shi, J. Ferrer, and J. Zander, "A model for aggregate adjacent channel interference in TV white space," in *Proc. IEEE VTC-Spring 2011*, Budapest, Hungary, May 2011, pp. 1–5.

[35] H.-C. Lee, D.-C. Oh, and Y.-H. Lee, "Mitigation of inter-femtocell interference with adaptive fractional frequency reuse," in *Proc. IEEE ICC'10*, Cape Town, South Africa, May 2010, pp. 1–5.

[36] A. M. A. Ahmed and I. Marsland, "Co-channel interference cancellation in multihop relay networks," in *Proc. IEEE ICC'08 Workshops*, Beijing, P.R. China, May 2008, pp. 62–67.

[37] ——, "Co-channel interference cancellation in wireless cellular networks," in *Proc. IEEE VTC-Spring 2008*, Marina Bay, Singapore, May 2008, pp. 698–702.

[38] G. Miao, Y. Li, N. Himayat, and S. Talwar, "Co-channel interference avoidance MAC in wireless cellular networks," *IEEE Trans. Commun.*, vol. 57, no. 11, pp. 3897–3405, Nov. 2009.

[39] O. Bendov and C. B. Patel, "Television receiver optimization in the presence of adjacent channel interference," *IEEE Trans. Broadcast.*, vol. 51, no. 1, pp. 38–42, Mar. 2005.

[40] D. Gidony and I. Kalet, "Adjacent channel interference cancellation for MSK-type signals using antenna diversity in Rayleigh fading environment," *IEEE Trans. Commun.*, vol. 52, no. 2, pp. 317–325, Feb. 2004.

[41] A. Babaei and B. Jabbari, "Interference modeling and avoidance in spectrum underlay cognitive wireless networks," in *Proc. IEEE Int Communications (ICC) Conf*, 2010, pp. 1–5.

[42] D. Hu, S. Mao, and J. Reed, "On video multicast in cognitive radio networks," in *Proc. IEEE INFOCOM'09*, Rio de Janeiro, Brazil, Apr. 2009.

[43] H. Nan, T.-I. Hyon, and S.-J. Yoo, "Distributed coordinated spectrum sharing MAC protocol for cognitive radio," in *Proc. IEEE DySPAN'07*, Dublin, Ireland, Apr. 2007, pp. 240–249.

[44] C. Corderio, K. Challapali, D. Birru, and S. Shankar, "IEEE 802.22: An introduction to the first wireless standard based on cognitive radios," *J. Commun.*, vol. 1, no. 1, pp. 38–47, Apr. 2006.

[45] S. Kompella, S. Mao, Y. T. Hou, and H. D. Sherali, "On path selection and rate allocation for video in wireless mesh networks," *IEEE/ACM Trans. Netw.*, vol. 17, no. 1, pp. 212–224, Feb. 2009.

[46] D. Hu and S. Mao, "Resource allocation for medium grain scalable videos over femtocell cognitive radio networks," in *Proc. IEEE ICDCS'11*, Minneapolis, MN, June 2011, pp. 258–267.

[47] Y. Hou, Y. Shi, and H. Sherali, "Spectrum sharing for multi-hop networking with cognitive radios," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 146–155, Jan. 2008.

[48] M. S. Bazaraa, J. J. Jarvis, and H. D. Sherali, *Linear Programming and Network Flows*, 4th ed. New York, NY: John Wiley & Sons, Inc., 2010.

[49] S. Mao, S. Lin, Y. Wang, S. S. Panwar, and Y. Li, "Multipath video transport over wireless ad hoc networks," *IEEE Wireless Commun.*, vol. 12, no. 4, pp. 42–49, Aug. 2005.

[50] S. Mao, S. Kompella, Y. T. Hou, H. D. Sherali, and S. F. Midkiff, "Routing for multiple concurrent video sessions in wireless ad hoc networks," in *Proc. IEEE ICC'05*, Seoul, Korea, May 2005, pp. 1229–1235.

[51] IEEE, "Draft standard for wireless regional area networks part 22: Cognitive wireless RAN medium access control (MAC) and physical layer (PHY) specifications: Policies and procedures for operation in the TV bands," May 2007, IEEE P802.22 Draft Standard (D0.3).

