DESIGN METHODS FOR OPTIMAL RESOURCE ALLOCATION IN WIRELESS NETWORKS

Mohammad Faisal Uddin

A thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

Concordia University

Montréal, Québec, Canada

January 2012

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: MOHAMMAD FAISAL UDDIN

Entitled: DESIGN METHODS FOR OPTIMAL RESOURCE ALLOCATION IN WIRELESS NETWORKS

and submitted in partial fulfillment of the requirements for the degree of

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Brigitte Jaumard — Chair

Dr. Ekram Hossain — External Examiner

Dr. Mingyuan Chen — External to Program

Dr. Walaa Hamouda — Examiner

Dr. Dongyu Qiu — Examiner

Dr. Chadi Assi — Thesis Supervisor

Approved by

Chair of Department or Graduate Program Director

January, 2012

Dean of Faculty

Abstract

Design Methods for Optimal Resource Allocation in Wireless Networks

Mohammad Faisal Uddin, Ph.D.

Concordia University, 2012

Wireless communications have seen remarkable progress over the past two decades and perceived tremendous success due to their agile nature and capability to provide fast and ubiquitous internet access. Maturation of 3G wireless network services, development of smart-phones and other broadband mobile computing devices however have motivated researchers to design wireless networks with increased capacity and coverage, therefore un-leaching the wireless broadband capabilities. In this thesis, we address two very important design aspects of wireless networks, namely, interference management and control through optimal cross-layer design and channel fading mitigation through relay-assisted cooperative communications. For the former, we address, in the context of wireless network design, the problem of optimally partitioning the spectrum into a set of non-overlapping channels with non uniform spectrum widths and we model the combinatorially complex problem of joint routing, link scheduling, and spectrum allocation as an optimization problem. We use column generation decomposition technique (which decomposes the original problem into a master and a pricing subproblem) for solving the problem optimally. We also propose several sub-optimal methods for efficiently solving the pricing subproblems. For the latter problem, we study the joint problem of relay selection and power allocation in both wireless unicast and multicast cooperative cellular networks. We employ convex optimization technique to model this complex optimization problem and use branch

and bound technique to solve it optimally. We also present sub-optimal methods to reduce the problem complexity and solve it more efficiently.

Acknowledgments

I would like to express my sincere gratitude to all people who supported me throughout the years of my PhD studies and beyond. First and foremost, I am truly indebted to my thesis supervisor, Dr. Chadi Assi, who is not only the most hard working person I have ever come across but also a dedicated and extremely helpful professor. It would have been impossible for me to accomplish this endeavor without his constant support, guidance and motivations.

I'd also like to express my appreciation to all the faculty and people at Concordia University who contributed to my success one way or the other.

During this thesis, I collaborated with several scholars and researchers in the field of optimization and communication networks. These collaborations yielded to authoring high-quality and well received journal and conference papers. In this matter, I am grateful to my co-authors and collaborators, Dr. Hamed Alazemi, Dr. Ali Ghrayeb and Mr. Mohammad Nurujjaman. Furthermore, I was granted a warm and friendly atmosphere in our research lab at Concordia University throughout my PhD. I would like to express my warm thanks to all my office mates, especially Dr. Mohammad Kiaei. I am also grateful to all my friends in Montreal, and other places who supported me during my education and my personal life.

I also appreciate the funding support from the Quebec Funds for Research on Nature and Technology (FQRNT).

My deepest appreciation goes to my parents who raised me with love and highest

possible devotion. I could never achieve my ambitions without their encouragement, understanding, support, and true love. When I was feeling down and frustrated, they are the ones who always believed in me.

Last, my most tender thanks go to my loving wife, Saba. Aside from perhaps myself, no one has been more impacted by this dissertation. I would like to thank her for all her support, constant love and care.

Contents

List of Figures

List of Tables

List of Acronyms

3GPP	3rd Generation Partnership Project
AF	Amplify-and-Forward
AP	Access Point
B&B	Branch and Bound
BER	Bit Error Rate
BS	Base Station
CA	Channel Assignment
CDMA	Code Division Multiple Access
CG	Column Generation
CSI	Channel State Information
CSMA	Carrier sense multiple access
DF	Decode-and-Forward
DSSS	Direct Sequence Spread Spectrum
DSTC	Distributed Space Time Codes
FWFR	Fixed Width Fixed Rate
FWVR	Fixed Width Variable Rate
GSM	Global System for Mobile Communications
I-factor	Interference Factor
IL	Inner Loop
ILP	Integer Linear Programing

IPTV	Internet Protocol television
JRSVW	Joint Routing and Scheduling with Variable Width
KKT	Karush-Kuhn-Tucker
KMP	Knuth-Morris-Pratt
LAN	Local Area Network
LoS	Line of Sight
LP	Linear Programing
LTE	Long Term Evolution
MAC	Medium Access Control
MBMS	Multimedia Broadcast/Multicast Services
MILP	Mixed Integer Linear Programing
MIMO	Multiple Input Multiple Output
MINLP	Mixed Integer Non-Linear Programing
MP	Master Problem
NIC	Network Interface Card
NLP	Non-Linear Programing
NP	Nondeterministic Polynomial Time
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OL	Outer Loop
OPT	Optimal
QoS	Quality of Service
RMP	Restricted Master Problem
s-d	Source-Destination
SA	Simulated Annealing
SF	Sequential Fixing

SINR	Signal to Interference and Noise Ratio
SNR	Signal to Noise Ratio
TDMA	Time Division Multiple Access
UB	Upper Bound
VWFR	Variable Width Fixed Rate
VWVR	Variable Width Variable Rate
WF	Water-filling
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
WMN	Wireless Mesh Network

Chapter 1

Introduction

1.1 Overview

Currently, wireless broadband is growing at unprecedented rate and broadband access is considered to be the great infrastructure challenge of the early 21st century [1]. Key drivers for this growth include the maturation of third-generation (3G) wireless network services, development of smart phones and other mobile computing devices, the emergence of broad new classes of connected devices and the roll out of 4G wireless technologies such as Long Term Evolution (LTE) and WiMAX [2]. Major wireless operators in the US (e.g., AT&T and Verizon) have recently reported substantial growth in data traffic in their networks, which is driven in part by smart-phones (e.g., iPhone) usage. According to Cisco, wireless networks in North America carried approximately 17 petabytes per month in 2009, and it is projected that in 2014 they will carry around 740 petabytes, a 40-fold increase. This traffic growth is due to the increased adoption of Internet-connected mobile computing devices and increased data consumption per device [1]. Furthermore, a surge of machine-based wireless broadband communications is forecasted for the next few coming years, as more smart devices (e.g., electric vehicles, body sensors, wireless enabled cameras, smart

meters, etc.) take advantage of the ubiquitous wireless connectivity. The aggregate impact of these devices on demand for wireless broadband could be enormous [1].

Currently, improving both the wireless capacity and coverage has been the limiting factor for un-leaching the broadband capabilities; this has been a daunting task, despite the recent progress in wireless communication techniques, such as adaptive modulation and coding, multi-user decoding, MIMO and OFDM. Further, there have also been substantial advances in technical solutions for mitigating the effects of interference and fading, thus improving the wireless performance and reliability. For example, relay-based cooperative techniques try to mitigate detrimental propagation conditions by allowing communication to take place through a third party device acting as relay [3]. Recently, cognitive radio technology has shown to substantially increase the spectrum efficiency through dynamic spectrum assignment by allowing secondary (unlicensed) users to identify and exploit local and instantaneous spectrum availability in a non-intrusive manner. Here, the objective is to provide sufficient benefits to the secondary users while protecting spectrum licensees (i.e., licensed users) from interference [4], which severely restricts the reusability of the spectrum in space. Managing interference for increased spectrum spatial reuse, and thus higher throughput performance and enhanced reliability, has turned out to be a challenging task that designers for wireless communication systems are facing. Standard approaches for dealing with interference necessitate the use of power control techniques at the physical layer, proper link scheduling and activation, effective routing and transmission rate control. Thus, interaction among two or more layers of the network protocol stack becomes a key for achieving fair resource allocation and increased network utility performance; such framework is more commonly known as the cross-layer optimization design problem [5, 6] and fundamental limits on the impacts of layer-crossing on network performance have been obtained.

1.2 Problem Statements and Motivations

1.2.1 Cross-layer Design and Optimization

Although the layered architectures have served well for wired networks, they are not very suitable for wireless networks. Multi-hop wireless networks and data packet transmissions opened numerous transmission possibilities, and motivated to break the barriers imposed by layered transmissions [7]. Unlike layered designing where the protocols at different layers are designed independently, cross-layer design considers the dependencies between protocol layers to improve the performance of wireless networks [8].

Local adaptation of physical layer resources such as transmit power, coding rate, modulation etc. to achieve a target bit error rate (BER), restrains both routing and MAC (Medium Access Control) decisions by altering the topology graph, feasible transmission schedules and payload transmission rates [5]. Interference between concurrently transmitted links can be controlled by proper link scheduling and channel allocation; the MAC layer is responsible for that. The interference accumulated from simultaneous transmissions directly affects the physical layer performance in terms of successfully separating the desired signal from other unwanted ones. Moreover, high packet delay and/or low bandwidth might be the result of transmission scheduling, forcing the network layer to change its routing decisions [9]. On the other hand, the network layer decides on the routing path and different routing decisions alter the set of links to be scheduled and therefore will influence MAC layer performance. Additionally depending upon the routing path physical layer resources need to be allocated properly to achieve the target BER [9]. Hence, the network performance of a wireless network can be improved drastically if a cross-layer design approach is adopted to coordinate the network layer routing, MAC layer scheduling and controlling physical

properties such as, transmit power and rate. Note that, considering attributes from other layers (i.e., congestion control and queue management of transport layer) can further improve the network performance.

To obtain optimal cross-layer design, a joint optimization of physical, MAC and network layer attributes is needed. Note that, solving a single joint optimization problem considering all these attributes is rather complex. In this thesis, we perform cross-layer design for wireless networks by jointly considering routing, link scheduling, variable bandwidth allocation and rate control. We decompose this combinatorial complex optimization problem using column generation technique to solve it optimally.

1.2.2 Cross-layer Design with Flexible Spectrum Access

Wireless Mesh Networks (WMNs) have recently emerged as a solution for providing last-mile broadband Internet access [10]. A WMN typically consists of a number of stationary wireless routers interconnected by wireless links and provides a backbone over which end users can access the Internet. As more users depend on WMNs for their primary source of Internet access, there is an increasing expectation that these networks should provide both reliable and high end-to-end throughput. The end-to-end throughput of a multi-hop wireless network, however, is often limited by interference caused by concurrent neighboring transmissions and intra-path interference caused by transmissions on successive hops along a single path. Therefore, it is important to control interference while maintaining high concurrency to achieve higher aggregate throughput, which is a key design objective for any wireless system. While the intra-path interference problem can potentially be eliminated by equipping each node with multiple radios and assigning different channels to links along a path, the interference among simultaneous transmissions may significantly be reduced by

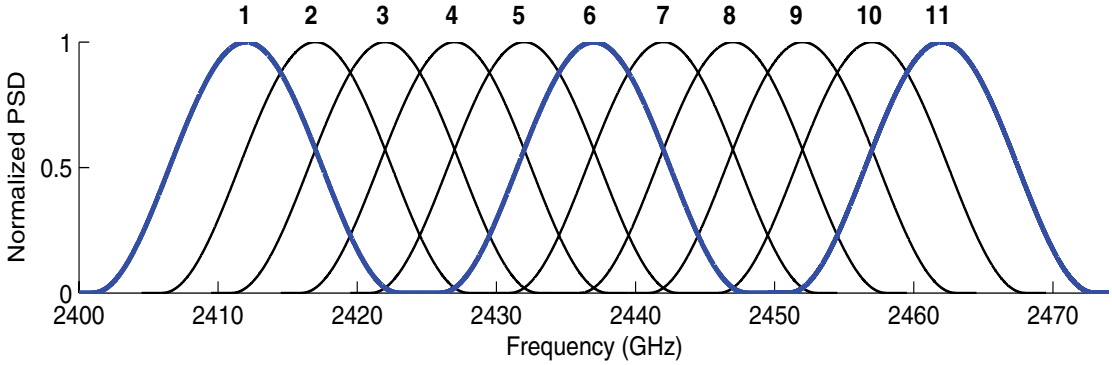


Figure 1.1: Channel bandwidth allocation in 802.11b/g; eleven partially overlapping channels, while channel 1, 6 and 11 are orthogonal

using orthogonal channels on adjacent links. This kind of channelization has already been deployed to improve the capacity of 802.11-based wireless LANs where neighboring access points (APs) are assigned different channels of fixed widths (each being 20 MHz wide) [11, 12]. This use of preset channel widths is the direct result of how the available spectrum is divided by existing wireless technologies. For example, in 802.11b/g, the entire available spectrum is divided into 11 overlapping (3 of which are orthogonal) channels, separated by 5 MHz, and of 20 MHz width each (Figure 1.1) [13]. In WiMAX networks, however, the spectrum block is divided into channels of different, but predetermined, widths. Further, the 802.11n introduces channel bonding, which allows users to form higher width channels (e.g., 40 MHz) from 20 MHz channels to achieve better performance gains.

Recent studies have shown that if the width of the spectrum band allocated for the available channels is configured dynamically, then higher capacity can be obtained over a preset width channel allocation [14–16]. In their original work [14], the authors have argued that while today WiFi nodes dynamically change many variables (e.g., power, transmission rate, channel center frequency, etc.) to improve their communication, the channel-width has been largely overlooked. They also showed, using

commodity hardware, that the channel-width may be adapted dynamically through some software modifications with very little overhead. They showed that such adaptation brings unique benefits and improves single link's throughput and energy efficiency. The authors concluded that WiFi networks should adapt the width of the communication channels based on their current needs and environmental conditions. The authors of [16] showed that variable-width channels provide significant theoretical capacity improvements and demonstrated a spectrum allocation algorithm, which assigns variable-width orthogonal channels, to improve the throughput of multiple interfering transmitters. The authors of [15] leveraged the findings in [14] and formulated the problem of variable-width channel allocation in infrastructure-based wireless LANs; they showed that by allocating more spectrum to highly loaded access points (APs), the overall spectrum utilization can be substantially improved.

Adaptation of variable bandwidth channels rather than fixed bandwidth provides wireless networks with some unique benefits to strike balance between interference control, maintaining higher concurrent transmissions and better spectrum reuse. While dividing the available spectrum among smaller bandwidth channels result in more orthogonal channels which allow for more concurrent non-interfering communications in the same area, each such communication has larger transmission range and smaller link capacity. On the other hand, a channel with wider spectrum band increases the link capacity and reduces the transmission range, but results in smaller number of orthogonal channels and stronger effect of intra-path interference. Given these conflicting objectives, it becomes clear that a variable spectrum width allocation (rather than fixed) may strike a good balance between interference control and maintaining both higher concurrency and better spectrum reuse for wireless networks.

Modern radios, e.g., software defined and cognitive radios, are frequency agile and

have recently received a lot of attention due to their ability of enabling very flexible spectrum access through their spectrum sensing capability and ability to dynamically reconfigure the allocated spectrum [17–20]. Frequency agile radios partition the spectrum into several sub-channels (e.g., OFDMA sub-carriers which are in turn grouped into sub-channels) of equal size and access the medium either through a block of contiguous number of sub-channels (1-agile radio) or through a set of non contiguous sub-channels which need not necessarily be frequency aligned [17]. This latter form of agile radio requires more sophisticated signal processing and hence increased hardware complexity. The former one, however, may be implemented through commodity WiFi hardware [14].

1.2.3 Resource Allocation in Cooperative Cellular Networks

Spatial diversity achieved from cooperative transmissions using relay nodes has shown great potential for combatting channel fading and enhancing the performance of wireless networks [3]. Indeed, efficient resource allocation plays a vital role in the performance of any wireless networks, and there has been a substantial amount of previous work done on this particular topic. However, most of the works done solve the joint resource allocation problem sub-optimally using heuristics or dividing the joint problem into multiple subproblems, which does not provide performance guarantees in terms of optimality.

Recent advances in broadband wireless and cellular access are attracting emerging multimedia applications, such as Internet Protocol television (IPTV) over WiMax [21] and multimedia broadcast/multicast services (MBMS) within 3GPP [22]. However, only little work has been done on cooperative multicast/broadcast over wireless networks [23].

In different cooperative communications schemes, relays may receive data from

a source, amplify it, and then forward it to a destination (amplify and forward) or they may decode a transmitted codeword, re-encode it and forward the re-encoded codeword to a destination (decode and forward). The spatial diversity achieved by this kind of transmission is based on the fact that a single transmission is received by the destination from multiple separate transmitters, one from the source and others from relays. This technique has proven to be a powerful physical layer technique to combat fading and increase the physical layer capacity in wireless relay networks.

Researchers have addressed the problem of resource allocation (relay, power, subcarrier) in relay-aided cooperative transmission wherein the authors have assumed that each user may get assistance from multiple relays [24–26]. Nonetheless, orthogonal transmissions from all relays are bandwidth inefficient. As an alternative, relays may use distributed space time codes (DSTC) [27], which is more bandwidth efficient and achieves full diversity gain. However, it requires symbol level synchronization [24]. As multiple relays transmit simultaneously to the destination, the propagation time of the signals from each relays to the destination is different. It is quite difficult to correct this potentially varying timing mismatch [28].

It has recently been shown and investigated in many different contexts [29–31] that the maximum benefit of cooperative diversity can be achieved with minimum overhead if a single best relay can be chosen for a particular source-destination (s-d) pair. Selecting a single relay limits the number of bandwidth channels as well as eliminate the need for synchronization. In the case of a single s-d pair, choosing only one (the best) relay is quite straightforward. However, in multiple data flow scenarios, the selection gets considerably more complicated and becomes a complex combinatorial problem. Moreover, in a multiple data flow scenario, a relay can assist more than one flow, thus the transmission power of the relay needs to be effectively shared between traversing flows. Since power is a valuable network commodity, the

relay power needs to be divided optimally to ensure a judicious use of this limited resource.

1.3 Thesis Contributions

- We address, in the context of wireless network design, the problem of optimally partitioning the spectrum into a set of non overlapping (orthogonal) channels with non uniform spectrum widths. While narrower bands split the total available spectrum into more non-overlapping channels allowing more parallel concurrent transmissions, wider bands yield links with larger transport capacity. Thus, we model the combinatorially complex (NP-complete) problem of joint routing, link scheduling, and variable-width channel allocation in both single and multi-rate multi-hop wireless networks as a mixed integer linear program (MILP), and present a solution framework using the column generation decomposition approach, where the problem is divided into a master problem and a pricing subproblem. Given the nature and complexity of the resulting pricing subproblem, we propose a greedy method for partitioning the spectrum and reduce the size of the subproblem, and hence obtain solutions for larger network instances. We present several numerical results and engineering insights suggesting both spectrum width and transmission rates as effective tunable knobs for combatting interference and promoting spatial reuse and thus achieving superior performance in multi-hop settings.

- Next, we investigate the problem of flexible spectrum access in multi-hop wireless networks with software defined radios. We assume radios that are capable of transmitting on channels of contiguous frequency bands and which do not

require any sophisticated processing. Because these radios can flexibly configure their transmissions anywhere in the available frequency band, the spectrum becomes vulnerable to fragmentation and interference. In this work, unlike the previous one, we do not impose optimal partitioning of the available spectrum band into a set of non-overlapping channels rather we let the cross-layer design decide on the channel bandwidth positions. In this way a more flexible allocation of bandwidth is possible, since the transmissions may use overlapping channels. When considering spectrum overlap, the design problem gets further complicated, and adjacent channel interference must be dealt with properly. The Interference factor (I-factor) captures the amount of overlap between a transmitting and an overlapping interfering channel. This I-factor may not be predefined, but rather it is jointly determined when performing channel assignment as it depends on the portion of overlap between two channels. We consider the joint problem of routing, link scheduling and spectrum allocation where scheduling feasibility is considered under the physical interference (SINR) constraint. We again present a column generation based decomposition for this complex optimization problem. We show that obtaining the optimal solution is computationally not feasible, except for very small networks. We thus adopt a two-fold method to circumvent the complexity while yielding practical solutions. First, we relax the SINR constraint and use a simplified graph-based model for the interference. Second, we use a simulated annealing (SA) approach to solve the pricing subproblem. Our SA approach however is augmented with a feasibility check so that only SINR-feasible schedules are passed back to the master problem. Results confirm that the column generation method using SA substantially reduces the computation time and achieves near optimal solutions. Our results also revealed that substantial improvement in network performance

is obtained with flexible spectrum assignment which results from its capability of better managing the interference in the network.

- Finally, we investigate the joint problem of optimal relay selection and power allocation in amplify-and-forward relay aided cooperative cellular wireless networks considering both unicast and multicast traffic. We first present mixed Boolean-convex optimization models for both unicast and multicast traffic scenarios to maximize the overall network capacity and solve these combinatorial problems optimally using the branch and bound technique. We then show that obtaining the optimal solution is computationally infeasible for large network sizes. To remedy this, for unicast, we present an efficient water-filling based technique to obtain near optimal solutions. We show that, unlike unicast traffic, water-filling does not yield near optimal solutions in multicast scenarios. We thus adopt an algorithm based on sequential fixing for the multicast case which substantially reduces the computation time and achieves near optimal solutions. Furthermore, in both unicast and multicast scenarios, we assume that the power levels are drawn from a continuous range. To make the proposed methods more practical, we also consider scenarios when the number of power levels is finite (i.e., discrete). We present optimal and sub-optimal methods for solving this optimization problem and compare their performance with previous methods.

1.4 Thesis Outline

The rest of the thesis is organized as follows. Chapter 2 presents the background and reviews the related work in the fields investigated throughout this thesis. Chapter 3 discuss optimization theories and methodologies which we use to implement

and solve the problems discussed in this thesis. In Chapter 4, we present cross-layer design for wireless mesh networks considering optimal partitioning of the spectrum into a set of non overlapping channels. Chapter 5 discusses the flexible spectrum assignment problem considering overlapping channels. We present the joint problem of optimal relay selection and power allocation in cooperative cellular wireless networks in Chapter 6. Chapter 7 summarizes our conclusions and presents some future research directions.

Chapter 2

Background and Related Work

In this chapter, we present the background and the literature survey for the topics investigated throughout this thesis.

2.1 Wireless Mesh Networks

Wireless Mesh Networks (WMNs) have emerged recently as an attractive option for increasing broadband penetration and providing inexpensive and reliable last-mile internet access [10]. These networks operate at the edge of the internet and consist of stationary wireless mesh routers interconnected through wireless links, which provide a backbone over which end users can access the internet. WMNs are different from traditional multi-hop wireless networks; they are expected to employ advanced communication technologies (e.g., adaptive modulation and coding, MIMO and OFDM) for enhancing the network throughput. Furthermore, WMNs are expected to be tightly coupled with the wired network and, to be competitive with other wired technologies, they must provide Quality of Service (QoS) support.

Fig. 2.1 shows a typical wireless mesh network deployed in a residential area. The

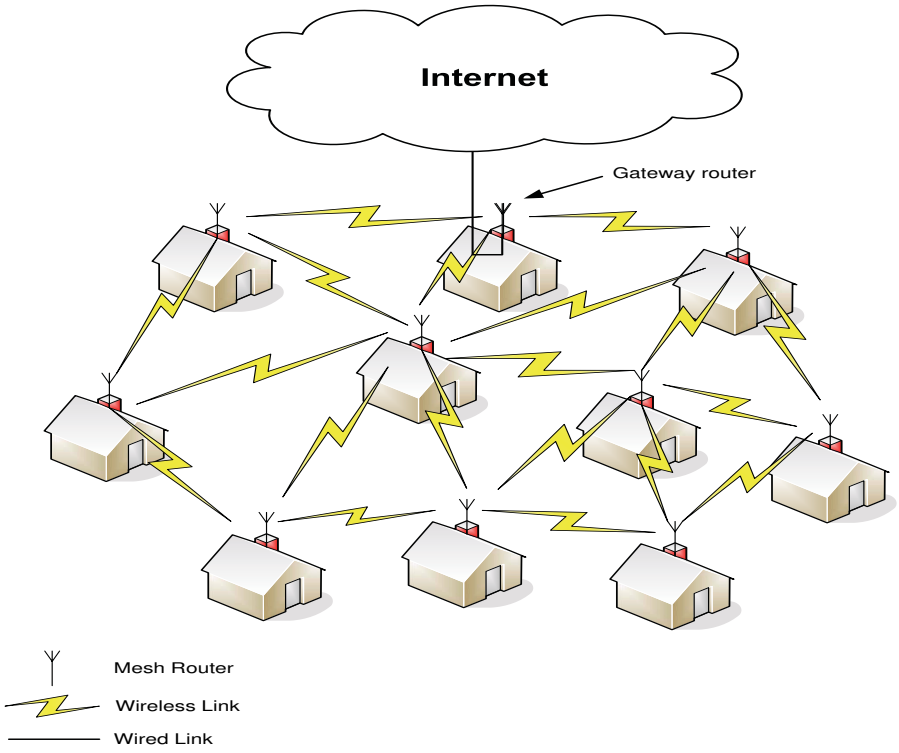


Figure 2.1: Typical wireless mesh networks in residential area

routers are placed on rooftops of the houses. Each of these mesh routers communicate with others and also with gateway via single-hop or multi-hop wireless links, depending upon the distance between them. Mesh clients or the home users usually use separate radio interface or wired Ethernet to connect themselves to the mesh routers.

2.1.1 Features

Since the infrastructure needed for WMNs are in form of small radio relaying devices and can be easily placed on rooftops of houses (Fig. 2.1), the investment needed for WMNs is much less than other networks (e.g., cellular networks). Moreover, network devices such as, mesh routers are also cheap and widely available and their price continue to decrease [32]. The built-in robustness of WMNs makes the network

maintenance much easier. Since the mesh topology of these networks provides many alternate routes for a particular traffic session, if an existing path fails due to router malfunction, quick reconfiguration of an alternate route can be done. Failures due to wire cut, which is widely experienced in other networks, is also not a possibility for these networks. In addition to lower deployment and maintenance costs, since there are only few routers which work as gateway to existing backbone networks, only a few wired internet subscriptions are shared among clients in a larger community; hence, subscription cost will also be lower [32].

2.1.2 Performance Challenges

Although WMNs exhibit nice and attractive features, they still lag in performance. These networks operate over unlicensed band and the achievable end-to-end throughput often is limited by external, internal and self interference. As the deployment and use of WMNs are increasing significantly (with many cities have planned and/or deployed WMNs [33–36]), more users depend on WMNs for their primary source of internet access and therefore, there is an increasing expectation that these networks should provide both reliable and high end-to-end throughput. Such high performance is a necessity for any attractive real-time multi-media applications (internet telephony, voice conference, IPTV etc.) envisioned to use the services of the WMNs.

2.1.3 Resource Allocation

To achieve a high aggregate throughput it is important to control interference while maintaining high concurrency. Several approaches are currently in use for managing interference. For example, MAC protocols which coordinate the access to the medium (e.g., CSMA) can reduce or eliminate collisions among concurrent transmissions. Other methods such as interference cancelation and interference alignment [37]

have great potential for mitigating the problems caused by interference; they however involve significant computational complexity and cannot be implemented with commodity hardware [16].

On the other hand, the performance of such networks can substantially be improved by controlling the interference if nodes are equipped with multi-radio capabilities and transmission links are assigned with orthogonal channels. Such multi-radio multichannel networks have received a lot of attention over the past few years and a large number of papers have been published on routing [38], joint routing and channel allocation [39, 40], joint routing and link scheduling [41], resource allocation [42], asymptotic capacity bounds [43, 44], topology control [45, 46], cross-layer design for rate allocation [47], among others. Partial overlapping channels have also been considered for WMNs to further improve their performance [48]. However, all these works considered the fixed channelization structure of existing wireless technology.

Recently, however, it has been shown that the channel width can be adapted dynamically purely in software [14]. The authors have shown that such adaptation brings unique benefits in improving the single's link throughput and energy efficiency. The authors of [15] have leveraged on this capability to improve the performance of infrastructure wireless networks and they proposed that access points (APs) adaptively adjust both the center frequency and spectrum width to match the traffic load.

In the context of WLANs, the authors of [12] presented a traffic-aware channel allocation where the observed traffic demand at APs is incorporated into the assignment process. This work [12] is motivated by recent studies [49], which showed that the traffic volume in enterprise WLAN deployments vary significantly both across APs and time. Similarly, the authors of [50] noted that WiFi APs must adjust their allocated bandwidth based on varying traffic demands in order to improve user

experience.

The authors of [18] have capitalized on the frequency agility of modern radios (e.g., cognitive radios (CR)), which can configure the center frequency and spectrum bands of their channels, and they proposed a cross-layer design approach for the problem of joint transport, routing, and spectrum sharing. They also proposed a distributed two-phase method for solving the complex optimization model, where flow routing and spectrum allocation are carried out in one phase and link scheduling is performed in another phase.

Cognitive radios [51] are capable of continuously sensing the spectrum and opportunistically utilizing blocks of spectrum unused by the primary users. Such blocks are referred to as white space and the authors of [52] have investigated the problem of spectrum allocation in CR networks and showed that it is more challenging (NP-hard) in these networks than networks with preset channel widths. Channel access in CR networks with joint power and rate control is discussed in [19] where the authors assumed that CR links may be assigned multiple non-contiguous channels of equal widths. Given the complexity of the problem, the authors presented an approximation algorithm using the rounding off method. Spectrum sharing in multi-hop networks with software defined radios is studied in [53]; the problem of routing and scheduling (in frequency domain) is modeled, using the protocol interference model, as a mixed integer non-linear program (MINLP) and the authors obtained a lower bound solution and proposed a method for approximating the near optimal solution. The authors assumed that radios can be configured to transmit on any band whose width is not fixed, and each band may be divided into sub-bands for optimal spectrum sharing.

In [54], the authors considered the resource optimization in OFDMA-based multi-hop wireless network with power and rate control. In OFDMA access systems, the

spectrum is divided into multiple preset-width sub-channels and therefore the problem of spectrum assignment reduces to allocating sub-channels to active links. The authors presented a greedy heuristic to obtain solutions in reasonable time.

In [55], the concept of spectrum partitioning is introduced; given a spectrum of total width W(Hz), the authors determined the optimal value of the number of orthogonal channels N, each of width W/N in order to maximize the number of transmission links (of fixed transmission data rate Rbps) in the network. The authors noted that while a larger N results in more orthogonal channels (hence better interference control), it however increases the SINR requirement since the data rate R must now be achieved over less bandwidth. They therefore studied this tradeoff in their paper.

2.2 Cooperative Wireless Communications

In [56], Sendonaris et al. first proposed the idea of user cooperation wherein mobile users cooperate by relaying each others data, thus exploiting the spatial diversity in a cellular network. Considering user pairs, where each user and the assigned partner receive, detect and retransmit each others data, the authors of [56] have presented information theoretic analysis and showed an increase in the capacity region with cooperation for a two user case. This work sparked further studies in this area and several cooperation schemes have, since then, been proposed and studied in the literature [28].

Cooperative communication can be best explained by a three node structure presented in Fig. 2.2 by Sharma et al. [57]. In cooperative communications, the transmission is usually done over two time slots (a frame). In the first time slot a source node s communicates with the destination d and due to the broadcast nature of the wireless medium the relay node r overhears the transmission and start processing

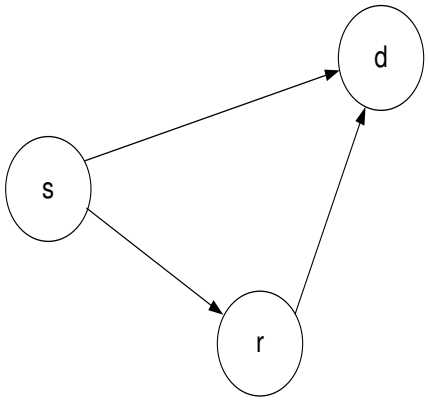


Figure 2.2: Three-node structure for cooperative wireless communications

the signal. In the second time slot the relay node r forwards the processed signal to the destination. The spatial diversity achieved by this kind of communication is based on the fact that a single transmission is received by the destination from two spatially separated transmitters a source and a relay. The two time slot structure of cooperative communications is the result of half-duplex nature of most wireless transceivers [58].

Depending upon the signal processing activity relay nodes in cooperative wireless communications may operate in different modes. Amplify-and-forward (AF) and Decode-and-forward (DF) are the two most common relaying strategies [3].

2.2.1 Amplify-and-forward (AF) Relaying

In AF relaying, when a relay node overhears the transmission from a source, it amplifies the signal and forwards it to the destination. In [57], the authors have shown that the single link cooperative capacity, between a source s and a destination d served by a relay node r with AF capabilities, can be written as follows:

$$C_{AF}(s,r,d) = \frac{W}{2}\log_2(1 + SNR_{sd} + \frac{SNR_{sr} \cdot SNR_{rd}}{SNR_{sr} + SNR_{rd} + 1}) \tag{2.1}$$

where, W is the channel bandwidth and SNR_{sd}, SNR_{sr}, SNR_{rd} are the signal to noise ratio at the destination and relay nodes. The multiplicative term $\frac{1}{2}$ is from the fact that the cooperative communication is done over two time slots.

2.2.2 Decode-and-forward (DF) Relaying

In DF relaying, when a relay node overhears the transmission from a source, it decodes the transmitted signal, re-encodes it and then forward the re-encoded signal to the destination. The achievable DF capacity under two time slot structure is shown in [3] as the following:

$$C_{DF}(s,r,d) = \frac{W}{2}\min\ \{\log_2(1 + SNR_{sr}), \log_2(1 + SNR_{sd} + SNR_{rd})\} \tag{2.2}$$

When relay aided cooperative communications is not considered the source can directly communicate with the destination over both the time slots an the achievable capacity can be written as Shannon Capacity:

$$C_D(s,d) = W\log_2(1 + SNR_{sd}) \tag{2.3}$$

In this thesis, we consider only AF relaying. In AF relay mode a relay only retransmits a scaled version of their received signals from the source node according to their power constraint [31]. Therefore, AF relay employed in this thesis is a reasonable strategy when relay nodes have limited power. Moreover, the complexity pertained to AF relaying is much simpler, since it does not require any signal processing at the

relay node for decoding and encoding process. Although we have only presented our work considering AF relaying, the techniques proposed in the thesis can also be used for DF relaying.

2.2.3 Relay Selection

Traditionally in a network with multiple relays, each destination gets assistance from all relays and the communication between them are done using orthogonal channels (orthogonal frequency division multiple access) [24–26]. However, this type of orthogonal transmissions from all the relays is not bandwidth efficient. A more practical alternative is that the relays may use distributed space time codes (DSTC) [27], which is more bandwidth efficient and achieves full diversity gain. However, it requires symbol level synchronization [24]. As multiple relays transmit simultaneously to the destination, the propagation time of the signals from each relays to the destination is different. It is quite difficult to correct this potentially varying timing mismatch [28]. However, it has recently been shown that most of the benefits of cooperative diversity can be achieved with minimum overhead if a 'single best' relay is selected to cooperate with a source-destination pair. This scheme is referred to as selection cooperation [29, 30]. The scheme has also been investigated in various contexts [31, 59–61]. In [30], the authors have shown that the selection cooperation achieves the same diversity order as DSTC and also provides a significant power gain over DSTC.

2.2.4 Resource Allocation

Resource allocation in cooperative relay networks has been an extremely active research area. Numerous works on relay selection, power allocation and sub-carrier

allocation have already been published. These network resources are optimized individually as well as jointly to improve the network performance. Although there have been many works done in this area, most of the joint optimizations are solved using heuristics or by dividing the joint problems into multiple sub-problems which do not confirm performance guarantees in terms of optimality .

Relay selection in the case of single s-d pair is quite straight forward and has been solved for both amplify and forward (AF) [31] and decode and forward (DF) [29,30]. In [29–31], the authors have selected the single best relay to cooperate with the s-d pair and showed that the selection maintains the full diversity order. An optimal relay selection algorithm for both AF and DF relaying has been proposed in [58] where the authors considered multiple s-d pairs. Then, they extended their work for multi-hop scenarios [62] where the authors considered selection cooperation in each hop jointly with flow routing. The authors in [63] considered relay assignment for cooperative networks comprising multiple source-destination pairs and multiple relays. They proposed assignment algorithms that achieve the maximum spatial diversity by all nodes, thus leading to fairness among the nodes. In [64], the authors considered the same system model considered in [63], but with two-way relaying. The relay nodes in this case use binary network coding and employ AF relaying, and threshold-based DF relaying.

A joint relay selection and power allocation considering AF relaying for both single and multiple s-d pairs is performed in [65]. In that paper, the authors have proposed a semi-distributed heuristic with no performance guarantee to the optimality of the achieved results. The authors of [66] have proposed a centralized solution for the same problem and proposed a suboptimal solution technique based on a rounding scheme. In [59], the authors proposed a water-filling based solution for the same problem considering DF relaying. They have also provided an upper bound solution

and showed that the water-filling solutions closely follow the upper bound solutions.

The authors of [67, 68] have used the concept of time-sharing condition, proposed in [69] and solved the joint problem of relay selection, power and subcarrier allocation. To employ the time-sharing condition both works assumed infinite number of channels (or subcarriers) which makes the duality gap of the non-convex problem zero. However, in practical scenarios, this assumption is very unlikely to hold. Moreover, the complexity of the proposed solution is not polynomial. There are also some recent works [70, 71] which solved the same problem sub-optimally by dividing the problem into multiple subproblems.

In [72], the authors proposed fully distributed cross-layer frameworks for multihop cooperative networks. They first solved the joint problem of routing, relay selection and power allocation to minimize the network power consumption and then they extend the framework by incorporating congestion control to optimize the sum rate utility and power tradeoff.

Although the authors of [73–75] have shown that energy accumulation in broadcast/multicast relaying leads to significant energy savings over conventional relaying, there has been little work done in terms of resource allocation considering broadcast/multicast traffic. Recently, the authors of [23] have analyzed the performance of cooperative networks in multicast scenarios by individually optimizing the power and relay location.

Chapter 3

Optimization Methods

Many problems of practical interest in communications and signal processing can be formulated as constrained optimization problems of the following standard form:

$$\begin{aligned} &\text{Minimize} && f_0(x) \\ &\text{Subject to,} && f_i(x) \leq 0 \qquad i = 1, 2, ..., m \end{aligned} \tag{3.1}$$

Here, f_0 is the objective function and f_i $(i = 1, 2, ..., m)$ are the constraint functions. The vector $x = (x_1, x_2, ..., x_n)$ is the optimization variable of the problem. The optimal solution of the problem (3.1) defined by $f_0(x^\star)$ is the smallest value achievable while a vector $x^\star$ satisfies all the constraints. $x^\star$ is the optimal solution for the variables. This implies that no other vector (say y) with $f_i(y) \leq 0$ will have $f_0(y) < f_0(x^\star)$.

By the characteristics and forms of objective function and constraints, optimization problem can be divided into many different classes and families. In this chapter we will discuss some of them which we have implemented and solved throughout this thesis.

3.1 Linear Programming Problem

If the objective function f_0 and all the constrains f_i $(i = 1, 2, ..., m)$ are linear, that is, they satisfy the following equations

$$f_i(\alpha x + \beta y) = \alpha f_i(x) + \beta f_i(y) \quad i = 0, 1, ..., m, \quad \forall x, y \in \mathbf{R}^n, \ \alpha, \beta \in \mathbf{R}, \tag{3.2}$$

the problem is called Linear Programming Problem or Linear Program (LP) [76]. If the objective or any of the constraints are not linear then the problem is called Non-linear Program (NLP).

There are number of effective methods for solving LP. The Simplex algorithm [77] is the most widely used LP solution method implemented in many commercial optimization softwares (such as CPLEX). Given a polyhedron representing the solution space and a real-valued linear function representing the objective function defined on this polyhedron, the Simplex method finds a vertex on the polyhedron where this function has the smallest (or largest, depending upon the objective of the problem) value [78]. The Simplex algorithm is quite efficient and is guaranteed to find the global optimum of the LP optimization problem [79].

3.2 Integer Linear Programming Problem

An Integer Linear Programming (ILP) model is the same as LP model with addition restriction that variables must have integer values. If only some of the variables are required to have integer values the problem is referred to as Mixed Integer Linear

Programming (MILP). The Problem takes the following standard form:

$$\begin{aligned} &\text{Minimize} && f_0(x) \\ &\text{Subject to,} && f_i(x) \leq 0 \qquad i = 1, 2, ..., m \\ & && x \quad \text{is integer} \end{aligned} \tag{3.3}$$

In contrast to LP problems, which can be solved efficiently within a limited number of iterations, ILP problems are NP-hard [80]. Indeed, LP solution gives a bound (upper-bound if maximization problem and lower-bound solution if minimization problem) on the optimal value of the ILP, and may give the optimal solution to ILP if the LP solution is integer. ILP solution methods require formulation whose LP relaxation gives a good approximation to the hull of integer feasible solutions. Different methods have been proposed to close the gap between LP and ILP, and to boost the performance of the ILP methods [81,82]. Some of the very common solution methods for ILP are: Branch and Bound techniques, Cutting Plane techniques, Branch and Cut techniques etc. In the next section, we will discuss Branch and Bound (B&B) techniques as we have used this throughout this thesis to solve integer programming problems both with linear and non-linear objectives and constraint functions.

3.3 Branch and Bound Techniques

Branch and Bound technique is a non-heuristic method. It searches for a globally optimal solution of integer programming problems with predefined precision of optimality [83]. This approach maintains an upper and lower bound on the optimal objective value during the optimization process; and terminates with a certificate proving that the suboptimal point found is ϵ-optimal. If globally optimal solution is defined as $f_0(x^\star)$, and ϵ has a value between 0 and 1, then the B&B algorithm

searches for ϵ-optimal solution $f_0(x)$, which satisfies $f_0(x) \leq \epsilon f_0(x^\star)$.

B&B algorithms are basically based on a divide and conquer technique. The branching is performed by dividing the feasible solution space into smaller subsets. The bounding is performed by discarding the current subset if its bound indicates that it cannot contain an optimal solution for the original problem. The efficiency of the B&B technique depends critically on the branching and bounding strategies. Some branching choices could lead to repeated branching, without any pruning of the B&B tree until the full exploration of the sub solution space. In such a case the branching method would become an exhaustive enumeration of the solution domain, which is often huge.

For more elaborate explanation on B&B algorithm the readers may refer to Stephen Boyd and Jacob Mattingley's lecture notes on Branch and Bound Method [83]. Note that, efficient Branch and Bound algorithms are already successfully implemented in many commercial optimization softwares such as: CPLEX, TOMLAB etc.

3.4 Convex Optimization

We now focus on the following standard form of optimization problem:

$$\begin{aligned} \text{Minimize} \quad & f_0(x) \\ \text{Subject to,} \quad & f_i(x) \leq 0 \qquad i = 1, 2, ..., m \\ & h_j(x) = 0 \qquad j = 1, 2, ..., p \end{aligned} \tag{3.4}$$

This optimization problem is said to be a convex optimization problem if the objective function $f_0(x)$ and all inequality constraint functions $f_i(x)$ $(i = 1, 2, ..., m)$ are convex

that is,

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y) \quad i = 0, 1, ..., m, \quad \forall x, y \in \mathbf{R}^n, \quad \alpha, \beta \in \mathbf{R}, \tag{3.5}$$

and also all equality constrains $h_j(x) \quad (j = 1, 2, ..., p)$ are affine, i.e., $h_j(x)$ can be written as: $h_j(x) = a_j^T x + b_j$ for some a_j and b_j.

Note that, the conditions above imply that the corresponding feasible set $\mathcal{S}$ associated with any convex optimization problem is a convex set as it is the intersection of the convex domain $\mathcal{D} = \cap_{i=0}^{m} \mathrm{dom} f_i$, with the m convex sub-level sets $\{x| \quad f_i(x) \leq 0\}$ and the p hyperplanes $\{x| \quad a_j^T x + b_j = 0\}$. A convex optimization problem consists therefore of minimizing a convex objective function over a convex set [76].