[52] M. van der Schaar, S. Krishnamachari, S. Choi, and X. Xu, "Adaptive cross-layer protection strategies for robust scalable video transmission over 802.11 WLANs," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 10, pp. 1752–1763, Dec. 2003.

[53] M. Wien, H. Schwarz, and T. Oelbaum, "Performance analysis of SVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1194–1203, Sept. 2007.

[54] N. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 3062–3080, Nov. 2004.

[55] R. Urgaonkar and M. J. Neely, "Opportunistic scheduling with reliability guarantees in cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 8, no. 6, pp. 766–777, June 2009.

[56] T. Shu and M. Krunz, "Throughput-efficient sequential channel sensing and probing in cognitive radio networks under sensing errors," in *Proc. ACM MobiCom'09*, Beijing, China, Sept. 2009, pp. 37–48.

[57] Q. Zhao, S. Geirhofer, L. Tong, and B. Sadler, "Opportunistic spectrum access via periodic channel sensing," *IEEE Trans. Signal Process.*, vol. 36, no. 2, pp. 785–796, Feb. 2008.

[58] A. Fattahi, F. Fu, M. van der Schaar, and F. Paganni, "Mechanism-based resource allocation for multimedia transmission over spectrum agile wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 601–612, Apr. 2007.

[59] H.-P. Shiang and M. van der Schaar, "Dynamic channel selection for multi-user video streaming over cognitive radio networks," in *Proc. IEEE ICIP'08*, San Diego, CA, Oct. 2008, pp. 2316–2319.

[60] H. Luo, S. Ci, and D. Wu, "A cross-layer design for the performance improvement of real-time video transmission of secondary users over cognitive radio networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 8, pp. 1040–1048, Aug. 2011.

[61] S. Ali and F. Yu, "Cross-layer QoS provisioning for multimedia transmissions in cognitive radio networks," in *Proc. IEEE WCNC'09*, Budapest, Hungary, Apr. 2009, pp. 1–5.

[62] Z. Guan, L. Ding, T. Melodia, and D. Yuan, "On the effect of cooperative relaying on the performance of video streaming applications in cognitive radio networks," in *Proc. IEEE ICC'11*, Kyoto, Japan, June 2011, pp. 1–6.

[63] S. Mao, X. Cheng, Y. Hou, and H. Sherali, "Multiple description video multicast in wireless ad hoc networks," *ACM/Kluwer Mobile Netw. Appl. J.*, vol. 11, no. 1, pp. 63–73, Jan. 2006.

[64] W. Wei and A. Zakhor, "Multiple tree video multicast over wireless ad hoc networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 1, pp. 2–15, Jan. 2007.

[65] S. Deb, S. Jaiswal, and K. Nagaraj, "Real-time video multicast in WiMAX networks," in *Proc. IEEE INFOCOM'08*, Phoenix, AZ, Apr. 2008, pp. 1579–1587.

[66] Y. Hou, Y. Shi, and H. Sherali, "Optimal spectrum sharing for multi-hop software defined radio networks," in *Proc. IEEE INFOCOM'07*, Anchorage, AK, Apr. 2007, pp. 1–9.

[67] Z. Feng and Y. Yang, "Joint transport, routing and spectrum sharing optimization for wireless networks with frequency-agile radios," in *Proc. IEEE INFOCOM'09*, Rio de Janeiro, Brazil, Apr. 2009, pp. 1665–1673.

[68] D. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.

[69] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Nashua, NH: Athena Scientific, 1999.

[70] H. Mahmoud, T. Yücek, and H. Arslan, "OFDM for cognitive radio: Merits and challenges," *IEEE Wireless Commun.*, vol. 16, no. 2, pp. 6–14, Apr. 2009.

[71] F. Kelly, A. Maulloo, and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *J. Operational Research Society*, vol. 49, no. 3, pp. 237–252, Mar. 1998.

[72] Y. Hou, Y. Shi, and H. Sherali, "Optimal base station selection for anycast routing in wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 3, pp. 813–821, May 2006.

[73] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Stable throughput of cognitive radios with and without relaying capability," *IEEE Trans. Commun.*, vol. 55, no. 12, pp. 2351–2360, Dec. 2007.