It is clear from the above discussions that linear programming problem (LP) is a spacial case of general convex optimization problem. Although LP is a special case of convex optimization problem, convex optimization is not limited to LP but it also covers non-linear programs (NLP) with the properties discussed above. In our thesis when we mention convex optimization problem we mean non-linear convex optimization problem. If it is linear we simply mention it as LP.

A convex optimization problem can be solved to optimality if something called strong duality holds. In the next two sections we will briefly discuss on duality theory, weak and strong duality, Slater's condition and Karush-Kuhn-Tucker(KKT) Conditions which will help the readers to follow the convex optimization problem formulations done in Chapter 6.

3.5 Duality Theory

The standard form problem (3.4) presented in the previous section is called the primal optimization problem and the optimization variable x is the primal variable. Let us

assume the domain $\mathcal{D}$ of the problem (3.4) is nonempty and the optimal value of the primal problem is referred to as $P^\star = f_0(x^\star)$. The primal problem can be any generic optimization problem which is not necessarily a convex optimization problem. The basic idea for Lagrangian duality is to take the constraints of the primal problem into account by augmenting the objective function with a weighted sum of the constraint functions. The Lagrangian defined with $\mathcal{L}$ associated with the problem (3.4) can be written as the following:

$$\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_i h_i(x) \tag{3.6}$$

here, λ and ν are the Lagrange multipliers, where, λ and ν are associated with inequality and equality constraints respectively.

The Lagrange dual function referred to as $g(\lambda, \nu)$ is the minimum value of the Lagrangian over x and can be written as:

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} (f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_i h_i(x)) \tag{3.7}$$

The function $g(\lambda, \nu)$ is a concave function as it is point-wise infimum of a family of affine functions of (λ, ν) even when the primal problem (3.4) is not convex [76].

Let us now assume that, $\tilde{x}$ is a feasible point for the problem (3.4). When $\lambda \geq 0$ and from (3.4), $f_i(\tilde{x}) \leq 0$, and $h_j(\tilde{x}) = 0$, we can write,

$$\sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{j=1}^{p} \nu_i h_i(\tilde{x}) \leq 0 \tag{3.8}$$

Then, adding $f_0(\tilde{x})$ in the both sides of Eq. (3.8) we get the following expression:

$$\mathcal{L}(\tilde{x}, \lambda, \nu) = \sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{j=1}^{p} \nu_i h_i(\tilde{x}) + f_0(\tilde{x}) \leq f_0(\tilde{x}) \tag{3.9}$$

Hence, we can write,

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda, \nu) \leq \mathcal{L}(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x}) \tag{3.10}$$

Since, Eq. (3.10) holds for every feasible point $\tilde{x}$, we can say that the Lagrangian dual function provides lower bound on the optimal value of the primal problem $P^\star$ for any feasible x and $\lambda \geq 0$.

$$g(\lambda, \nu) \leq P^\star \tag{3.11}$$

As the Lagrange dual function provides us with the lower bound on the optimal value of the primal problem $P^\star$, the natural question that arises here is to find the best (largest) lower bound value from Lagrange dual function. That leads to the dual optimization problem of the primal problem and is defined as follows:

$$\begin{aligned} &\text{Maximize} && g(\lambda, \nu) \\ &\text{Subject to,} && \lambda \geq 0 \end{aligned} \tag{3.12}$$

Since, $g(\lambda, \nu)$ is concave and the constraint is convex the dual problem (3.12) is always a convex optimization problem regardless of the convexity of the primal problem. This convex problem can be efficiently solved to optimality and let us denote the optimal value of the dual problem as $D^\star$. From Eq. (3.11) we can then write:

$$D^\star \leq P^\star \tag{3.13}$$

Eq. (3.13) holds even if the primal problem is not convex. This property is called weak duality. The nonnegative quantity $P^\star - D^\star$, also known as the optimal duality gap. If the duality gap is zero that is,

$$D^\star = P^\star, \tag{3.14}$$

then strong duality for the problem holds. If the primal problem is a convex optimization problem most of the time (but not always) the strong duality holds and it is possible to get the global optimal solution efficiently. The strong duality holds for the convex optimization problems satisfying some conditions called constraint qualifications, one of which is Slater's condition [76]. Slater's condition requires the existence of a strictly feasible interior point $x \in \textbf{relint } \mathcal{D}$ such that:

$$\begin{aligned} f_i(x) &< 0 \qquad i = 1, 2, ..., m \\ h_j(x) &= 0 \qquad j = 1, 2, ..., p \end{aligned} \tag{3.15}$$

where **relint** is the relative interior, which is a set contains all points which are not on the 'edge' of the set, relative to the smallest subspace in which this set lies. If some of the inequality constraints are affine (let us assume $i = 1, 2, ..., k$), Slater's condition requires the existence of a strictly feasible interior point $x \in \textbf{relint } \mathcal{D}$ such that:

$$\begin{aligned} f_i(x) &\leq 0 \qquad i = 1, 2, ..., k \\ f_i(x) &< 0 \qquad i = k+1, k+2, ..., m \\ h_j(x) &= 0 \qquad j = 1, 2, ..., p \end{aligned} \tag{3.16}$$

In other words, Slater's condition requires the existence of variables which satisfy the inequality constraints strictly and also satisfy the equality constraints [28].

3.6 Karush-Kuhn-Tucker (KKT) Conditions for Convex Problems

When the primal problem is convex and the strong duality holds (constrains satisfy Slater's condition), the KKT conditions are also sufficient for the points to be primal

and dual optimal. In other words, if f_i are convex and h_i are affine, for optimal value of primal and dual variables defined as $x^\star$, $\lambda^\star$ and $\nu^\star$, respectively, must satisfy the following conditions:

$$\begin{aligned} f_i(x^\star) &< 0 \qquad i = 1, 2, ..., m \\ h_j(x^\star) &= 0 \qquad j = 1, 2, ..., p \\ \lambda_i^\star &\geq 0 \qquad i = 1, 2, ..., m \\ \lambda_i^\star f_i(x^\star) &= 0 \qquad i = 1, 2, ..., m \\ \nabla f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(x^\star) + \sum_{j=1}^{p} \nu_i^\star \nabla h_i(x^\star) &= 0 \end{aligned} \tag{3.17}$$

3.7 Framework for Cross-layer Design

In [84], the authors have formulated a general cross-layer throughput optimization problem for multi-hop wireless networks as following:

$$\text{Maximize} \quad U(\mathbf{r}) \tag{3.18}$$

$$\text{Subject to,} \quad (\mathbf{r}, \mathbf{f}) \in \mathcal{N} \tag{3.19}$$

$$(\mathbf{c}, \mathbf{p}) \in \mathcal{C} \tag{3.20}$$

$$\sum_{i \in S} f_i^\ell \leq c^\ell \tag{3.21}$$

Given that,

$\mathbf{r}$= Throughput vector, where, r_i is the throughput for a session $i \in S$.

S= Set of sessions.

$\mathbf{f}$= Flow rate vector, where, f_i^ℓ be the flow rate of session i going through a link $\ell \in E$.

E =Set of links.

$\mathbf{c}$= Capacity vector, where, c^{ℓ} is the capacity of link ℓ.

$\mathbf{p}$= Maximum node power vector.

$\mathcal{N}$= Routing region.

$\mathcal{C}$=Capacity region.

The objective of the optimization problem in Eq. (3.18) is given by a utility function $U(\mathbf{r})$, which aims to increase the network throughput, where,

$$U(\mathbf{r}) = \sum_{i \in S} \log(1 + r_i) \tag{3.22}$$

Eq. (3.19) shows the interdependence between achievable throughput $\mathbf{r}$ and flow routing $\mathbf{f}$. Here, $(\mathbf{r}, \mathbf{f})$ is defined by a routing region $\mathcal{N}$, so that the through $\mathbf{r}$ can be supported by the flow rate $\mathbf{f}$ [84]. Eq. (3.20) shows the interdependence of capacity vector and node power constraint. The capacity region $\mathcal{C}$ defines ($\mathbf{c}$,$\mathbf{p}$) in a way that power $\mathbf{p}$ can support capacity $\mathbf{c}$ [84]. Lastly, Eq. (3.21) enforces that the flow rate of a link can not exceed the link capacity.

In our cross-layer optimization framework we however do not directly use this general formulation to maximize the network throughput, rather we assume a TDMA (Time Division Multiple Access) scheme to divide the time into slots and schedule links to be active in different time slots to manage interference from adjacent concurrent communications. Our objective is to improve the network performance by minimizing the total scheduling period (system activation time).

3.8 Column Generation Technique for Large Scale Optimization

In this section, we discuss Column Generation (CG) technique [79, 85, 86] which we use to successfully solve large scale MILP problems. Recently, column generation technique is attracting a lot of attentions and several large scale cross-layer optimization problem has been successfully solved using this technique [41, 47, 87–91]. Although CG technique was first developed to solve large scale linear and integer linear problems, recently the authors of [91] have successfully implemented CG technique to solve large scale convex optimization problems. However, the underlying concept of CG in convex optimization problem is a bit different than that of linear and integer linear problems. While CG in convex optimization problem exploits Lagrange dual function to check the optimality of the solution (primal-dual column generation), CG in LP uses the Simplex theory to find the optimal solution (classical column generation). In this thesis we use CG technique to solve large scale LP and in this section we will discuss column generation technique strictly for the case of large scale linear programming problems.

3.8.1 Motivation

The main motivation behind using CG is that many LP models are too huge to explicitly consider all their variables and constraints. Furthermore, in the final optimal solution, most of the variables will be nonbasic and assume a value of zero, and only a small subset of variables will be basic (assume a value different from zero). The idea behind CG is to consider only the variables which have the potential to improve the objective function, by generating them on demand by means of new columns [79].

3.8.2 Modeling a Problem using Column Generation

CG technique models a problem by decomposing an optimization problem into a master problem and a pricing subproblem. Instead of having all columns (or possible solutions) inserted in the master problem, the master problem is initialized with a subset of columns. The master problem is then referred as the restricted master problem (RMP). The pricing subproblem generates new columns to be added in the basis of the master problem. The pricing subproblem keeps generating and adding columns to the master problem as long as there exists one that can improve the solution of the restricted master problem.

Let us assume a linear programming problem of the following form:

$$\begin{aligned} \text{Minimize} \quad & \sum_{i \in \mathcal{I}} c_i x_i \\ \text{Subject to,} \quad & \sum_{i \in \mathcal{I}} \mathbf{a}_i x_i \geq \mathbf{b} \qquad i = 1, 2, ..., m \\ & x_i \geq 0 \qquad i \in \mathcal{I} \end{aligned} \tag{3.23}$$

We refer to this problem as the master problem (MP). Let us assume $\boldsymbol{\mu}$ is the non-negative vector of dual variable. To solve the problem, in each iteration of the Simplex method, we try to find a nonbasic variable to price out and enter the basis. In other words, we wish to find a column $i \in \mathcal{I}$ which minimizes $\overline{c_i} = c_i - \boldsymbol{\mu}^T \mathbf{a}_i$. An explicit pricing can become an impractical operation when $|\mathcal{I}|$ is huge. As a result of that we work with a reasonably small subset $\overline{\mathcal{I}} \subseteq \mathcal{I}$ of columns, thus the master problem is transformed into a restrictive master problem. If $\boldsymbol{x}^\star$ and $\boldsymbol{\mu}^\star$ are the optimal primal and dual variable solutions of the current RMP, respectively, and columns $\mathbf{a}_i$, $i \in \mathcal{I}$ are given as element of a nonempty set $\mathcal{A}$ then the following subproblem performs

the pricing (pricing subproblem) [92]:

$$\text{Minimize} \qquad \overline{c}^{\star} = \{c(\mathbf{a}) - \boldsymbol{\mu}^{\star T}\mathbf{a} \;\; | \; \mathbf{a} \in \mathcal{A}\} \tag{3.24}$$

If $\overline{c}^{\star} \geq 0$, then no negative $\overline{c_i}$, $i \in \mathcal{I}$ exists and the optimal solution of variable $\boldsymbol{x}^{\star}$ to the RMP optimally solves the MP as well. Otherwise, we extend the basis of the RMP by adding the column $\mathbf{a}$ derived from the pricing subproblem (3.24) solution and repeat the process by re-optimizing the RMP. As we mentioned earlier, the iteration process between restricted master problem and the pricing subproblem starts with a small subset of columns $\overline{\mathcal{I}} \subseteq \mathcal{I}$, the initial subset of columns $\overline{\mathcal{I}}$ can be any feasible artificial solution of the problem and they can be obtained using a heuristic.

Chapter 4

Cross-layer Design with Optimal Spectrum Partitioning

In this chapter, we study in the context of wireless mesh networks design, the benefits of variable-width channel allocation on the end-to-end throughput performance. We assume total system spectrum can be divided into multiple orthogonal channels of variable bandwidths and multiple spatially separated transmissions can be assigned the same channel if the interference between them are low enough that all the transmissions can successfully transmit their data. We consider single-radio multichannel networks and study the joint problem of routing and transmission scheduling. We present a cross-layer problem formulation which incorporates multi-path routing and link layer scheduling; link layer scheduling refers to the problem of determining a set of transmission links (their channel assignment and spectrum allocation) which can be concurrently active without violating the signal to interference plus noise ratio (SINR) requirement at their intended receivers. The set of transmission links, their channel assignments, and their spectrum allocation, is referred to as a configuration. In our formulation, we assume a fixed link spectral efficiency (bps/Hz), therefore the link capacity will only depend on the spectrum-width of the channel used on the link

and the level of interference. Later, we incorporate transmission rate adaptation into this model. We assume different modulation and coding schemes and each of these schemes results into a particular data rate.

We mathematically model this combinatorially complex (NP-complete) problem as a mixed integer linear program (MILP) and we present a decomposition approach based on column generation (CG) for solving it; a pricing subproblem (ILP) enumerates only those configurations that contribute towards determining the optimal solution of the master problem [41, 87–89]. The CG technique however allows us to solve exactly the design problem only for small size networks; the difficulty arises from solving the pricing subproblem (ILP) in an efficient manner and the fact that this subproblem is solved repeatedly.

To overcome this issue, we propose to reduce the size of the ILP pricing subproblem by introducing a greedy heuristic for partitioning the spectrum and thus fixing some of the integer variables in the model. Then, the ILP is more effectively solved with smaller subset of the binary variables. Our method yields a much faster and very close to optimal solution.

Finally, we present an alternate design model for the variable bandwidth adaptation problem, where we assume, regardless of the spectrum widths on the allocated channels, all links must satisfy a fixed capacity requirement. Numerical results confirm that variable-width spectrum assignment achieves significant improvement over fixed-width allocation in all the instances of the problem. Note that, like most other centralized problem formulations, we do not claim that our centralized solutions are the best practical way to operate WMNs, rather it will provide benchmark design solutions which can be used as a lower-bound on determining the system activation time (i.e., upper-bound on performance) for WMNs using random access protocols or different distributed models [90, 93].

In the rest of the chapter, the system model is discussed in Section 4.1. In Section 4.2, we present the different constraints for the problem and in Section 4.3 we present the formulation. Section 4.4 presents the problem decomposition. We present the joint variable width and rate adaptation model in Section 4.5. The heuristic is presented in Section 4.6. In Section 4.7 we present an alternate design approach. Numerical results and discussions are given in Section 4.8 and we conclude in Section 4.9.

4.1 System Model

We model a multi-hop wireless mesh network as a set of $\mathcal{N}(|\mathcal{N}| = N)$ stationary nodes. Each node is equipped with a single radio and relays data to neighboring nodes. We consider a total system spectrum B(=80MHz), which can be divided into smaller spectrum blocks. We assume 5, 10, 20 and 40MHz spectrum blocks. The total number of orthogonal channels (k) vary between 16 and 2 (when all channels are 5 and 40MHz wide respectively). We also assume that each radio can dynamically switch across different available channels and dynamically configure the spectrum width. A node can either transmit or receive at a time to or from nodes within its communication range. Consider a single transmission link between two neighboring nodes i and j that are $d_{i,j}$ distant from each other. In the absence of any interference and assuming a simple 2-ray propagation model, the signal to noise ratio (SNR) at the intended receiver is determined as:

$$SNR_{i,j} = \frac{P_t d_{i,j}^{-\alpha}}{\eta_0 W} \tag{4.1}$$

where, P_t is the transmission power, α is the path loss exponent (varies between 2 and 4), W is channel width and η_0 is the power spectral density of thermal noise.

For fixed transmission power and η_0, the SNR depends on two factors, $d_{i,j}$ and W. The expression in Eq. (4.1) indicates that for a given transmission link, a smaller channel width yields a higher SNR at the intended receiver than a larger channel width. This SNR must be larger than a given threshold (β) to meet the transmission requirement of a particular data rate. Denote by T_k the transmission range on a particular channel width (W_k) which corresponds to $SNR = \beta$; T_k can be derived as follows:

$$T_k = \sqrt[\alpha]{\frac{P_t}{\eta_0 W_k \beta}} \tag{4.2}$$

This expression shows that, for the same transmission power, smaller width channels have higher transmission range than larger width channels. This result has been already observed by the authors of [14] from empirical measurements. Given 4 available spectrum bands (5, 10, 20, 40MHz), the maximum transmission range (T_{max}) corresponds to the narrowest band ($W_k = 5$). Therefore, a directed link may exist between any two nodes i and j if and only if, $d_{i,j} \leq T_{max}$. The set of all such links is represented by ε. It is important to note that $d_{i,j} \leq T_{max}$ does not imply $d_{i,j} \leq T_k$ and hence a link between nodes i and j may only exist for smaller widths.

Now, in the presence of interference from neighboring transmitters, the signal to interference plus noise ratio (SINR) at receiver j of link (i, j) is given by:

$$SINR_{i,j} = \frac{P_t d_{i,j}^{-\alpha}}{\eta_0 W_k + \sum\limits_{a \neq (i,j)} P_t d_{a,j}^{-\alpha}} \tag{4.3}$$

Here, W_k is the width of the channel assigned to link (i, j) and $d_{a,j}$ is the distance from interfering transmitter a to receiver j on the same channel. Since we are assuming orthogonal channels, transmission on other neighboring links (but different channels) would not interfere with the transmission of link (i, j). For a successful transmission, the $SINR$ at receiver j must satisfy:

$$\frac{P_t d_{i,j}^{-\alpha}}{\eta_0 W_k + \sum_{a \neq (i,j)} P_t d_{a,j}^{-\alpha}} \geq \beta \tag{4.4}$$

The capacity of a wireless link (i, j) which is allocated a spectrum of width W_k, is a function of $SINR_{i,j}$; we assume each link can be viewed as a single user Gaussian channel, and the Shannon Capacity of the link (i, j) is given by:

$$C_{i,j} = W_k \log_2(1 + SINR_{i,j}) \tag{4.5}$$

It follows from Eq. (4.5) that a larger width results in larger capacity (although in practice most communication links achieve lower rates than Eq. (4.5)). A larger width, however, results in smaller transmission range, which in turn results in more hops along the end-to-end route of a flow (Eq. 4.2). The more hops a route contains, the stronger is the effect of intra-path interference.

To summarize, smaller widths result in more orthogonal channels which allow for more concurrent non-interfering communications in the same area, each such communication has larger transmission range but smaller link capacity. On the other hand, a wider spectrum width increases the link capacity and reduces the transmission range; this however results in smaller number of orthogonal channels (thus fewer parallel concurrent transmissions) and stronger effect of intra-path interference. Given these conflicting objectives, it is clear that a variable spectrum width allocation (rather than fixed) may strike a good balance between interference control and maintaining both higher concurrency and better spectrum reuse for wireless networks.

4.2 Problem Formulation

We consider multi-hop wireless mesh networks with a maximum of K orthogonal channels, each is allocated a spectrum from (5, 10, 20, 40MHz). We formulate the joint routing, scheduling and channel and spectrum allocation in such networks as an optimization problem. We consider M concurrent multi-hop sessions, each of which corresponds to a source destination (s-d) pair in the network. The traffic demand for each session m ($1 \leq m \leq M$) is given by R_m (bits) which is to be transmitted from a source s_m to a destination d_m. Traffic of a particular session may be split to sub-flows routed over different paths. The choice of these routing paths depends on the underlying schedule of different concurrent transmissions. The objective of our model is to minimize the system activation time for delivering the M sessions without violating the minimum SINR requirement (Eq. (4.4)). We assume a TDMA access scheme where time is divided into slots and a link may be active in one or more time slots to meet the traffic requirement. Several links may be active in the same time slot without violating the SINR requirement; such a set of links and their spectrum allocation is referred to as a configuration. A configuration p is defined as a set of elements $W^p_{i,j}$, $(\forall (i,j) \in \varepsilon)$.