[74] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, UK: Cambridge University Press, 2005.

[75] V. Cadambe and S. A. Jafar, "Interference alignment and the degrees of freedom for the k user ntererence channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, May 2008.

[76] L. E. Li, R. Alimi, D. Shen, H. Viswanathan, and Y. R. Yang, "A general algorithm for interference alignment and cancellation in wireless networks," in *Proc. IEEE INFOCOM 2010*, San Diego, CA, Mar. 2010.

[77] G. Strang, *Introduction to Linear Algebra*, 4th ed. Wellesley, MA: Wellesley Cambridge Press, 2009.

[78] T. Cover and A. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. on Info. Theory*, vol. 25, no. 5, pp. 572–584, Sept. 1979.

[79] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity - Part I: System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

[80] ——, "User cooperation diversity - Part II: Implementation aspects and performance analysis," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1939–1948, Nov. 2003.

[81] M. Khojastepour, A. Sabharwal, and B. Aazhang, "On capacity of Gaussian 'cheap' relay channel," in *Proc. IEEE GLOBECOM'03*, San Francisco, CA, Dec. 2003, pp. 1776–1780.

[82] Y. Zhao, R. Adve, and T. Lim, "Improving amplify-and-forward relay networks: Optimal power allocation versus selection," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 3114–3123, Aug. 2007.

[83] A. Bletsas, A. Khisti, D. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 659–672, Mar. 2006.

[84] T. C.-Y. Ng and W. Yu, "Joint optimization of relay strategies and resource allocations in cooperative cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 328–339, Feb. 2007.

[85] J. Cai, X. Shen, J. Mark, and A. Alfa, "Semi-distributed user relaying algorithm for amplify-and-forward wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1348–1357, Apr. 2008.

[86] Y. Shi, S. Sharma, Y. Hou, and S. Kompella, "Optimal relay assignment for cooperative communications," in *Proc. ACM MobiHoc'08*, Hong Kong, P. R. China, May 2008, pp. 3–12.

[87] L. Ding, T. Melodia, S. Batalama, and J. Matyjas, "Distributed routing, relay selection, and spectrum allocation in cognitive and cooperative ad hoc networks," in *IEEE SECON'10*, Boston, MA, June 2010, pp. 1–9.

[88] L. Ding, S. Pudlewski, T. Melodia, S. Batalama, J. Matyjas, and M. Medley, "Distributed spectrum sharing for video streaming in cognitive radio ad hoc networks," in *Intl. Workshop on Cross-layer Design in Wireless Mobile Ad Hoc Networks*, Niagara Falls, Canada, Sept. 2009, pp. 1–13.

[89] S. Katti, S. Gollakota, and D. Katabi, "Embracing wireless interference: Analog network coding," in *Proc. ACM SIGCOMM'07*, Kyoto, Japan, Aug. 2007, pp. 397–408.

[90] S. Gollakota, S. David, and D. Katabi, "Interference alignment and cancellation," in *Proc. ACM SIGCOMM'09*, Barcelona, Spain, Aug. 2009, pp. 159–170.

[91] I. K. Son and S. Mao, "Design and optimization of a tiered wireless access network," in *Proc. IEEE INFOCOM 2010*, San Diego, CA, Mar. 2010, pp. 1–9.

[92] R. Kim, J. S. Kwak, and K. Etemad, "WiMAX femtocell: requirements, challenges, and solutions," *IEEE Commun. Mag.*, vol. 47, no. 9, pp. 84–91, Sept. 2009.

[93] I. Guvenc, S. Saunders, O. Oyman, H. Claussen, and A. Gatherer, "Femtocell networks," *EURASIP J. Wireless Comm. and Networking*, 2010, article ID 367878, 2 pages, doi:10.1155/2010/367878.

[94] A. Goldsmith, *Wireless Communications*. Cambridge, UK: Cambridge University Press, 2006.

[95] L. Li, R. Alimi, R. Ramjee, H. Viswanathan, and Y. R. Yang, "muNet: Harnessing multiuser capacity in wireless mesh networks," in *Proc. IEEE INFOCOM'09 Mini Symp.*, Rio de Janeiro, Brazil, Apr. 2009, pp. 2876–2880.

[96] V. Chandrasekhar, J. G. Andrews, T. Muharemovic, Z. Shen, and A. Gatherer, "Power control in two-tier femtocell networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4316–4328, Aug. 2009.

[97] F.-S. Chu and K.-C. Chen, "Mitigation of macro-femto co-channel interference by spatial channel separation," in *Proc. IEEE VTC-Spring'11*, Budapest, Hungary, May 2011, pp. 1–5.

[98] S. Rangan, "Femto-macro cellular interference control with subband scheduling and interference cancelation," in *Proc. IEEE GLOBECOM'10 Workshops*, Miami, FL, Dec. 2010, pp. 695–700.

[99] Z. Bharucha, H. Haas, G. Auer, and I. Cosovic, "Femto-cell resource partitioning," in *Proc. IEEE GLOBECOM Workshops*, Honolulu, HI, Dec. 2009, pp. 1–6.

[100] R. Madan, A. Sampath, A. Khandekar, J. Borran, and N. Bhushan, "Distributed interference management and scheduling in LTE-A femto networks," in *Proc. IEEE GLOBECOM'10*, Miami, FL, Dec. 2010, pp. 1–5.

[101] S.-M. Cheng, S.-Y. Lien, F.-S. Chu, and K.-C. Chen, "On exploiting cognitive radio to mitigate interference in macro/femto heterogeneous networks," *IEEE Wireless Commun. Mag.*, vol. 18, no. 3, pp. 40–47, 2011.

[102] S. Kaimaletu, R. Krishnan, S. Kalyani, N. Akhtar, and B. Ramamurthi, "Cognitive interference management in heterogeneous femto-macro cell networks," in *Proc. IEEE ICC'11*, Kyoto, Japan, June 2011, pp. 1–6.

[103] S. Sen, N. Santhapuri, R. R. Choudhury, and S. Nelakuditi, "Successive interference cancellation: A back-of-the-envelope perspective," in *Proc. ACM Hotnets'10*, Monterey, CA, Oct. 2010, pp. 1–6.

[104] C. A. S. Jean and B. Jabbari, "On game-theoretic power control under successive interference cancellation," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1655–1657, Apr. 2009.

[105] C. S. Park and K. B. Lee, "Transmit power allocation for successive interference cancellation in multicode MIMO systems," *IEEE Trans. Commun.*, vol. 56, no. 12, pp. 2200–2213, Dec. 2008.

[106] N. Benvenuto, G. Carnevale, and S. Tomasin, "Joint power control and receiver optimization of CDMA transceivers using successive interference cancellation," *IEEE Trans. Commun.*, vol. 55, no. 3, pp. 563–573, Mar. 2007.

[107] A. Agrawal, J. G. Andrews, J. M. Cioffi, and T. Meng, "Iterative power control for imperfect successive interference cancellation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 878–884, May 2005.

[108] J. G. Andrews and T. H. Meng, “Optimum power control for successive interference cancellation with imperfect channel estimation,” *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 375–383, Mar. 2003.

[109] U. Ewaldsson, “Cut your network’s electricity bill and carbon footprint,” *Global Telecoms Business*, Feb. 28, 2010.

[110] T. Rappaport, *Wireless Communications: Principles & Practice*, 2nd ed. Indianapolis, IN: Prentice Hall PTR, 2001.

[111] Q. Zhang and S. Kassam, “Finite-state Markov model for Rayleigh fading channels,” *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, Nov. 1999.

[112] Z. Wang, L. Lu, and A. C. Bovik, “Video quality assessment using structural distortion measurement,” *Signal Processing: Image Commun.*, no. 2, pp. 121–132, Feb. 2004.

Appendices

Publications from This Dissertation Work

Book Chapter

1. Donglin Hu and Shiwen Mao, "Video over cognitive radio networks: when compression meets the radios," in *Advanced Video Communications over Wireless Networks*, C. Zhu and Y. Li (editors), New York, NY: CRC Press, in press. (39 pages)

Journal Publications

1. Donglin Hu and Shiwen Mao, "A Successive Interference Cancellation approach to multicast in femtocell networks," *IEEE Transactions on Wireless Communications*, under submission.