We introduce link binary variable, $x^{k,p}_{i,j}$, which takes a value of 1 if link (i,j) is active (on channel k) in configuration p and 0 otherwise. Given a single radio per node, a maximum of one channel per radio, and the fact that a node can either transmit or receive at a time, we write the radio transmission constraint as follows:

$$\sum_{k=1}^{K} \sum_{j:(n,j)\in\varepsilon} x^{k,p}_{n,j} + \sum_{k=1}^{K} \sum_{j:(j,n)\in\varepsilon} x^{k,p}_{j,n} \leq 1 \qquad \forall n \in \mathcal{N} \tag{4.6}$$

For spectrum partitioning in a configuration p, we define a spectrum partitioning variable $b^{k,p}$, which can take values from the set {0, 5, 10, 20, 40} depending upon

the spectrum block allocated to a channel k. To assure that, the following two constraints are introduced:

$$b^{k,p} = \sum_{v=1}^{4} 2^{(v-1)} \times 5 \times z_v^{k,p} \qquad k = 1, 2, ..., K \tag{4.7}$$

$$\sum_{v=1}^{4} z_v^{k,p} \leq 1 \qquad k = 1, 2, ..., K \tag{4.8}$$

where $z_v^{k,p}$ is a binary indicator variable. Given a total spectrum of B and a maximum of K orthogonal channels, the spectrum partitioning constraint is written as:

$$\sum_{k=1}^{K} b^{k,p} \leq B \tag{4.9}$$

Note that, in a configuration p, when $b^{k,p} = 0$, then no link $(i, j) \in \varepsilon$ can use that particular channel k and $x_{i,j}^{k,p}$ must be 0. Further if $b^{k,p} \neq 0$, then any link $(i, j) \in \varepsilon$ may or may not be active using that particular channel (i.e., $x_{i,j}^{k,p}$ can be 0 or 1). While $x_{i,j}^{k,p}$ is a binary variable and $b^{k,p}$ can only take values from (0, 5, 10, 20 and 40), the relation between link variables and spectrum partitioning variables can be written by the following equation.

$$x_{i,j}^{k,p} \leq b^{k,p} \qquad \forall (i, j) \in \varepsilon, k = 1, 2, ..., K \tag{4.10}$$

Accordingly, the element of a configuration p can be calculated as follows,

$$W_{i,j}^{p} = \begin{cases} b^{k,p} & \text{if } (i, j) \text{ is active, (i.e., } x_{i,j}^{k,p} = 1) \\ 0 & \text{if } (i, j) \text{ is inactive, (i.e., } \sum_{k=1}^{K} x_{i,j}^{k,p} = 0) \end{cases}$$

To linearize this above definition, we introduce the following expressions:

$$W_{i,j}^{p} \leq b^{k,p} + L(1 - x_{i,j}^{k,p}) \qquad \forall (i,j) \in \varepsilon, k = 1, 2, ..., K \qquad (4.11)$$

$$W_{i,j}^{p} \leq L \sum_{k=1}^{K} x_{i,j}^{k,p} \qquad \forall (i,j) \in \varepsilon \qquad (4.12)$$

$$W_{i,j}^{p} \geq b^{k,p} - L(1 - x_{i,j}^{k,p}) \qquad \forall (i,j) \in \varepsilon, k = 1, 2, ..., K \qquad (4.13)$$

$$W_{i,j}^{p} \geq 0 \qquad \forall (i,j) \in \varepsilon \qquad (4.14)$$

where, L is a large positive constant.

Here, when $x_{i,j}^{k,p} = 1$, Eq. (4.12) becomes redundant and $W_{i,j}^{p} = b^{k,p}$; when $\sum_{k=1}^{K} x_{i,j}^{k,p} = 0$, both Eqs. (4.11) and (4.13) become redundant and $W_{i,j}^{p} = 0$.

To maintain the interference in check for a particular configuration, the SINR requirement (β) for any link $(i,j) \in \varepsilon$ operating on a particular channel k with a spectrum band $W_{i,j}^{p}$, is guaranteed as follows:

$$P_t d_{i,j}^{-\alpha} x_{i,j}^{k,p} + L(1 - x_{i,j}^{k,p}) \geq \beta(\eta_0 W_{i,j}^{p} + \sum_{(a,b) \in \varepsilon, a \neq i} (P_t d_{a,j}^{-\alpha} x_{a,b}^{k,p})) \; \forall (i,j) \in \varepsilon, \; k = 1, 2, ..., K \qquad (4.15)$$

If link (i,j) is active on a certain channel k (i.e., $x_{i,j}^{k,p} = 1$), then Eq. (4.15) reduces to Eq. (4.4). Otherwise L ensures that Eq. (4.15) is always satisfied.

4.3 Joint Routing and Scheduling Formulation

Given a set of $\mathcal{M}(|\mathcal{M} = M|)$ concurrent sessions, our objective is to jointly determine routes for these sessions, feasible link schedules (or configurations) and spectrum allocation using the minimum system activation time. Let $\mathcal{P}$ denote the set of all feasible configurations, and $p = \{W_{i,j}^{p}, \forall (i,j) \in \varepsilon\}$ where $p \in \mathcal{P}$ is a certain feasible

link schedule and let $|\mathcal{P}| = \bar{P}$. Define λ_p to be the time (in second) during which configuration p is active. When a configuration p is scheduled, then $\lambda_p > 0$, otherwise $\lambda_p = 0$. Let $f_{i,j}^m$ denote the flow of traffic (bits) for session $m \in \mathcal{M}$ passing through link $(i,j) \in \varepsilon$. The LP formulation of the joint problem is presented as following:
Objective:

$$\text{Minimize} \sum_{p=1}^{\bar{P}} \lambda_p \tag{4.16}$$

Subject to:

$$\sum_{j \in \mathcal{N}:(i,j)\in\varepsilon} f_{i,j}^m - \sum_{j \in \mathcal{N}:(j,i)\in\varepsilon} f_{j,i}^m = 0 \qquad \forall i \in \mathcal{N} - \{s_m, d_m\} \quad \forall m \in \mathcal{M} \tag{4.17}$$

$$\sum_{j \in \mathcal{N}:(s_m,j)\in\varepsilon} f_{s_m,j}^m - \sum_{j \in \mathcal{N}:(j,s_m)\in\varepsilon} f_{j,s_m}^m = R_m \qquad \forall m \in \mathcal{M} \tag{4.18}$$

$$\sum_{j \in \mathcal{N}:(d_m,j)\in\varepsilon} f_{d_m,j}^m - \sum_{j \in \mathcal{N}:(j,d_m)\in\varepsilon} f_{j,d_m}^m = -R_m \qquad \forall m \in \mathcal{M} \tag{4.19}$$

$$\sum_{p \in \bar{P}} \lambda_p \times W_{i,j}^p \times c_{i,j} - \sum_{m=1}^{M} f_{i,j}^m \geq 0 \qquad \forall (i,j) \in \varepsilon \tag{4.20}$$

$$f_{i,j}^m \geq 0, \lambda_p \geq 0.$$

Here, the objective function in Eq. (4.16) aims to minimize the total system activation time to satisfy all the traffic demands. Eqs. (4.17-4.19) present the flow conservation constraints in the network. Eq. (4.20) indicates that the total traffic routed through a link (i,j) can not exceed the total transport capacity of the link. The term $W_{i,j}^p \times c_{i,j}$ indicates the transport capacity of a link assigned a spectrum band of $W_{i,j}^p$ width (Hz) and $c_{i,j}$ is the spectral efficiency (bps/Hz). To keep this model simple, we assume a fixed value for $c_{i,j} = c$ and thus the transport capacity of a feasible link depends only on the size of the spectrum band allocated. It is to be noted that in a configuration p, a link is assumed feasible when the SINR requirement

is satisfied at the intended receiver.

Finding the solution of the above model relies on determining the set of all feasible configurations $\mathcal{P}$. According to the number of links in the network, the number of orthogonal channels, and the number of possible channel widths, the size of $\mathcal{P}$ can be extremely large. This makes the above LP computationally infeasible, since it may not be possible to enumerate all such configurations. A more effective solution is to solve the problem without having to explicitly enumerate all possible configurations [41, 87–89]. Such a method will be presented next.

4.4 Column Generation Decomposition Method

Column generation (CG) [79, 85, 86] is an exact optimization technique that decomposes an optimization problem into a master model and a pricing model. The restricted master model is initialized with a subset of columns, the basis, P_0 (in our case, configurations, $P_0 \subseteq \mathcal{P}$) of the original problem and solved to obtain a feasible solution to the problem. The pricing subproblem is solved to identify whether the master should be enlarged with additional columns or not. Therefore, as opposed to the original problem where all the columns are needed to be used at the same time to obtain the optimal solution, CG alternates between the master (LP) and the pricing (ILP) models, until the former contains the necessary columns to find the optimal solution of the problem [90].

MASTER PROBLEM

$$\text{Objective:} \qquad \text{Minimize} \sum_{p \in P_0} \lambda_p \tag{4.21}$$

Subject to: Equations (4.17-4.19),

$$\sum_{p \in P_0} \lambda_p \times W_{i,j}^p \times c - \sum_{m=1}^{M} f_{i,j}^m \geq 0 \quad \forall (i,j) \in \varepsilon \tag{4.22}$$

$$f_{i,j}^m \geq 0, \lambda_p \geq 0.$$

During every iteration, when the master problem is solved, we need to verify the optimality of the solution. If it is optimal, we conclude our search, otherwise decide a new column to join in its current basis that can improve the current solution. This can be achieved by examining whether any new column that is not currently in P_0, has a negative reduced cost. Denoting the dual variables corresponding to Eq. (4.22) by $u_{i,j}$, the reduced cost ($\overline{cost}$) for any new column that is not in P_0 can be expressed as

$$\overline{cost} = 1 - \sum_{(i,j) \in \varepsilon} u_{i,j} \times W_{i,j} \times c \tag{4.23}$$

When the objective is to minimize, the standard pivoting rule of the Simplex method is to choose the column (in $\mathcal{P}$) where $\sum_{(i,j) \in \varepsilon} u_{i,j} \times W_{i,j}$ is maximum; if (max ($\sum_{(i,j) \in \varepsilon} u_{i,j} \times W_{i,j} \times c$) ≥ 1) (i.e., a negative reduced cost), the column (in $\mathcal{P}$) that is found is added to the basis (P_0). The master model is solved, again, with the new basis to obtain a new solution, and the dual variable is passed to the pricing which is again solved. The two subproblems are solved iteratively until there is no off-basis column with a negative reduced cost found and therefore the solution is optimal [79]. Indeed, this requires that the last Simplex iteration of the pricing model is solved to optimality to ensure that there is no off basis column with negative reduced cost remaining in $\mathcal{P}$.

PRICING PROBLEM

$$\text{Objective:} \quad \text{Minimize} \quad \overline{cost} = 1 - \sum_{(i,j)\in\varepsilon} u_{i,j} \times W_{i,j} \times c \tag{4.24}$$

Subject to: Equations (4.6-4.15)

Note that, in the model, $W^p_{i,j}$ (the elements of a configuration p) are used as fixed parameters in the master problem and as variables in the pricing, and the dual variables ($u_{i,j}$) corresponding to Eq. (4.22) are used as fixed parameters in the pricing subproblem. Also note that, since the pricing subproblem deals with a particular configuration at a time, $W^p_{i,j}$, $x^{k,p}_{i,j}$, $b^{k,p}$ and $z^{k,p}_v$ are replaced by $W_{i,j}$, $x^k_{i,j}$, b^k and z^k_v respectively in the pricing. For convenience, we present the list of variables and parameters in Table 4.1.

4.5 A Joint Model with Spectrum & Transmission Rate Allocation

This section extends our previous model to consider, in addition to variable spectrum-width allocation, the allocation of link transmission rates. Note that 802.11-based networks support data rates with discrete values and a particular transmission rate can be achieved from a particular modulation/coding scheme. To achieve these discrete set of data rates for a certain bit error rate (BER) requirement, a transmission needs to satisfy a discrete set of SINR threshold requirements. In our work, we consider 4 modulation/coding schemes ($|\mathcal{R}| = 4$); Table 4.2 shows the mapping of SINR requirements to corresponding data rates for our considered modulation/coding

Table 4.1: A List of all parameters and variables

Parameters/Variables	Definition
P_t	Transmission power
R_m	Total traffic for session m
$d_{i,j}$	Distance between nodes i and j
s_m/d_m	Source/destination node for session m
η_0	Background noise
$H_k/S_{i,j}^k$	Multiplication variables
W_k	Channel bandwidth
K	Number of orthogonal channels
α	Path loss exponent
y_v^k/z_v^k	Binary indicator variables
T_k	Transmission range using channel width W_k
B	Total bandwidth
β	SINR threshold
λ_p	System activation time
C	Capacity
R	Number of transmission rates
p	Configuration
L	Large constant
$x_{i,j}^{k,p}$	Link binary variable
M	Number of concurrent multi-hop sessions
$b^{k,p}$	Spectrum partitioning variable
c	Spectral Efficiency
$f_{i,j}^m$	Traffic flow variable
$\overline{cost}$	Reduced cost for pricing problem
$W_{i,j}^p$	Channel bandwidth variable (variable in pricing parameter in master)
$u_{i,j}$	Dual variable (variable in master parameter in pricing)

Table 4.2: SINR-threshold values for different transmission rates in IEEE 802.11

Modulation	Coding Rates	Tran. Rates C^r (Mbps)	SINR Threshold β^r(dB)
BPSK	1/2	6	6.02
BPSK	3/4	9	7.78
QPSK	1/2	12	9.03
QPSK	3/4	18	10.79

schemes [94]. Note that, the table shows SINR requirements for BER less than or equal to 10^{-5}. Denote by β^r the SINR requirement corresponding to a transmission rate of $r, r \in \mathcal{R}$. Since β^r varies with the corresponding modulation/coding scheme used, the transmission range also will vary :

$$T_k^r = \sqrt[\alpha]{\frac{P_t}{\eta_0 W_k \beta^r}} \tag{4.25}$$

This expression shows that, the maximum transmission range (T_{max}) corresponds to the combination of the narrowest band ($W_k = 5$MHz) and lowest SINR requirement ($\beta^r = 6.02$dB). In other words, a denser network topology is obtained when selecting both smaller widths and transmission rates, and a sparser topology is obtained with larger spectrums and higher transmission rates. In the latter one (i.e., when transmission range is smaller), the SINR requirement on each link becomes higher and fewer orthogonal channels are available (thus, limiting transmission concurrency); however, the transport capacity per link is higher (due to the larger bandwidth per link). The shorter range indeed implies that sessions will be routed through multi-hop routes, which results in higher intra-path interference. Now, since each communication link requires higher SINR for correct reception, the communication becomes very prone to interference from neighboring links. On the other hand, both smaller width and transmit rate result in more orthogonal channels with smaller transport capacity each and smaller SINR requirement. The smaller SINR requirement and higher number

of orthogonal channels result in better spectrum reuse; however, a smaller spectrum width yields to links with smaller capacity resulting in higher system activation time. Clearly, there are conflicting objectives and our model attempts to determine the best combination of spectrum and transmission rates per link, to route the multi-hop demands while minimizing the system activation time. In this formulation we assume, as before, that a link between nodes i and j exists if the distance between them ($d_{i,j}$) is smaller or equal to T_{max}. The set of all such links is represented by ε. $x_{i,j}^{k,r}$ and $W_{i,j}^{r}$ are defined as link binary variables and link bandwidth variables for different modulation/coding schemes $r \in \mathcal{R}$.

MASTER PROBLEM

Objective: Equation (4.21)

Subject to: Equations (4.17-4.19)

$$\sum_{p \in P_0} (\lambda_p \times \sum_{r=1}^{R} W_{i,j}^{p,r} c^r) - \sum_{m=1}^{M} f_{i,j}^{m} \geq 0 \qquad \forall (i,j) \in \varepsilon \tag{4.26}$$

$$f_{i,j}^{m} \geq 0, \lambda_p \geq 0$$

where, c^r in Eq. (4.26) is the spectral efficiency (bps/Hz) which, unlike before, varies according to the modulation/coding scheme used as shown in Table 4.2. In the 802.11 standard, the transmission rates (C^r) for different modulation/coding schemes are given assuming a fixed channel bandwidth of 20 MHz. Therefore, the spectral efficiency is normalized as: $c^r = \frac{C^r}{20}$.

PRICING PROBLEM

Objective:

$$\text{Minimize } \overline{cost} = 1 - \sum_{(i,j) \in \varepsilon} (u_{i,j} \sum_{r=1}^{R} W_{i,j}^{r} c^r) \tag{4.27}$$

Subject to:

$$\sum_{k=1}^{K}\sum_{r=1}^{R}\sum_{j:(n,j)\in\varepsilon} x_{n,j}^{k,r}+\sum_{k=1}^{K}\sum_{r=1}^{R}\sum_{j:(j,n)\in\varepsilon} x_{j,n}^{k,r}\leq 1 \qquad \forall n\in\mathcal{N} \quad (4.28)$$

$$\sum_{k=1}^{K} b^k \leq B \quad (4.29)$$

$$\sum_{r=1}^{R} x_{i,j}^{k,r} \leq b^k \qquad \forall (i,j)\in\varepsilon,\ \ k=1,2,...,K \quad (4.30)$$

$$W_{i,j}^{r} \leq b^k + L(1-x_{i,j}^{k,r}) \qquad \forall (i,j)\in\varepsilon,\ \ k=1,2,...,K,\ \ \forall r\in\mathcal{R} \quad (4.31)$$

$$W_{i,j}^{r} \leq L\sum_{k=1}^{K} x_{i,j}^{k,r} \qquad \forall (i,j)\in\varepsilon,\ \ \forall r\in\mathcal{R} \quad (4.32)$$

$$W_{i,j}^{r} \geq b^k - L(1-x_{i,j}^{k,r}) \qquad \forall (i,j)\in\varepsilon,\ \ k=1,2,...,K,\ \ \forall r\in\mathcal{R} \quad (4.33)$$

$$P_t d_{i,j}^{-\alpha} x_{i,j}^{k,r} + L(1-x_{i,j}^{k,r}) \geq \beta^r(\eta_0 W_{i,j}^{r} + \sum_{(a,b)\in\varepsilon, a\neq i}(P_t d_{a,j}^{-\alpha}\sum_{q=1}^{R} x_{a,b}^{k,q})) \qquad \forall (i,j)\in\varepsilon,\ \ k=1,2,...,K,\ \ \forall r\in\mathcal{R} \quad (4.34)$$

$$b^k = \sum_{v=1}^{4} 2^{(v-1)}\times 5\times z_v^k \qquad k=1,2,...,K \quad (4.35)$$

$$\sum_{v=1}^{4} z_v^k \leq 1 \qquad k=1,2,...,K \quad (4.36)$$

$$W_{i,j}^{r}\geq 0,\quad x_{i,j}^{k,r}=\{0,1\},\quad z_v^k=\{0,1\}.$$

In this design formulation of joint channel width and transmission rate adaptation, both the master and pricing subproblems are modeled similar to the previous design. The only difference is that we have added another index r to our variables $(W_{i,j}^{r}, x_{i,j}^{k,r})$. Since, a link can not use more than one transmission rate at a time, constraint (4.28) of the pricing problem restricts a link to a single transmission rate (along with single

radio constraint).

4.6 A Greedy Heuristic for Pricing

The CG decomposition technique allows us to solve the joint design problem exactly. The approach taken by CG is to repeatedly iterate between master and pricing subproblems; the pricing subproblem generates promising configurations/schedules and pass them back to the master. As long as promising configurations are found (in the pricing), the iteration continues. Observe that the master problem deals with non-integer variables ($f_{i,j}^m, \lambda_p$) and usually is easy to solve (LP). The pricing however deals mostly with integer variables ($x_{i,j}^k, z_v^k$) which is commonly more difficult to solve (ILP). As the network size gets larger, the number of links and thus the number of integer variables ($x_{i,j}^k$) increase exponentially and the pricing subproblem becomes very difficult to solve. Therefore, and we verify later, our decomposition approach for solving this joint optimization problem is limited to only small size networks. To solve the pricing subproblem more efficiently, we introduce a greedy heuristic to reduce the number of integer variables (e.g., $x_{i,j}^k$) and thus speed up the solution of the ILP pricing. The heuristic works by observing that in our pricing problem, a maximum of K ($K = 16$) bands are available in the spectrum (i.e., 16 channels with 5MHz bandwidth each), and thus the maximum number of link integer variable $x_{i,j}^k$ is K times the number of links. However, depending on the spectrum partitioning, the number of spectrum bands (and hence the number of binary link variables) may be much smaller than K. In our greedy heuristic (Algorithm 4.1) we start with an initial bandwidth partitioning $b_{initial}^k$ with as few as two bands ($K = 2$) of 40MHz each and solve the pricing problem. If we obtain a configuration $p_{initial}$ with negative reduced cost ($cost_{initial} < 0$), we add that configuration as a new column into the basis of the master problem. Otherwise, we generate a new candidate bandwidth

partitioning solution b^k_{can} from the current bandwidth partitioning solution b^k_{cur} (initially, $b^k_{cur} = b^k_{intial}$) by removing at random one spectrum band (i.e., channel) from b^k_{cur} and adding one or more spectrum bands instead, such that $\sum_{k=1}^{K} b^k_{can} = B$ (Line 9 of Algorithm 4.1). We iteratively, generate different b^k_{can} from b^k_{cur} and compare the current ($cost_{cur}$) and candidate ($cost_{can}$) cost as long as we get a candidate configuration (p_{can}) with negative reduced cost ($cost_{can} < 0$) or exhaust the maximum number of iterations ($Loop$). If we get a configuration with negative reduced cost before the number of iterations is exhausted, we add that configuration as new column into the basis of the master problem, otherwise we terminate the program. As before, the alternation between master and pricing problem continues as long we can get configurations with negative reduced cost from the heuristic.

Algorithm 4.1 Greedy Heuristic

1: Initialize $b^k_{initial} = \{40, 40\}$
2: Get $cost_{initial}$ and $p_{initial}$ by solving PRICING ILP by fixing b^k to $b^k_{initial}$
3: **if** $cost_{initial} < 0$ **then**
4: Add $p_{initial}$ as new column into the basis of the master problem
5: **else**
6: $b^k_{cur} = b^k_{initial}$, $cost_{cur} = cost_{initial}$ and $p_{cur} = p_{initial}$
7: Initialize $Loop$=Maximum number of iterations
8: **while** $i \leq Loop$ **and** $cost_{can} \geq 0$ **do**
9: Generate b^k_{can} from b^k_{cur}
10: Get $cost_{can}$ and p_{can} by solving PRICING ILP by fixing b^k to b^k_{can}
11: **if** $cost_{can} \leq cost_{cur}$ **then**
12: $b^k_{can} = b^k_{cur}$, $cost_{can} = cost_{cur}$ and $p_{can} = p_{cur}$
13: **end if**
14: $i++$
15: **end while**
16: **if** $cost_{cur} < 0$ **then**
17: Add p_{cur} as new column into the basis of the master problem
18: **end if**
19: **end if**

4.7 An Alternative Variable Width Adaptation Design Approach

We now turn our attention to an alternative design of WMNs, where regardless of the spectrum widths on the allocated channels, all links must satisfy a fixed transmission data rate (link capacity) requirement. Note that, this design is not meant to limit a link to a fixed capacity, but rather it ensures that all links in the system must achieve at least a certain link transmission capacity that ensures a target data rate for each user, which is appropriate either to support certain applications with stringent delay constraints or due to some user expectations [55]. Similar to before, our objective is to do joint routing, scheduling and optimal spectrum allocation. From Eq. (4.5), it follows that to maintain a predetermined capacity per link, as the channel bandwidth varies, the required SINR (or SINR threshold) must vary accordingly. For a predetermined transmission capacity C, let β^k be the SINR threshold corresponding to a channel width b^k ($b^k \in [0, 5, 20, 40]MHz$):

$$\beta^k = 2^{\frac{C}{b^k}} - 1 \tag{4.37}$$

From the above expression, we see that smaller channel widths need to satisfy much higher SINR requirements to achieve the same transmission capacity. Substituting Eq. (4.37) into Eq. (4.2), the transmission range (T_k) can be written as follows:

$$T_k = \sqrt[\alpha]{\frac{P_t}{\eta_0 \times b^k \times (2^{\frac{C}{b^k}} - 1)}} \tag{4.38}$$

Table 4.3 shows the different values obtained for β^k and T_k as we vary the channel bandwidth b^k for a 10Mbps link capacity. Transmission power P_t and spectral density of the thermal noise η_0 are assumed 10mWatt and 10^{-6}Watt/MHz respectively.