2. Donglin Hu and Shiwen Mao, "Cooperative relay with interference alignment for video streaming in cognitive radio networks," *IEEE Journal on Selected Areas in Communications - Cognitive Radio Series*, vol.30, no.11, Nov.2012, under review. (11 pages)

3. Donglin Hu and Shiwen Mao, "On co-channel and adjacent channel interference mitigation in cognitive radio networks," *Elsevier Ad Hoc Networks Journal*, under review. (18 pages)

4. Donglin Hu and Shiwen Mao, "On cooperative relay in cognitive radio networks," *ICST Transactions on Mobile Communications and Applications*, in press. (9 pages)

5. Donglin Hu and Shiwen Mao, "A sensing error aware medium access control protocol for cognitive radio networks," *ICST Transactions on Mobile Communications and Applications*, vol.1, no.2, 2012, in press. (10 pages)

6. Donglin Hu and Shiwen Mao, "On medium grain scalable video streaming over femtocell cognitive radio networks," *IEEE Journal on Selected Areas in Communications*, Special Issue on Femtocell Networks, vol.30, no.4, Apr. 2012, in press. (11 pages)

7. Donglin Hu and Shiwen Mao, "Streaming scalable videos over multi-hop cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol.9, no.11, pp.3501–3511, Nov. 2010.

8. Shiwen Mao and Donglin Hu, "Video over cognitive radio networks: when compression meets the radios," invited paper, *E-Letter of Multimedia Communications Technical Committee* (MMTC), IEEE Communications Society, Special Issue on Advances in Network-Adaptive Multimedia Communications, vol.5, no. 6, pp.14–16, Nov. 2010.

9. Donglin Hu, Shiwen Mao, Y. Thomas Hou, and Jeffrey H. Reed, "Fine grained scalability video multicast in cognitive radio networks," *IEEE Journal on Selected Areas in Communications*, Special Issue on Wireless Video Transmission, vol.28, no.3, pp.334–344, Apr. 2010.

Conference Publications

1. Donglin Hu, Shiwen Mao, and Nedret Billor, "On compressed sensing in wireless sensor networks," under review for *IEEE Global Communications Conference* (GLOBECOM) 2012, Anaheim, CA, Nov. 2012. (6 pages)

2. Donglin Hu and Shiwen Mao, "Cooperative relay with interference alignment for video over cognitive radio networks," in *Proc. IEEE IEEE International Conference on Computer Communications* (INFOCOM) 2012, pp.1–9, Orlando, FL, Mar. 2012.

3. Donglin Hu and Shiwen Mao, "Multicast in femtocell networks: a successive interference cancellation approach," in *Proc. IEEE Global Communications Conference* (GLOBECOM) 2011, Houston, TX, pp.1–6, Dec. 2011.

4. Donglin Hu and Shiwen Mao, “Co-channel and adjacent channel interference mitigation in cognitive radio networks,” in *Proc. IEEE Military Communications Conference* (MILCOM) 2011, Baltimore, MD, pp.1–6, Nov. 2011.

5. Donglin Hu and Shiwen Mao, “Resource allocation for medium grain scalable videos over femtocell cognitive radio networks,” in *Proc. The 31st IEEE International Conference on Distributed Computing Systems* (ICDCS) 2011, Minneapolis, MN, pp.258-267, June 2011.

6. Donglin Hu and Shiwen Mao “Cooperative relay in cognitive radio networks: decode-and-forward or amplify-and-forward?” in *Proc. IEEE Global Communications Conference* (GLOBECOM) 2010, Miami, FL, pp.1–5, Dec. 2010.

7. Donglin Hu and Shiwen Mao, “Design and analysis of a sensing error-aware MAC protocol for cognitive radio networks,” in *Proc. IEEE Global Communications Conference* (GLOBECOM) 2009, Honolulu, HI, pp.1–6, Nov./Dec. 2009.

8. Donglin Hu, Shiwen Mao, and Jeffrey H. Reed, “On video multicast in cognitive radio networks,” in *Proc. IEEE International Conference on Computer Communications* (INFOCOM) 2009, Rio de Janeiro, Brazil, pp.2222–2230, Apr. 2009.