Table 4.3: Values of β^k and T_k while varying b^k for C =10Mbps

b^k	β^k	$T_k(m)$
0	Inf	0
5	3.0	25.8199
10	1.0	31.6228
20	0.4142	34.7434
40	0.1892	36.3497

We can see from Table 4.3 that transmitting at a fixed data rate over narrower bandwidth translates into a higher SINR requirement which also results in lower transmission range. Clearly, although narrower bands yield more orthogonal channels that can be used concurrently, the reuse of the same channel is however limited due to the higher SINR requirement. Alternatively, wider bands have lower SINR requirement but partitions the spectrum into smaller number of orthogonal channels. With lower SINR requirement, the spectrum spatial reuse is improved as the same channel may be used concurrently by different transmissions. Note that since narrower widths have shorter transmission ranges (higher SINR requirement), the likelihood of multi-hop routing is higher, which results in stronger intra-path interference. Evidently, there are various conflicting objectives and our design in this section will try to find a good balance to achieve optimal system performance.

4.7.1 Problem Formulation

Given a set of $\mathcal{M}$ ($|\mathcal{M}| = M$) concurrent sessions, our objective is to determine routes for these sessions and feasible link schedules using minimum system activation time. We assume fixed transmission data rate and we formulate the joint routing and scheduling problem. The objective of the master problem in Eq. (4.21) and the flow conservation constraints (4.17-4.19) remain unchanged from the original problem.

The capacity constraint in Eq. (4.22) is written as:

$$\sum_{p \in P_0} \lambda_p \times C \times a_{i,j}^p - \sum_{m=1}^{M} f_{i,j}^m \geq 0 \qquad \forall (i,j) \in \varepsilon \tag{4.39}$$

Where, C is the fixed link capacity (a constant) and $a_{i,j}^p$ represents the elements in configuration p. Here, a configuration p is the set of all links which can be active simultaneously without violating the SINR requirement of Eq. (4.4) with β being replaced by β^k. $a_{i,j}^p$ can be written as a function of the link binary variable, $x_{i,j}^{k,p}$:

$$a_{i,j}^p = \sum_{k=1}^{K} x_{i,j}^{k,p} \tag{4.40}$$

For the pricing subproblem, the constraints (4.6), (4.9) and (4.10) of the original problem remain unchanged in the alternate formulation. The SINR constraint (4.15) is rewritten as follows:

$$P_t d_{i,j}^{-\alpha} x_{i,j}^k + L(1 - x_{i,j}^k) \geq \beta^k (\eta_0 b^k + \sum_{(a,b) \in \varepsilon, a \neq i} (P_t d_{a,j}^{-\alpha} x_{a,b}^k)) \quad \forall (i,j) \in \varepsilon, k = 1, 2, ..., K \tag{4.41}$$

Here, β^k, b^k and $x_{a,b}^k$ are all variables and hence Eq. (4.15) becomes quadratic. To linearize the first term in the right hand side of the inequality (i.e., product of β^k and b^k) we consider a multiplication variable H^k. Since there are only 5 channel bandwidth (b^k) options (0, 5, 10, 20 and 40 MHz) and since a particular channel bandwidth has a predetermined SINR-threshold (β^k), H^k is obtained by multiplying the value of β^k which corresponds to the value of b^k (Table 4.3). Now, we can rewrite Eq. (4.41) as follows:

$$P_t d_{i,j}^{-\alpha} x_{i,j}^k + L(1 - x_{i,j}^k) \geq \eta_0 H^k + \beta^k \sum_{(a,b) \in \varepsilon, a \neq i} (P_t d_{a,j}^{-\alpha} x_{a,b}^k)) \quad \forall (i,j) \in \varepsilon, k = 1, 2, ..., K \tag{4.42}$$

where the following expressions linearize the definition of H^k:

$$b^k = \sum_{v=1}^{4} 2^{(v-1)} \times 5 \times y_v^k \qquad k = 1, 2, ..., K \tag{4.43}$$

$$\beta^k = \sum_{v=1}^{4} (2^{\frac{C}{2^{(v-1)} \times 5}} - 1) \times y_v^k + L(1 - \sum_{v=1}^{4} y_v^k) \qquad k = 1, 2, ..., K \tag{4.44}$$

$$H^k = \sum_{v=1}^{4} (2^{\frac{C}{2^{(v-1)} \times 5}} - 1) \times 2^{(v-1)} \times 5 \times y_v^k \qquad k = 1, 2, ..., K \tag{4.45}$$

$$H^k \geq 0 \qquad k = 1, 2, ..., K \tag{4.46}$$

$$\sum_{v=1}^{4} y_v^k \leq 1 \qquad k = 1, 2, ..., K \tag{4.47}$$

$$y_v^k = \{0, 1\}$$

Here, L is a large positive constant. β^k=Inf when, b^k=0. To linearize the second term in the right hand side of Eq. (4.41) (i.e., product of β^k and $x_{i,j}^k$), we substitute the product by a new variable $S_{i,j}^k$, defined as follows:

$$S_{i,j}^k = \begin{cases} \beta^k & \text{if link } (i,j) \text{ is active, on channel } k \text{ (i.e., } x_{i,j}^k = 1) \\ 0 & \text{Otherwise (i.e., } x_{i,j}^k = 0) \end{cases}$$

The above definition can be linearized with the following constraints:

$$S_{i,j}^k \leq \beta^k + L(1 - x_{i,j}^k) \qquad \forall (i,j) \in \varepsilon, k = 1, 2, ..., K \tag{4.48}$$

$$S_{i,j}^k \leq L x_{i,j}^k \qquad \forall (i,j) \in \varepsilon, k = 1, 2, ..., K \tag{4.49}$$

$$S_{i,j}^k \geq \beta^k - L(1 - x_{i,j}^k) \qquad \forall (i,j) \in \varepsilon, k = 1, 2, ..., K \tag{4.50}$$

$$S_{i,j}^k \geq 0 \qquad \forall (i,j) \in \varepsilon, k = 1, 2, ..., K \tag{4.51}$$

Hence, the SINR constraint can be rewritten as follows:

$$P_t d_{i,j}^{-\alpha} x_{i,j}^k + L(1 - x_{i,j}^k) \geq \eta_0 H^k + \sum_{(a,b)\in\varepsilon, a\neq i} (P_t d_{a,j}^{-\alpha} S_{a,b}^k) \tag{4.52}$$

The Pricing problem for the alternate formulation is presented as follows:

PRICING PROBLEM

Objective:

$$\text{Minimize } \overline{cost} = 1 - C \times \sum_{(i,j)\in\varepsilon} u_{i,j}.a_{i,j}(= \sum_{k=1}^{K} x_{i,j}^k) \tag{4.53}$$

Subject to: Equations (4.6), (4.9), (4.10) and (4.43-4.52)

$$x_{i,j}^k = \{0,1\}, y_v^k = \{0,1\}$$

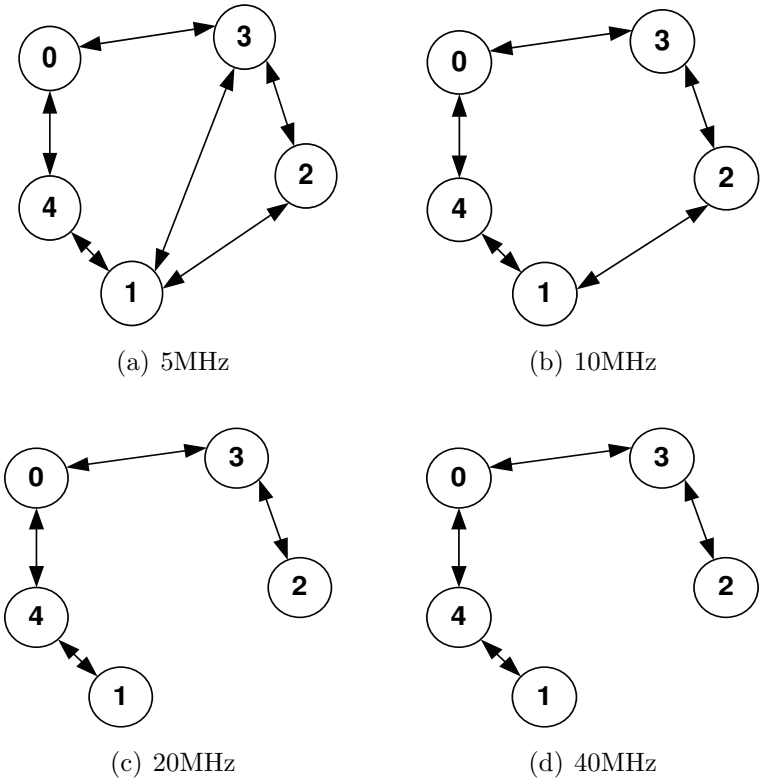
4.8 Numerical Results

We present numerical results to evaluate the two formulations presented earlier in Sections 4.4 and 4.7 and we refer to them as JRSVW1 and JRSVW2 (for Joint Routing and Scheduling with Variable Width). We also evaluate our enhanced model which considers optimal transmission rate selection, presented in Section 4.5. In our evaluation, we assume constant power (P_t = 10mWatt), unless mentioned otherwise. Path loss exponent is chosen as $\alpha = 2$ and $\eta_0 = 10^{-6}$Watt/MHz. The CG is implemented in C++ and solved using the CPLEX Concert Technology (version 9.1.3) [95].

4.8.1 Evaluation of JRSVW1

Evaluation on a small network

In our evaluation, the SINR threshold is chosen as $\beta = 1.3$ and the link spectral efficiency is fixed to $c = 1.206$ bps/Hz. We consider first a small network (5 nodes)



(a) 5MHz (b) 10MHz

(c) 20MHz (d) 40MHz

Figure 4.1: A 5-nodes network

shown in Fig. 4.1. As we mentioned earlier, the topology of the network depends on whether a link exists between any two nodes or not. A link exists if the receiver is within the transmission range of a transmitter. Fig. 4.1(a)-(d) shows the different topologies that result for different fixed channel widths (5-40 MHz). We consider three parallel multi-hop sessions, each corresponds to a particular s-d pair in the network. The three sessions are (3,1), (0,3) and (1,2) with traffic demands of (27.4085), (6.914) and (9.72211) in Mbits respectively randomly generated.

We present the numerical solutions of JRSVW1 for this 5-node network in Tables 4.4 and 4.5. Table 4.4 compares the performance of the fixed channel width allocation (5MHz and 40MHz are shown) to that of variable channel width adaptation model in terms of optimal routing. When the spectrum width is fixed to 5MHz, the resulting network topology is shown in Fig. 4.1(a) and Table 4.4 shows that all sessions are

Table 4.4: Routing paths and traffic flows in a 5-nodes network with 3-sessions

Ses.	Fixed 5MHz		Fixed 40MHz		Variable width	
	Routing Path	**Traffic (Mbits)**	**Routing Path**	**Traffic (Mbits)**	**Routing Path**	**Traffic (Mbits)**
1	(3, 1)	27.4085	(3, 0, 4, 1)	27.4085	(3, 0, 4,1)	27.4085
2	(0, 3)	6.9143	(0, 3)	6.9143	(0, 3)	6.9143
3	(1, 2)	9.72211	(1, 4, 0, 3, 2)	9.72211	(1, 4, 0, 3, 2)	0.760948
					(1, 2)	8.96117

Table 4.5: Configurations for a 5-nodes network with 3-sessions

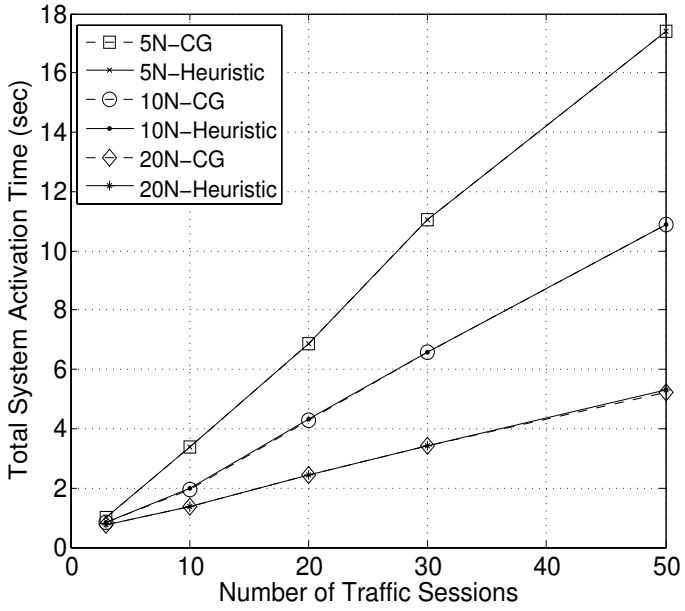
Channel Width	Conf.	Active Links	Link Act. Time (sec)	Sys. Act. Time (sec)
Fixed 5MHz	1	(1,2)	0.467333	6.18002
	2	(3,1)	4.56187	
	3	(0,3), (1,2)	1.15082	
Fixed 40MHz	1	(0,3)	0.143852	1.68886
	2	(0,4)	0.367965	
	3	(4,0)	0.202269	
	4	(0,3), (1,4)	0.202269	
	5	(3,0), (4,1)	0.570233	
	6	(0,4), (3,2)	0.202269	
Variable Width	1	(1, 4, 40 MHz) (3, 0, 40 MHz)	0.0158315	1.33181
	2	(3, 2, 40 MHz) (4, 1, 40 MHz)	0.0158315	
	3	(3, 0, 40 MHz) (4, 1, 40 MHz)	0.554402	
	4	(1, 2, 10 MHz) (4, 0, 40 MHz)	0.0158315	
	5	(0, 3, 40 MHz) (1, 2, 10 MHz)	0.0159684	
	6	(0, 4, 40 MHz) (1, 2, 10 MHz)	0.0570233	

routed through a one-hop path each. When the spectrum width is fixed to 40MHz, the transmission range gets shorter and the network is modified as shown in Fig. 4.1(d). Clearly, one-hop routing is not feasible for the sessions (3-1 and 1-2) as opposed to 5MHz bandwidth; however, the transport capacity of the links is increased due to the increase in the spectrum band allocation. As mentioned earlier, more hops along the route result in stronger intra-path interference. To mitigate the effect of this interference, links on successive hops must be scheduled at different times. Table 4.5 compares the performance in terms of total system activation time. Here, we can see that, the 40MHz case significantly outperform the 5MHz case (more than 3 times less). This is mainly due to the large capacity links provided by the 40MHz bands. Intra-path interference is mitigated by scheduling interfering links (along a single path) at different times and this is illustrated in the table where the optimal solution of 40MHz case yielded 6 different configurations where 5MHz case only yielded 3. Further, due to the single radio constraints, two successive links can not be scheduled simultaneously (e.g., links (3-0) and (0-4) along the route of session 3-1). The variable spectrum allocation selects a mixture of small and large bands available in the system to achieve the optimal solution. This is shown in Table 4.5 (variable width) where both 10MHz and 40MHz are used in the optimal solution. We can also see from Table 4.4 that multi-path routing on the same session is performed for optimal routing. For example, 0.760948 Mbits of session 3 are routed over a multi-hop path of larger transport capacity links, while a 8.96117 Mbits of the same session is routed over a single hop path (1-2) with 10MHz spectrum band. This shows that a mixture of large capacity short links with high concurrency and smaller capacity longer links are selected in the optimal solution. Therefore, longer high-capacity routes or shorter smaller-capacity routes do not result in best system throughput but rather there is a balance and the variable width adaptation model achieves this

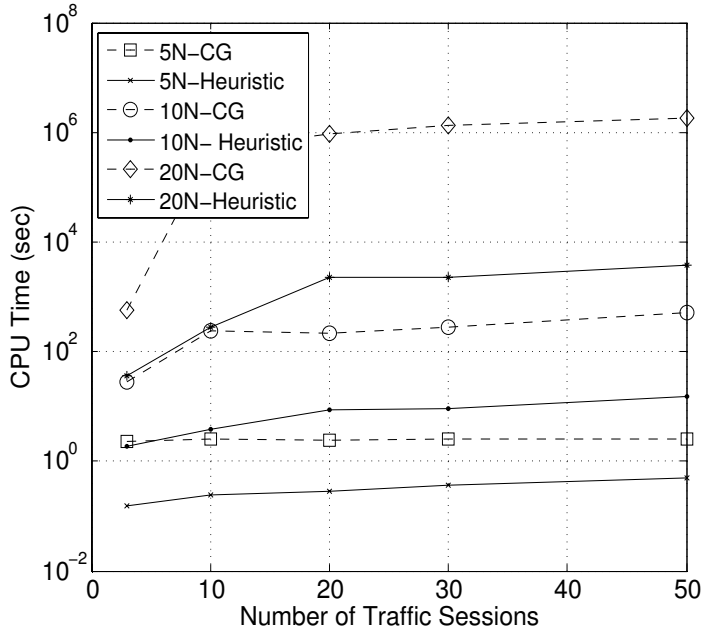
balance. The overall performance of variable spectrum-width allocation is superior to all other schemes, resulting in a system activation time of 1.33181(s) to deliver the demands; this is an improvement of 21.14% over the best of the fixed bandwidth schemes.

Evaluation of the Greedy pricing

We evaluate the results obtained by the greedy heuristic (presented in Section 4.6) for effectively solving the pricing subproblem and show that indeed the greedy yields results close to optimal. We consider four sets of randomly generated networks of 5, 10 and 20 nodes (5N, 10N and 20N) and different traffic instances (3-50 sessions). These networks are deployed over a $100m \times 100m$ area and each session m has a traffic demand randomly generated in the range of $0 < R_m \leq 35$Mbits. To minimize the influence of network topology, all the results presented in this section are averaged over four sets of results achieved from four sets of network topologies. Fig. 4.2(a) shows the system activation time obtained from JRSVW1 using both exact (CG) and greedy heuristic methods; we observe that constantly the greedy heuristic yields solutions very close to those obtained using the exact method, with a highest optimality gap being below 2%. The two methods however differ in their computation times, as shown in Fig. 4.2(b). As the size of the network and number of sessions increase, the exact model becomes unscalable with the greedy algorithm enjoying fast running times. For example, for a 20-nodes network with 50 sessions, the heuristic obtains the solution after 1 hour of CPU time while the exact pricing took close to 21 days of CPU time (i.e., more than 500× than that of greedy). Therefore, in the rest of the chapter, all the results from JRSVW1 are achieved using the heuristic model, unless otherwise stated.

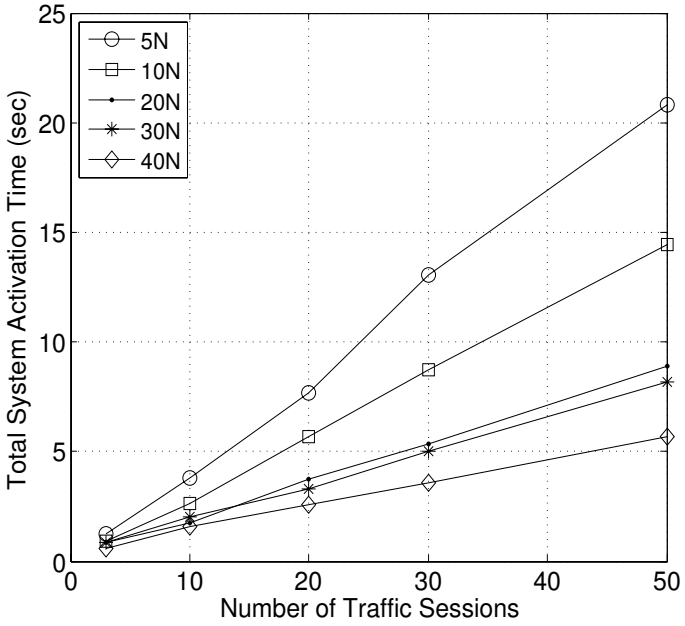


(a) Accuracy

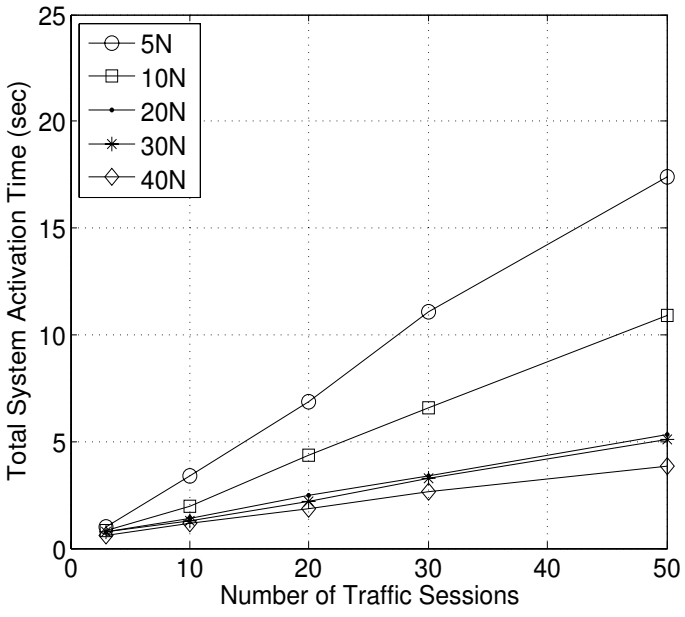


(b) Scalability

Figure 4.2: Comparison between exact and greedy heuristic model

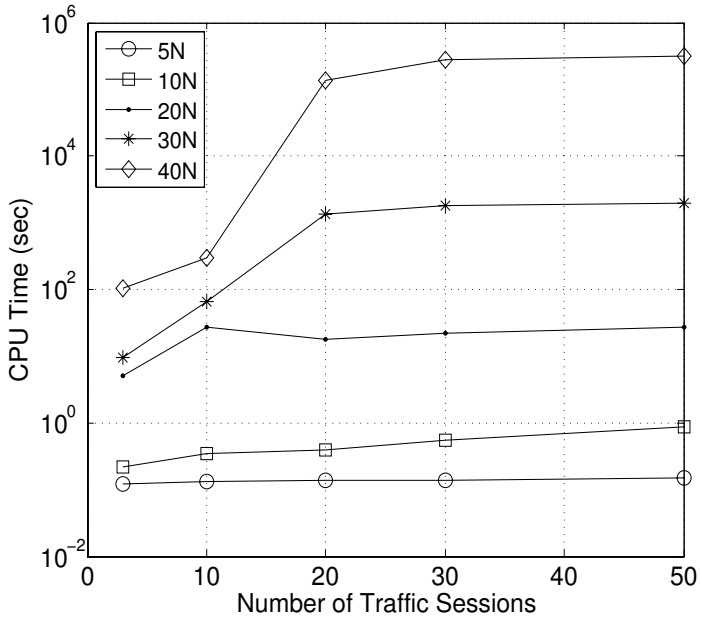


(a) Best fixed bandwidth (40MHz)

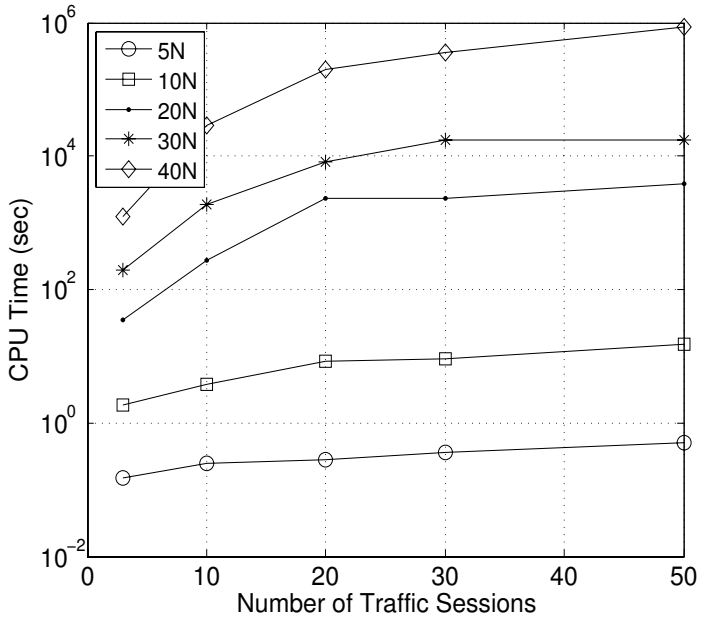


(b) Variable bandwidth

Figure 4.3: System activation time comparison between fixed and variable bandwidth



(a) Best fixed bandwidth (40MHz)



(b) Variable bandwidth

Figure 4.4: CPU time comparison between fixed and variable bandwidth

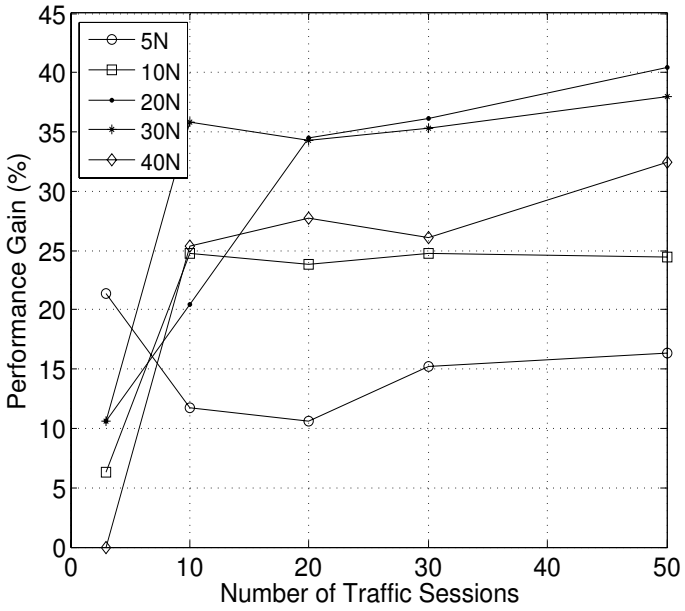


Figure 4.5: Variable bandwidth performance gain

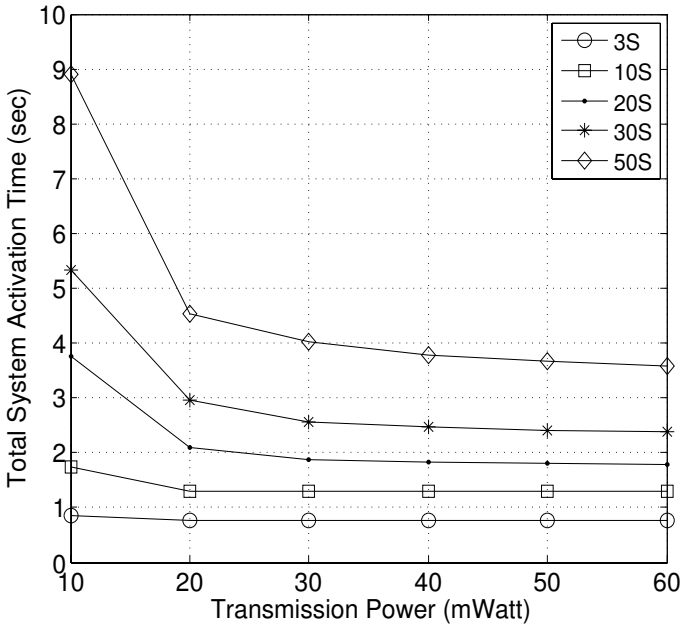
Evaluation of JRSVW1 on larger networks

We evaluate the performance of JRSVW1 on different networks with varying sizes (5N - 40N) and traffic loads (3S - 50S) (5N and 3S stand for 5 nodes and 3 sessions); note that results from each network size is averaged over 4 instances. We compare the performance of JRSVW1 with the best fixed-width spectrum allocation (in our case it is partitioning the spectrum into two channels of 40MHz each). The results are presented in Figs. 4.3, 4.4 and 4.5. We observe that although JRSVW1 requires more computation time (as expected) than the fixed band allocation, it results in much better performance, shown by the lower system activation time; this is due to its capability of better managing interference in the network by effective partitioning and allocation of the spectrum. This performance gain is illustrated in Fig. 4.5; although there is no straight forward pattern for the gain, we can see that the gain improves with the traffic load. The reason we observe smaller gains when the load is lighter is due to the fact that with small number of sessions, partitioning into

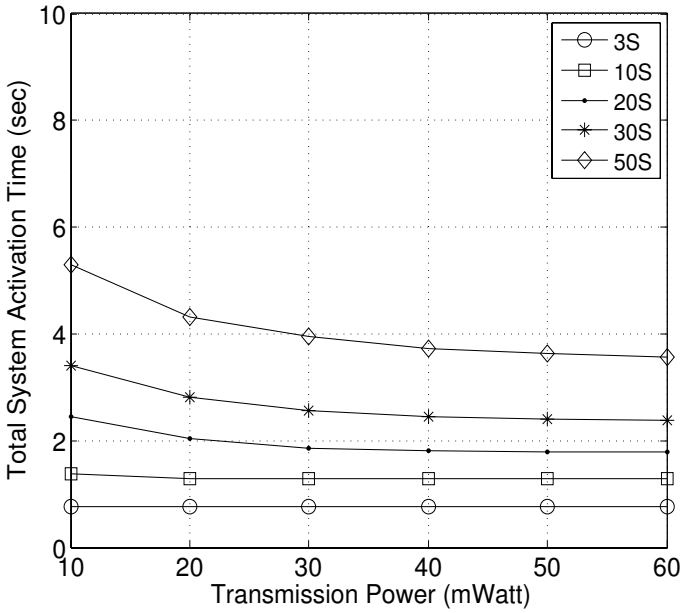
large spectrum blocks would yield optimal results because the interference is not a limiting factor (flows may be routed apart from each other to avoid interference) and there is not a strong contention on the available spectrum. For larger networks (and higher loads), the interference becomes a serious threat on the performance. Indeed, JRSVW1 effectively partitions the spectrum into channels of varying widths to be allocated to communication links to combat the interference and satisfy the traffic demands. Our results indicate, an average performance gain of 23.5% for JRSVW1 achieved over the joint routing and scheduling problem with best fixed spectrum partitioning and a maximum performance gain of 40.4% in terms of system activation time.

Impact of transmission power

We consider a network of 20 nodes and vary the number of sessions between 3 and 50 (3S - 50S); the transmission power varies between 10mW to 60mW. We compare the performance of JRSVW1 with best fixed band allocation. We observe that as the power increases, the system activation time decreases (i.e., better performance) (Figs. 4.6(a) and 4.6(b)) and the computation time increases (Figs. 4.7(a) and 4.7(b)). Clearly, as the transmission power increases, the transmission range increases and therefore the network connectivity increases as well. This results in denser networks and hence more links. The increase in the number of links results in increase in the number of link integer variables in the pricing subproblem and therefore an increase in the computation time to solve the model (as the figure shows). On the other hand, more links in the network give the optimizer freedom to schedule more links in the same time-slots and thus increase the number of concurrent transmissions that satisfy the SINR constraint and improve the system performance. Fig. 4.8(a) shows (for a fixed 40MHz 20N-50S System) that when transmission power increases

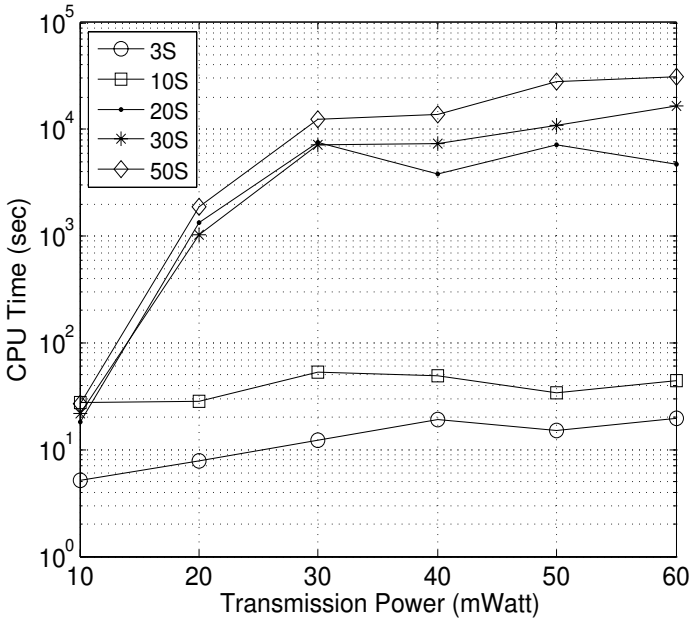


(a) Best fixed bandwidth (40MHz)

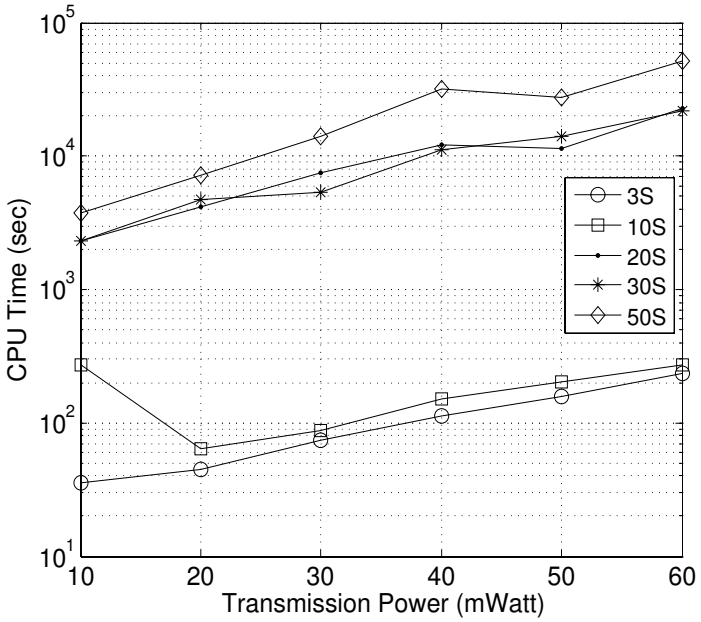


(b) Variable bandwidth

Figure 4.6: System activation time with variable transmission power



(a) Best fixed bandwidth (40MHz)



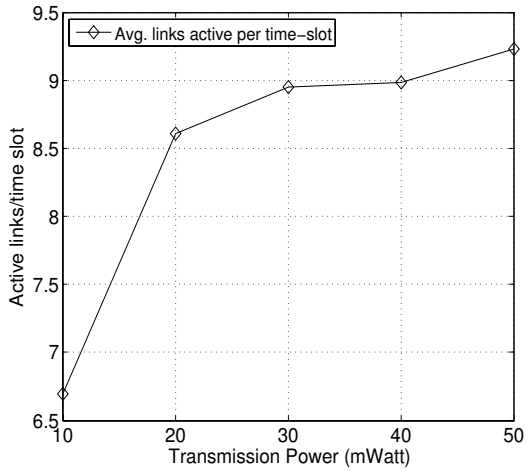
(b) Variable bandwidth

Figure 4.7: CPU time with variable transmission power

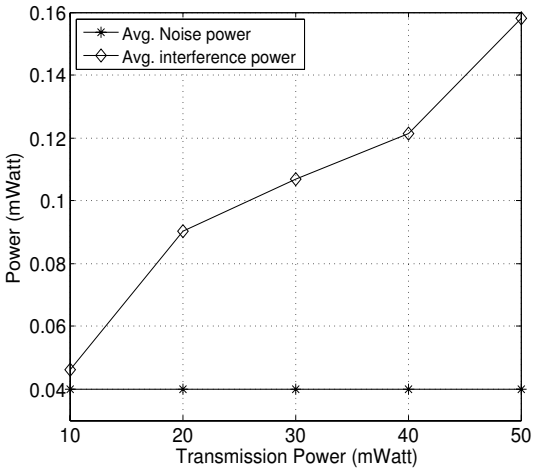
the average number of active links per time-slot also increases. Note that, we have considered a single radio system (i.e., radio constraint), so for a 20-nodes network no more than 10 links can be active at a time in a single time-slot. As a result of that, as we further increase the transmit power, there is no further enhancement in the system performance (Figs. 4.6(a), 4.6(b) and 4.9) because most of the time-slots already have 10 links concurrently active in them. Although radio constraint is one of the main limiting factor, it is not the only one. We can see from Fig. 4.8(b) that as transmission power increases the interference power is getting more dominant over the noise power and while the number of concurrently active links per time slot increase the average SINR per link decreases (Fig. 4.8(c)). Therefore, even when the radio constraint allows, it is not always feasible to add more concurrently active links in the same time-slot without violating the SINR constraint.

Evaluation of JRSVW1 with transmission rate adaptation

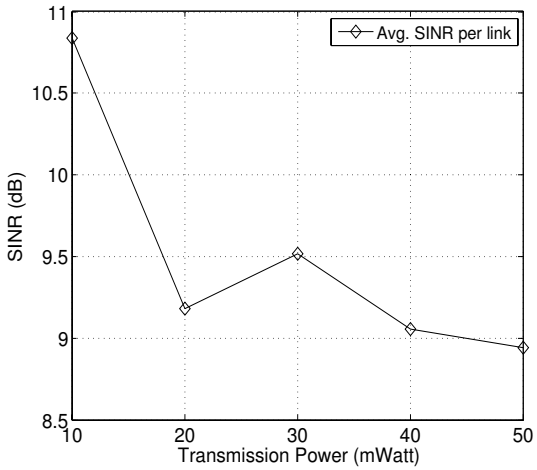
We evaluate the performance of the model presented in Section 4.5 where in addition to varying the channel width, we also vary the transmission data rate (VWVR). The mapping between transmission rates and SINR threshold is shown in Table 4.2. We consider networks of 5, 10 and 20 nodes and vary the traffic load between 3 and 40 sessions and each of the results presented is averaged over four instances. The base of our comparison is the model where both the spectrum width and the transmission rates are fixed (FWFR) and we select the one that yields best results to compare with. We also consider two other models where in one the spectrum width is variable but transmission rate is fixed (VWFR) and in the other one the spectrum width is fixed and the rate is variable (FWVR). We first consider a 20 nodes network with traffic load of 20 sessions and we evaluate (Fig. 4.10) the performance of FWFR, for different spectrum widths and different transmission rates, to determine the best



(a) Transmission power vs. Avg. active links per time-slot



(b) Comparison between noise and interference power



(c) Transmission power vs. Avg. SINR per link

Figure 4.8: Variable transmission power effect on network

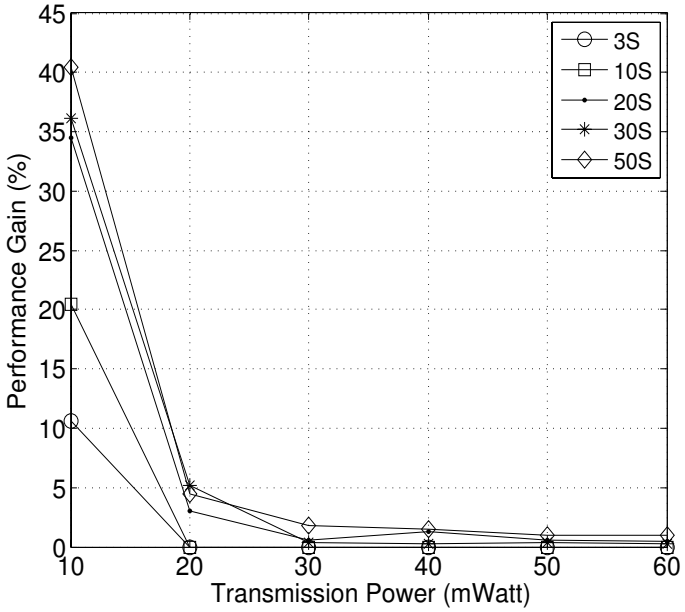


Figure 4.9: Variable bandwidth performance gain with variable transmission power

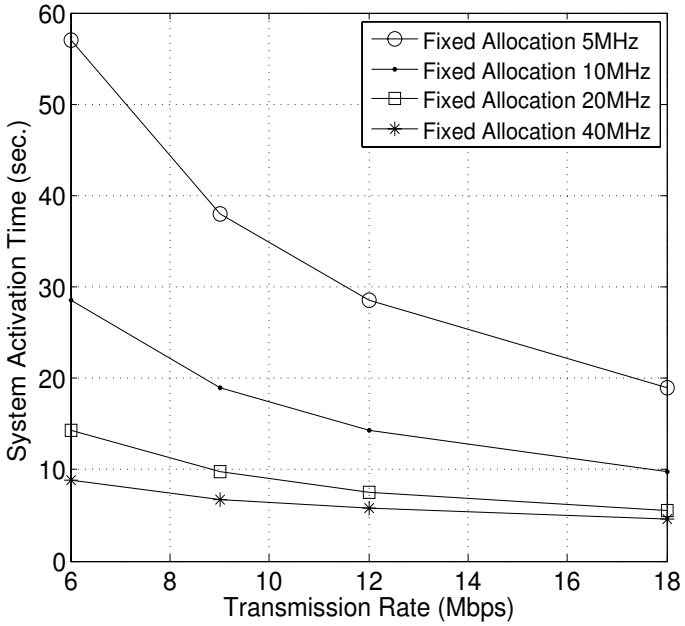
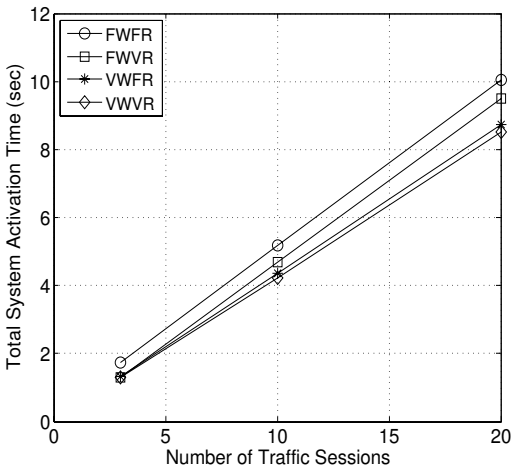
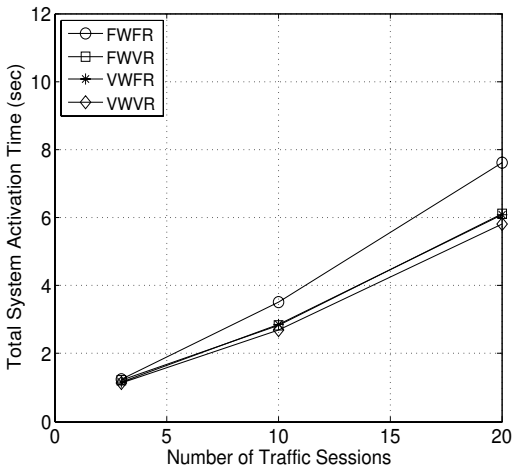


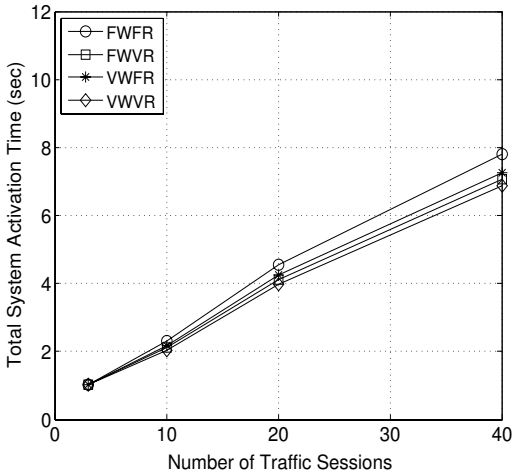
Figure 4.10: FWFR with different spectrum widths and transmission rates



(a) 5-nodes network

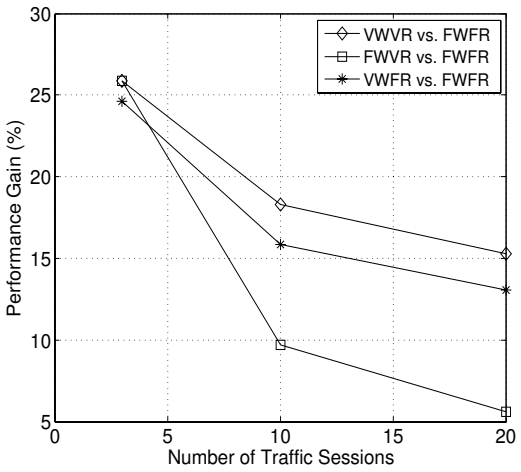


(b) 10-nodes network

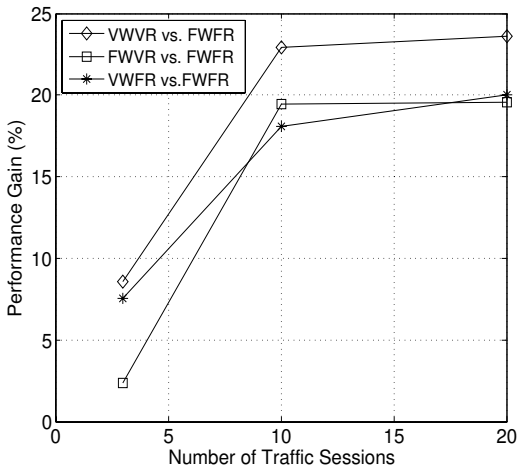


(c) 20-nodes network

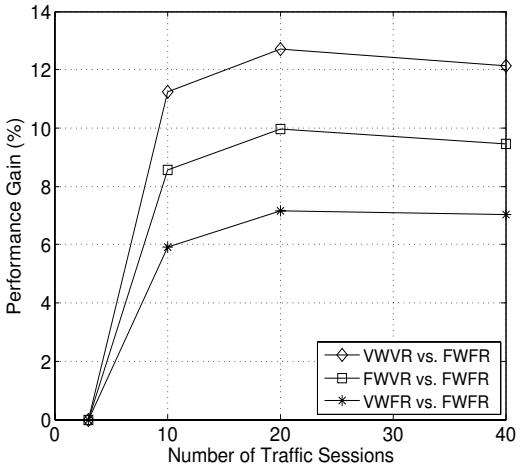
Figure 4.11: System activation time with variable bandwidth and rate adaptation



(a) 5-nodes network



(b) 10-nodes network



(c) 20-nodes network

Figure 4.12: Variable bandwidth and rate adaptation performance gain

(width-rate) combination to operate our base model FWFR. Clearly, FWFR with 40MHz and 18Mbps yielded the best results and will be used as the base model in the rest of our evaluation.

Figs. 4.11(a)-(c) show the system activation vs. traffic load for the 4 different models and different networks we consider and Figs. 4.12(a)-(c) depict the performance gain the three models achieve over the base model (FWFR). Observe that transmission rate adaptation and spectrum width adaptation are both shown to be effective mechanisms for managing interference and effectively sharing the spectrum resource among concurrent transmissions. Both FWVR and VWFR achieve higher performance than FWFR, with VWVR outperforming all models due to the higher degree of freedom (2 dimensions) in adapting the proper parameters. This indeed suggests the benefits of developing corresponding MAC protocols with capability of jointly tuning transmission rate and spectrum width to improve the transmission quality.

4.8.2 Evaluation of Alternative Variable Width Adaptation Model

In the first scenario, we consider 2 different networks, the first consists of 6 nodes (3 single hope flows) and the other consists of 14 nodes (7 single hop flows) both deployed over a $25m \times 25m$ area. Table 4.6 shows the performance results for the different design methods. We can see a clear contrast between these results and those obtained in Section 4.8.1. Here, the 5MHz fixed channel width model outperforms the 40MHz channel width model in both network instances. Recall that a smaller width partitions the spectrum into more orthogonal channels that may be used concurrently. Considering a spectrum of 80MHz, we have a total to 16 channels of 5MHz width and 2 channels of 40MHz width, each providing the same transmission capacity but with

Table 4.6: System activation time in scenario 1

Number of Nodes	**Total System Activation Time (in sec)**		
	Variable width	**Fixed 5MHz**	**Fixed 40MHz**
6	2.94066	2.94066	3.53092
14	3.42802	3.42802	4.1423

Table 4.7: System activation time in scenario 2

Number of Nodes	**Total System Activation Time (in sec)**				
	Variable Width	**Fixed 5MHz**	**Fixed 10MHz**	**Fixed 20MHz**	**Fixed 40MHz**
6	3.23882	6.14107	6.099	3.23882	4.6058
14	3.19077	3.50789	3.34113	3.19077	3.64105

different SINR requirement. With 5MHz fixed channel width, all single hop flows (in both network instances) can be active and scheduled concurrently; alternatively with 40MHz widths, only 2 channels are available and hence the flows must be scheduled at different times, leading to a longer schedule length. We should note however, that although only two 40MHz channels are available, the SINR requirement for transmission on each such channel is much smaller (derived in Table 4.3), which means the same channel may be reused on different links with enough spatial separation. The alternative variable width adaptation model dynamically searches for the optimal spectrum allocation and yields a minimal system activation time (similar to that obtained with 5MHz) with a 16.7% (17.24%) reduction in system activation time for the 6-node (14-node) networks over fixed 40MHz spectrum allocation.

In scenario 2, we consider both 6-node and 14-node networks over a $34m \times 34m$ area. For a 20MHz and 40MHz channel widths, all nodes are within transmissions range of each other, as can be seen from Table 4.3; however, with 5 and 10MHz widths, some sessions may fall outside each other's transmission range and require intermediate forwarding. Clearly, in addition to inter-path interference, with multi-hop routing, intra-path interference becomes also a limiting factor. The numerical

Table 4.8: System activation time in scenario 3

Number of Nodes	**Total System Activation Time (in sec)**		
	Variable width	**Fixed 5MHz**	**Fixed 40MHz**
10	8.71983	9.84742	8.71983

Table 4.9: System activation time in scenario 4

Number of Nodes	**Total System Activation Time (in sec)**				
	Variable Width	**Fixed 5MHz**	**Fixed 10MHz**	**Fixed 20MHz**	**Fixed 40MHz**
6	3.23882	6.14107	6.099	4.6058	4.6058

results are presented in Table 4.7. We see that (for the 6-nodes network), both 5 and 10MHz are out performed by the other design methods; this is mainly due to the fact that with these widths, flows are delivered to their destinations through multi-hop routes and links on successive hops along a path are scheduled separately (the single radio constraint), although more orthogonal channels are available. With 40MHz, only two orthogonal channels are available and hence all three flows cannot be concurrently active leading to a larger schedule length than 20MHz (where all three flows are active concurrently). We observe that our optimal selection corresponds to that obtained with 20MHz channel width. Similar observations can be seen for the 14-nodes network as shown in Table 4.7 with 40MHz being outperformed by the other schemes, due to the limit in the available channels.

We now consider a 10-node random network deployed over a $50m \times 50m$ area with 10 randomly generated sessions. We present the numerical results obtained from fixed 5MHz, 40MHz and variable width allocation design in Table 4.8. The network is not fully connected for any of the channel widths; both 40MHz and alternative model achieve optimal schedule length. Here, wider widths have longer reach and hence the sessions are more likely to either be delivered on single hop routes (if possible) or on routes with smaller hops (than those selected by the 5MHz); the

performance of the 5MHz width allocation in hence more affected by the intra-path interference. In addition, since the SINR requirement for 40MHz channel is much smaller, inter-path interference is not a major issue and the same channel can be reused on multiple links concurrently. Finally we consider the same 6-node network of scenario 2 with 3 single hope sessions and change the distance of one of the links such that the destination could not be reached by the source directly with a 20MHz channel but it can be reached in a single hop by using 40MHz channel. We define this scenario as scenario 4. Here, our model achieves the best performance; with 40MHz width, only 2 orthogonal channels are available and hence all three flows cannot be concurrently active. With 20MHz, one flow requires multi hop routing and hence the intra-path interference limits the performance. Our alternative design approach for variable width adaptation uses mixture of both 20MHz and 40MHz widths to yield an optimal schedule length of 3.23882 (s). A significant reduction of almost 30% (over the best fixed scheme) in the schedule length.

4.9 Conclusion

In this chapter, we presented a cross-layer formulation for the joint routing, scheduling and spectrum allocation problem in multi-channel multi-rate wireless mesh networks. While narrower spectrum widths divide the total available spectrum into more non-overlapping channels allowing more parallel concurrent transmissions, wider bands have the effect, to either increase the transport capacity per link or reduce the SINR requirement to achieve the same link capacity. We presented two different formulations for solving the same problem; in the first one, we assume the link capacity as a function of channel bandwidth and in the second one, we assume fixed link capacity where the SINR threshold is a function of the channel bandwidth. Contrasting results are obtained from each design method with predetermined channel widths; while in

the first design, smaller widths always yield worse performance (due to the small achievable link capacities), in the second design, the performance depends on the network and traffic scenarios under study. In all studied cases, however, the variable channel-width allocation achieved optimal performance; this is due to the capability of the model in achieving optimal partitioning to the spectrum which simultaneously maximizes spatial reuse and minimizes intra-path interference. Our work shows that larger system throughput can be achieved when both the spectrum width and data rate are considered as adaptable, rather than fixed, control parameters in network design.

Chapter 5

Optimal Flexible Spectrum Allocation

In this chapter we consider a more flexible optimal spectrum allocation technique. Unlike the previous chapter, here, we do not impose optimal partitioning of the available spectrum band into a set of non-overlapping channels rather we let the cross-layer design decide on the channel bandwidth positions. In this way a more flexible allocation of bandwidth is possible, since the transmissions may use overlapping channels. When considering spectrum overlap, the design problem gets further complicated, and adjacent channel interference must be dealt with properly. The Interference factor (I-factor) captures the amount of overlap between a transmitting and an overlapping interfering channel. This I-factor may not be predefined, but rather it is jointly determined when performing channel assignment as it depends on the portion of overlap between two channels.

We again adopt the widely used cross-layer optimization to formulate the joint problem of routing, optimal link scheduling and spectrum assignment and we use a realistic model for characterizing the interference. Under this (centralized) approach, link scheduling determines the set of transmission links which can be concurrently

active without violating the interference constraints. Under optimal scheduling, one needs to allocate spectrum on active links to satisfy the traffic demands. In this work, the spectrum allocation is done by properly finding the location and the size of the spectrum block assigned to each active link.

Similar to previous chapter, the set of active links together with their allocated spectrum blocks are referred to as a transmission configuration and the joint optimization can then be obtained by constructing the whole set of such configurations. We model mathematically this combinatorial complex problem which requires the enumeration of all configurations, and we present a decomposition method, based on column generation, for solving it without exhaustive enumerations. We show that this cross layer design is computationally very complex to solve, except for very small network instances. The complexity arises as a result of the mathematical function characterizing the physical interference as well as the large combinatorial nature of the problem. To circumvent the first difficulty, researchers have adopted a more simplified, but scalable, interference model commonly known as the protocol model. This simplified model is shown to underestimate the interference in the network and thus results in schedules which may not be feasible under physical interference constraints. The difficulty arising from the second problem is attributed to the fact that the pricing subproblem deals with a large number of integer variables and usually is quite hard (ILP) to solve. To overcome this problem, we propose a heuristic based on simulated annealing (SA) to solve the pricing subproblem. To overcome the first problem, we assume the simplified protocol model but we only allow our SA to generate configurations which are feasible under physical interference constraints.

The rest of the chapter is organized as follows: System model is presented in Section 5.1. In Section 5.2 we present the problem formulation and its decomposition in Section 5.3. In Section 5.4 we present an illustrative example. Section 5.5 presents

the simulated annealing based heuristic model for solving the pricing. Numerical results and discussions are given in Section 5.6 and conclusions in Section 5.7.

5.1 System Model

5.1.1 Network Model

We consider a network with a set of $V = \{v_1, v_2, ..., v_N\}$ nodes($|V| = N$); the Euclidean distance between two nodes v_i and v_j is denoted by d_{ij}. We assume each node is equipped with a single frequency-agile radio and can dynamically configure the channel center frequency and width. These parameters can be adjusted within 10s of microseconds on commodity (WiFi) hardware [52]. We assume a total accessible target spectrum band of BMHz, which may be partitioned into several variable-width contiguous blocks; due to hardware limitations [14], the bandwidth values are assumed to be discrete from the set $W = \{5, 10, 20, 40\}$MHz, and each spectrum block is characterized by its start and end frequencies. Let E denote the set of links in the network. A link $l = (v_i, v_j) \in E$ iff v_j can successfully receive and decode a signal from v_i. In the absence of any interference, the signal to noise ratio (SNR) at the intended receiver is:

$$SNR_{ij}^{k} = \frac{P_t g_{ij}}{\eta_0 w_k} \tag{5.1}$$

where, P_t is the sender transmission power, g_{ij} is the channel propagation gain ($g_{ij} = g(d_{ij})$), η_0 is the power spectral density of the thermal noise and w_k is the width of the spectrum block allocated on link (v_i, v_j). A transmission is thus successful iff the SNR at the receiving node exceeds a certain threshold β and we say that v_j is within the transmission range of v_i. Therefore, a link $l = (v_i, v_j) \in E$ iff v_j is within the transmission range of v_i. Let T_k denote the transmission range when node v_i

transmits using a power P_t and a radio spectrum of width w_k. Setting $SNR = \beta$, we obtain:

$$g(T_k) = (\frac{P_t}{\eta_0 w_k \beta})^{-1} \quad \text{or} \quad T_k = g^{-1}((\frac{P_t}{\eta_0 w_k \beta})^{-1}) \tag{5.2}$$

where $g(.)$ is the channel gain function. For the widely used uniform channel model, $g_{ij} = d_{ij}^{-\alpha}$ where α is the path loss exponent; then, $T_k = (\frac{P_t}{\eta_0 w_k \beta})^{\frac{1}{\alpha}}$. For a fixed P_t, η_0, and β, the transmission range T_k depends on the spectrum width w_k; a wider channel spectrum implies a shorter transmission range and a smaller spectrum yields a larger range. Thus, Eq. (5.1) shows that for a particular radio link (v_i, v_j) and a fixed g_{ij}, when a smaller spectrum band is allocated, a higher SNR is observed at the receiver and when a wider spectrum is used, a lower SNR is observed. It is to be noted, however, that the wider bandwidth results in higher link capacity. Since the transmission range of a node depends on the allocated spectrum, the existence of a communication link between two nodes depends, in addition to the Euclidean distance, on the size of the spectrum allocated to the link. That is, while v_j may not be in the transmission range of node v_i for spectrum band w_{k_1}, both nodes may communicate using a spectrum band $w_{k_2} < w_{k_1}$.

Let E_k denote the set of radio links in the network corresponding to spectrum band w_k; thus, G_k is a graph representing the radio connectivity in the network, $G_k = (V, E_k)$. A link $l = (v_i, v_j) \in E_k$ iff v_j is within the transmission range of v_i when the radio link is assigned a spectrum w_k. We now define E as the set of all possible radio links, i.e., $E = \cup_{w_k \in W} E_k$, and $G = (V, E)$. G is therefore a multi-layer graph, each layer is a subgraph G_k representing the connectivity in the network corresponding to a particular spectrum width w_k (we note that when the size of a spectrum block takes *any* arbitrary value in the permissible band (e.g., [53]), then modeling the system using graph G would not be feasible). The connectivity in each subgraph is determined according to the transmission range given by Eq. (5.2) and

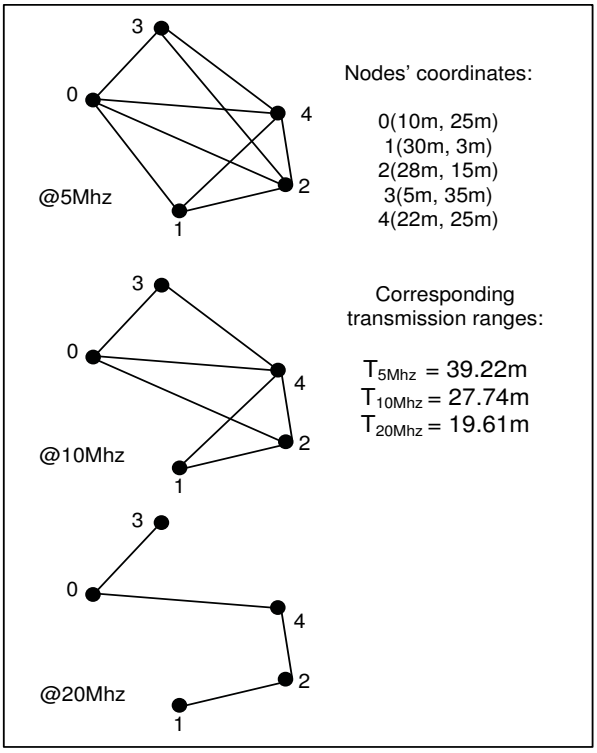


Figure 5.1: Multi-layer graph representing radio connectivity at different channel widths: 5MHz (up) 10MHz (middle) 20MHz (bottom)

the Euclidean distances between each pair of nodes. A link $\ell \in E$ then corresponds to a pair of nodes v_i and v_j within transmission range of each other *and* a particular spectrum block b_ℓ ($b_\ell = w_k \in W$). b_ℓ is characterized by its start and end frequencies 's_ℓ' and 'e_ℓ'.

Fig. 5.1 shows the multi-layer graph representing a network of 5 nodes, each layer represents the connectivity in the network for a particular spectrum width (5, 10 and 20MHz). We assume $\alpha = 2$, $\eta_0 = 10^{-6}$ watt/MHz, $P_t = 1mW$, $\beta = 1.3$; for different channel widths, we obtain the following transmission ranges: $T_5 = 39.22m$, $T_{10} = 27.74m$, $T_{20} = 19.61m$. Fig. 5.1(up) shows the directed graph corresponding to 5MHz width; notice here that $d_{1,3} > T_5$ and thus node 3 is outside the transmission range of node 1 (and vice-versa). $d_{1,4} \leq T_5$, and nodes 1 and 4 are within transmission range of each other. For $w_k = 20$MHz (Fig. 5.1(bottom)), $T_{20} = 19.61m$, hence

$d_{1,4} > T_{20}$ and nodes 1 and 4 fall outside each other's transmission range.

5.1.2 Interference Model

In the presence of concurrent transmission on neighboring links, a transmission may be corrupted as a result of strong interference caused by active links on the overlapping portions of the spectrum. Considering the cumulative effects of interference, a transmission is successful if the $SINR$ at the intended receiver is above a certain threshold:

$$SINR_\ell = \frac{P_t \cdot g_\ell}{\eta_0 b_\ell + \sum\limits_{t(\ell') \neq t(\ell), r(\ell)} P_t \cdot g_{(t(\ell'), r(\ell))} \cdot I_{(\ell,\ell')}} \geq \beta \tag{5.3}$$

where, $t(\ell)$, $t(\ell')$ and $r(\ell)$, $r(\ell')$ are the transmitters and receivers of links ℓ and ℓ' respectively. $I_{(\ell,\ell')}$ ($I_{(\ell,\ell')} \in [0,1]$) is the normalized interference factor (I-factor) [48] which captures the amount of overlap between a transmission (on link ℓ') on a certain spectrum block and reception on link ℓ. This model, which takes into account cumulative interference from all other links is widely known as ***physical interference model***. The value of $I_{(\ell,\ell')}$ depends on the spectrum assignment on links ℓ and ℓ', which is dependent on the values of s_ℓ, e_ℓ, $s_{\ell'}$ and $e_{\ell'}$. The amount of overlap is captured quantitatively by calculating the area of intersection between interfering signal's spectrum $S_{t(\ell'),b_{\ell'}}$ and receiver's band-pass filter $B_{r(\ell),b_\ell}$ [48, 96].

$$IF_{\ell,\ell'}(\tau) = \int_{-\infty}^{\infty} S_{t(\ell'),b_{\ell'}}(f) B_{r(\ell),b_\ell}(f-\tau) df \tag{5.4}$$

Here, τ is the difference between the center frequencies of the channels used on links ℓ and ℓ'. $S_{t(\ell'),b_{\ell'}}(f)$ denotes the power distribution of the interfering signal (on interfering link ℓ') across the frequency spectrum and $B_{r(\ell),b_\ell}(f)$ represents the frequency response of the band-pass filter at the receiver of link ℓ. Similar to [48, 96], we approximate the signal power distribution with the corresponding transmit

spectrum mask; the normalized I-factor between the spectrum of interfering signal on ℓ' and the signal of interest on ℓ can be written as:

$$I_{(\ell,\ell')} = \frac{\Lambda_{\ell,\ell'}}{\Lambda_{\ell'}} \tag{5.5}$$

where $\Lambda_{\ell,\ell'}$ is the area of overlap between the spectrum of the interfering signal and the frequency response of the band pass filter at the receiver of ℓ and $\Lambda_{\ell'}$ is the area under the interfering signal spectrum. Thus, $I_{(\ell,\ell')}$ captures the fraction of power of the interfering signal which affects the transmission on link ℓ.

According to the spectrum assignment (both size and location), a signal on link ℓ' may either completely or partially interfere with a transmission on link ℓ or may cause no interference if the assigned spectrums have no overlap. Thus, depending on the values of s_ℓ, e_ℓ, $s_{\ell'}$ and $e_{\ell'}$, 6 different scenarios may be distinguished to determine the values of the normalized I-factor, as shown in Table 5.2 [1].

Note that the value of $I_{(\ell,\ell')}$ depends on the spectrum assignment on links ℓ and ℓ' (s_ℓ, e_ℓ, $s_{\ell'}$, and $e_{\ell'}$), which is not predetermined; rather in our work, it is jointly determined when performing link scheduling to achieve better interference management and reduce the impact of spectrum fragmentation. This makes the computation of the SINR quite difficult to obtain, particulary when dealing with cross-layer optimization, making the physical interference model only practical for very small network instances [97].

To overcome the scalability problems associated with the SINR model, a more simplified and widely used approach for characterizing the interference is known as the ***protocol interference model***; this model describes interference constraints according to a conflict graph, where a pair of nodes within transmission range can

[1] One other way to look at the I-factor is to consider the ratio of the spectral density of the overlapped spectrums to the spectral density of the interfering signal's spectrum.

successfully communicate as long as the receiver is separated by a distance $IR_k = T_k(1+\delta)$ from any active transmitter on a conflicting spectrum band; spectrum bands on two different links are said to conflict if they completely or partially overlap. IR_k is referred to as the interference range and δ is a small constant. The conflict graph is denoted by $G_{CC} = (V_{CC}, E_{CC})$, where V_{CC} is the set of vertices corresponding to all links in E and E_{CC} is the set of edges. Two vertices in V_{CC} are connected by an edge if the corresponding links in E falls in each others interference range. Let $F_{CC}(= [f_{CC}(\ell, \ell')])$ denote the adjacency matrix of G_{CC}; $f_{CC}(\ell, \ell') = 1$ if links ℓ and ℓ' do not interfere and 0 otherwise.

The capacity of the wireless link ℓ (with spectrum width b_ℓ) is a function of the $SINR_\ell$; we assume each link can be viewed as a single user Gaussian channel, and the Shannon Capacity of the link (ℓ) is given by

$$C_\ell = b_\ell \log_2(1 + SINR_\ell) \tag{5.6}$$

We assume a fixed transmission rate on all links utilizing the same spectrum width, even when the $SINR$ of a particular link exceeds the threshold β [89]. Thus, C_ℓ simplifies to the following lower-bound:

$$C_\ell = b_\ell \log_2(1 + \beta) \tag{5.7}$$

5.2 Problem Formulation

We will formulate the joint routing, scheduling and spectrum band allocation as an optimization problem. We consider M concurrent multi-hop sessions, each corresponds to a source-destination pair (t_m,r_m) in the network. The traffic demand for

Table 5.1: A list of all parameters and variables

Parameter/Variable	Definition
P_t	Transmission power
η_0	Background Noise
g_{ij}	Channel Gain
w_k	Width of spectrum block
T_k	Transmission range
d_{ij}	Euclidean Distance
α	Path loss exponent
β	SINR Threshold
s_ℓ / e_ℓ	start/end frequency of spectrum block
b_ℓ	Bandwidth of spectrum block
B	Total permissible spectrum
$IF_{\ell,\ell'}$	I-Factor
$I_{\ell,\ell'}$	Normalized I-factor
IR	Interference Range
C_ℓ	Shannon capacity
x_ℓ^p	Link binary variable
p	Configuration
$C_{\ell,\ell'}^i$	Binary variable indicator for Scenario i
f_{CR}/f_{CC}	Binary parameter for Radio/Channel constraint
L	Large positive constant
$y_{\ell,\ell'}$	Decision variable for spectrum assignment
λ_p	activation time for configuration p
f_ℓ^m	Flow variable
R_m	Bit rate for session m
u_ℓ	Dual variable
p_{curr}	Current configuration
p_{init}	Initial configuration
$Cost_{curr}$	Cost of current configuration
OL/IL	Outer/Inner loop parameter for SA
Pr	Selection probability parameter for SA

each session m ($1 \leq m \leq M$) is given by R_m (bits). Traffic of a particular session may be split to sub-flows routed over different paths. The choice of these routing paths depends on the underlying schedule of different concurrent transmissions. The objective of our model is to minimize the system activation time for delivering the M sessions without violating the interference constraints. We assume a TDMA access scheme where time is divided into slots and a link may be active in one or more time slots to meet the traffic requirement. We further assume a total spectrum width of B(MHz) and links ($\ell \in E$) assigned spectrum blocks (channel) of widths ($b_\ell \in W$) with s_ℓ and e_ℓ being the start and end frequencies of each allocated block.

5.2.1 Scheduling and Spectrum Assignment

Recall from the multi-layer graph concept introduced earlier that the term link is used to identify a pair of adjacent nodes and the spectrum band b_ℓ. If a radio link between two nodes is active, then the spectrum is known (therefore b_ℓ is a parameter rather than a variable); what indeed remains to be determined is the location of the spectrum block (s_ℓ and e_ℓ) and the time during which a link ℓ should be active. Hence, a transmission configuration (p) is defined as the set of links, and their corresponding spectrum allocation, which may be active concurrently. We introduce link binary variable, x_ℓ^p, which is defined as follows:

$$x_\ell^p = \begin{cases} 1 & \text{if link } \ell \text{ (with spectrum width } b_\ell) \text{ is active in } p \\ 0 & \text{Otherwise} \end{cases}$$

Since a node can either transmit or receive at a time, the radio conflict constraint is:

$$x_\ell^p + x_{\ell'}^p \leq 1 + f_{CR}(\ell, \ell') \qquad \forall (\ell, \ell') \in E \quad \ell \neq \ell' \tag{5.8}$$

where,

$$f_{CR}(\ell, \ell') = \begin{cases} 1 & \text{if links } \ell \text{ and } \ell' \text{ do not have any common nodes} \\ 0 & \text{Otherwise} \end{cases}$$

Eq. (5.8) assures that only links with no common radio may be active simultaneously. When link ℓ is active, the start and end frequencies of its allocated spectrum are related by the following:

$$b_\ell x_\ell^p = e_\ell - s_\ell \qquad \forall \ell \in E \tag{5.9}$$

where, e_ℓ and s_ℓ have integer values. The following constraint ensures that any spectrum allocation to a link ℓ must fall inside the permissible spectrum band B.

$$e_\ell \leq B \times x_\ell^p \qquad \forall \ell \in E \tag{5.10}$$

Interference constraints for Physical Model

A transmission is successful if the SINR at the intended receiver is above a threshold (β):

$$P_t d_{t(\ell),r(\ell)}^{-\alpha} x_\ell^p + L(1 - x_\ell^p) \geq \beta(\eta_0 b_\ell + \sum_{t(\ell') \neq t(\ell), r(\ell)} (P_t d_{t(\ell'),r(\ell)}^{-\alpha} I_{\ell\ell'})) \quad \forall \ell \in E \tag{5.11}$$

Eq. (5.11) is enforced when $x_\ell^p = 1$ and redundant when $x_\ell^p = 0$; Eq. (5.12) forces $I_{\ell\ell'}$ to 0 when $x_{\ell'}^p = 0$.

$$I_{\ell\ell'} \leq x_{\ell'}^p \qquad \forall \ell' \in E \quad \ell' \neq \ell \tag{5.12}$$

We present the different constraints for normalized I-factor calculation in Table 5.2. Table 5.2 contains all the constraints which model the 6 scenarios ($S1 - S6$) we mentioned earlier. Please note that according to the assignment, only one of the scenarios is chosen:

$$\sum_{i=1}^{6} C_{\ell\ell'}^{i} = 1 \qquad \forall \ell \in E \;\; \forall \ell' \in E \;\; \ell \neq \ell' \tag{5.13}$$

$e_\ell \in \{0, 1, ..., B\}$, $s_\ell \in \{0, 1, ..., B\}$. $C_{\ell\ell'}^{i} = \{0, 1\}$ indicates if scenario i governs transmissions on link ℓ and interfering link ℓ' or not. $x_\ell^p \in \{0, 1\}$, $0 \leq I_{\ell\ell'} \leq 1$, L is a large positive constant. Note that all variables (e.g., s_ℓ, e_ℓ, $I_{\ell\ell'}$) in this (and next) section are particular to a configuration p.

Interference constraints for Protocol Model

Under the protocol model, when links ℓ and ℓ' are in each other's interference range (i.e., $f_{CC}(\ell, \ell') = 0$), and active concurrently, they must be allocated non-overlapping spectrum blocks. The following two constraints ensure that the channels used for links ℓ and ℓ' do not overlap:

$$e_\ell \leq s_{\ell'} + L(1 - y_{\ell,\ell'}) \qquad \forall f_{CC}(\ell, \ell') = 0 : (\ell, \ell') \in E, \;\; \ell \neq \ell' \tag{5.14}$$

$$e_{\ell'} \leq s_\ell + L y_{\ell,\ell'} \qquad \forall f_{CC}(\ell, \ell') = 0 : (\ell, \ell') \in E, \;\; \ell \neq \ell' \tag{5.15}$$

here, $y_{\ell,\ell'} \in \{0, 1\}$ is a decision variable. When $y_{\ell,\ell'} = 1$, then Eq. (5.14) forces the spectrum block assigned on link ℓ to precede (not overlap) that of link ℓ' (and thus ℓ and ℓ' may be active concurrently). Alternatively, when $y_{\ell,\ell'} = 0$, then Eq. (5.15) forces the spectrum block assigned on ℓ' to precede that of ℓ. Note that, when ℓ and ℓ' are not adjacent (or in interference range of each other), then both links may be assigned overlapping (partial or complete) spectrum blocks.

Table 5.2: Normalized I-factor and corresponding constraints for different scenarios

Scenario#	Conditions	$I_{(\ell,\ell')}$	Constraints ($\forall \ell, \ell' \in E, , \ell \neq \ell'$)
S_1	$e_\ell \leq s_{\ell'}$	0	$e_\ell \leq s_{\ell'} + L(1 - C^1_{\ell\ell'})$ $I_{\ell\ell'} \leq 0 + (1 - C^1_{\ell\ell'})$ $I_{\ell\ell'} \geq 0 - (1 - C^1_{\ell\ell'})$
S_2	$e_{\ell'} \leq s_\ell$	0	$e_{\ell'} \leq s_\ell + L(1 - C^2_{\ell\ell'})$ $I_{\ell\ell'} \leq 0 + (1 - C^2_{\ell\ell'})$ $I_{\ell\ell'} \geq 0 - (1 - C^2_{\ell\ell'})$
S_3	$e_{\ell'} \leq e_\ell$ $e_{\ell'} > s_\ell$ $s_{\ell'} \geq s_\ell$	$\frac{e_{\ell'} - s_{\ell'}}{b_{\ell'}} = 1$	$e_{\ell'} \leq e_\ell + L(1 - C^3_{\ell\ell'})$ $e_{\ell'} > s_\ell - L(1 - C^3_{\ell\ell'})$ $s_{\ell'} \geq s_\ell - L(1 - C^3_{\ell\ell'})$ $I_{\ell\ell'} \leq \frac{e_{\ell'} - s_{\ell'}}{b_{\ell'}} + L(1 - C^3_{\ell\ell'})$ $I_{\ell\ell'} \geq \frac{e_{\ell'} - s_{\ell'}}{b_{\ell'}} - L(1 - C^3_{\ell\ell'})$
S_4	$e_{\ell'} \leq e_\ell$ $e_{\ell'} > s_\ell$ $s_{\ell'} < s_\ell$	$\frac{e_{\ell'} - s_\ell}{b_{\ell'}}$	$e_{\ell'} \leq e_\ell + L(1 - C^4_{\ell\ell'})$ $e_{\ell'} > s_\ell - L(1 - C^4_{\ell\ell'})$ $s_{\ell'} < s_\ell + L(1 - C^4_{\ell\ell'})$ $I_{\ell\ell'} \leq \frac{e_{\ell'} - s_\ell}{b_{\ell'}} + L(1 - C^4_{\ell\ell'})$ $I_{\ell\ell'} \geq \frac{e_{\ell'} - s_\ell}{b_{\ell'}} - L(1 - C^4_{\ell\ell'})$
S_5	$e_{\ell'} > e_\ell$ $s_{\ell'} < e_\ell$ $s_{\ell'} \geq s_\ell$	$\frac{e_\ell - s_{\ell'}}{b_{\ell'}}$	$e_{\ell'} > e_\ell - L(1 - C^5_{\ell\ell'})$ $s_{\ell'} < e_\ell + L(1 - C^5_{\ell\ell'})$ $s_{\ell'} \geq s_\ell - L(1 - C^5_{\ell\ell'})$ $I_{\ell\ell'} \geq \frac{e_\ell - s_{\ell'}}{b_{\ell'}} - L(1 - C^5_{\ell\ell'})$ $I_{\ell\ell'} \geq \frac{e_\ell - s_{\ell'}}{b_{\ell'}} - L(1 - C^5_{\ell\ell'})$
S_6	$e_{\ell'} > e_\ell$ $s_{\ell'} < e_\ell$ $s_{\ell'} < s_\ell$	$\frac{e_\ell - s_\ell}{b_{\ell'}}$	$e_{\ell'} > e_\ell - L(1 - C^6_{\ell\ell'})$ $s_{\ell'} < e_\ell + L(1 - C^6_{\ell\ell'})$ $s_{\ell'} < s_\ell + L(1 - C^6_{\ell\ell'})$ $I_{\ell\ell'} \leq \frac{e_\ell - s_\ell}{b_{\ell'}} + L(1 - C^6_{\ell\ell'})$ $I_{\ell\ell'} \geq \frac{e_\ell - s_\ell}{b_{\ell'}} - L(1 - C^6_{\ell\ell'})$

5.2.2 Routing

We assume multi-path routing and the choice of routing depends upon the scheduling of active transmissions. The objective is to obtain an optimal routing and scheduling (and spectrum allocation) which minimize the total system activation time. Let $\mathcal{P}$ ($|\mathcal{P}| = \bar{P}$) denote the set of all feasible transmission configurations for a network, and x_ℓ^p is a link binary parameter that indicates whether ℓ is active in configuration p ($p \in \mathcal{P}$) or not. Define λ_p to be the time (in second) during which configuration p is active. Let f_ℓ^m denote the amount of traffic (bits) of session m passing through link ℓ. The capacity of a link (ℓ) depends on both the channel width and SINR threshold and is given by Eq. (5.7). $\omega^+(i)$ is the set of all outgoing links from node i and $\omega^-(i)$ is the set of all incoming links to node i; the problem is modeled as:

$$\text{Objective:} \qquad \text{Minimize} \sum_{p=1}^{\bar{P}} \lambda_p \tag{5.16}$$

$$\text{Subject to:} \sum_{\ell \in \omega^+(i): i \in V} f_\ell^m - \sum_{\ell \in \omega^-(i): i \in V} f_\ell^m = 0 \quad \forall i \in V - \{t_m, r_m\} \quad m = 1, 2, ..., M \tag{5.17}$$

$$\sum_{\ell \in \omega^+(t_m): t_m \in V} f_\ell^m - \sum_{\ell \in \omega^-(t_m): t_m \in V} f_\ell^m = R_m \qquad m = 1, 2, ..., M \tag{5.18}$$

$$\sum_{\ell \in \omega^+(r_m): r_m \in V} f_\ell^m - \sum_{\ell \in \omega^-(r_m): r_m \in V} f_\ell^m = -R_m \qquad m = 1, 2, ..., M \tag{5.19}$$

$$\sum_{p=1}^{\bar{P}} \lambda_p \times b_\ell \times x_\ell^p \times \log_2(1+\beta) - \sum_{m=1}^{M} f_\ell^m \geq 0 \qquad \ell \in E \tag{5.20}$$

$$f_\ell^m \geq 0, \lambda_p \geq 0.$$

Here, Eq. (5.16) aims to minimize the total system activation time. Constraints (5.17-5.19) present the flow conservation constraints. Eq. (5.20) indicates that the total traffic routed through link ℓ can not exceed the total transport capacity of ℓ.

We note that when the set of all possible configurations is given, then x_ℓ^p for each input configuration is predetermined and further b_ℓ is a parameter (link attribute) that indicates the spectrum size on link ℓ. Finding the solution of the above model then relies on determining the set $\mathcal{P}$. According to the number of links and spectrum partitions, the size of $\mathcal{P}$ can be extremely large. This makes the above model computationally infeasible, since it may not be possible to enumerate all such configurations. Additionally most of these configurations will not be used in the optimal solution. Our approach to solve this problem is to leverage on the knowledge that only a subset of $\mathcal{P}$ will used to obtain the optimal solution and thus avoid the explicit enumeration of $\mathcal{P}$ [79, 90]; we use a column generation decomposition, where routing and scheduling are separated into different subproblems (LP and ILP) [89, 90].

5.3 Solution Approach

Column generation (CG) [79, 85] is an optimization technique that decomposes a linear program (LP) into a master model and a pricing model. The restricted master model is initialized with a subset of columns, the basis, P_0 (in our case, configurations, $P_0 \subseteq \mathcal{P}$) of the LP and is easily solved to obtain a feasible solution to the main problem. The pricing, is solved to identify whether the master should be enlarged with additional columns or not (as explained in earlier chapters). Therefore, as opposed to an LP where all the columns are used at the same time to obtain the optimal solution, CG alternates between the master (LP) and the pricing (ILP) models, until the former contains the necessary columns to find the optimal solution of the original LP [90].

A. MASTER PROBLEM

$$\text{Objective:} \quad \text{Minimize} \sum_{p \in P_0} \lambda_p \tag{5.21}$$

Subject to: Equations (5.17-5.19)

$$\sum_{p \in P_0} \lambda_p \times b_\ell \times x_\ell^p \times \log_2(1+\beta) - \sum_{m=1}^{M} f_\ell^m \geq 0 \qquad \forall \ell \in E \tag{5.22}$$

$$f_\ell^m \geq 0, \lambda_p \geq 0.$$

Note that in Eq. (5.22), x_ℓ^p is not a variable, but rather a parameter (containing scheduling information) which is obtained after solving the pricing subproblem. b_ℓ is a parameter, spectrum attribute of a link ℓ. Thus the resulting problem is simply a linear program.

During every iteration, when the master problem is solved, we need to verify the optimality of the solution. If it is optimal, we conclude our search, or else decide a new column to join in its current basis that can improve the current solution. This can be achieved by examining whether any new column that is not currently in P_0, has a negative reduced cost. Denoting the dual variables corresponding to Eq. (5.22) by u_ℓ, the reduced cost ($\overline{cost}$) for any new column that is not in P_0 can be expressed as:

$$\overline{cost} = 1 - \log_2(1+\beta) \times \sum_{\ell \in E} u_\ell \times b_\ell \times x_\ell \tag{5.23}$$

Therefore, the pricing subproblem can be written as follows:

B. PRICING PROBLEM

Objective: Minimize $\overline{cost}$

Subject to:

1)- (**For physical model**): Constraints (5.8)-(5.10); Constraints (S1)-(S6) and

Constraints (5.11)-(5.13). Note that variable x_ℓ^p is replaced by x_ℓ since the pricing subproblem deals with a particular configuration at a time. $e_\ell = \{0, 1, ..., B\}$, $s_\ell = \{0, 1, ..., B\}$, $C^i_{\ell\ell'} = \{0, 1\}$, $x_\ell = \{0, 1\}$, $0 \leq I_{(\ell,\ell')} \leq 1$.

2)- (**For protocol model**): Constraints (5.8)-(5.10) and Constraints (5.14)-(5.15), $e_\ell = \{1, 2, ..., B\}$, $s_\ell = \{1, 2, ..., B\}$, $x_\ell = \{0, 1\}$, $y_{\ell,\ell'} = \{0, 1\}$

5.4 Physical Model vs. Protocol Model

Recently, the authors of [97] noted that although the physical model is realistic in capturing the interference, its associated complexity renders it less attractive. The protocol model instead relies on a more simplistic approach to model interference and thus overcome the complexity of the SINR model. This simplified model could either overly estimate or underestimate the interference because it does not accurately capture the physical layer characteristics. As a result, solutions obtained under the protocol model could be practically not feasible. In this section we study the performance of our method and show that the SINR-based design model is indeed not scalable and the protocol model yields solution which are not feasible.

We consider a network of 10 nodes (Fig. 5.2(left)) with 5 single-hop sessions. The links are labeled $\ell_1, \ell_2, \ell_3, \ell_4$ and ℓ_5 respectively. The traffic demands for all sessions are 35Mbits. We assume that any spectrum width (5,10,20 or 40MHz) may be allocated to satisfy the demands. The transmission power is P_t = 1mW; α=2, η_0=10^{-6}W/MHz and δ = 0.2. The SINR threshold is fixed to β=1.3 and B=80MHz. We use both the protocol and physical models to determine the minimum system activation time along with optimal spectrum allocation. The CG method is implemented in C++ and solved using CPLEX Concert Technology. We assume that under protocol constraints, ℓ_1 interferes with ℓ_2; ℓ_5 interferes with both ℓ_2 and ℓ_4; ℓ_3 interferes with ℓ_4.

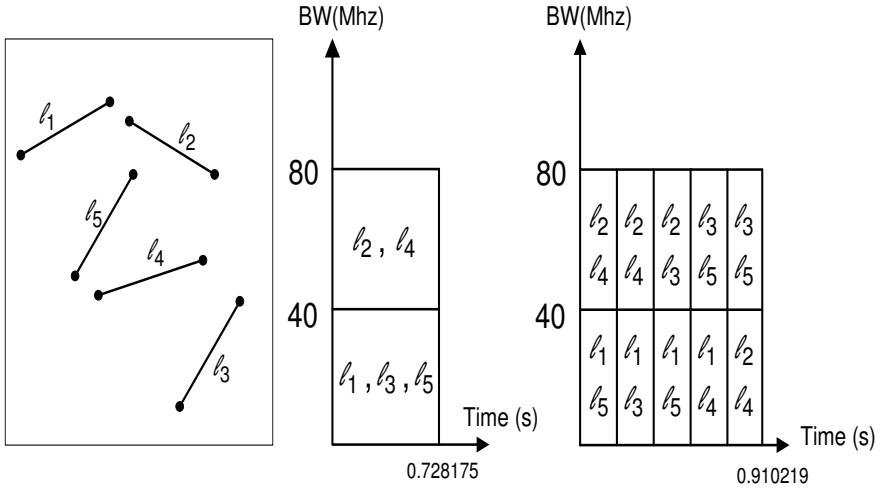


Figure 5.2: Optimal spectrum allocation: protocol model (middle) and physical model (right)

Table 5.3: SINR on active links

Active Links	SINR	Feasibility
ℓ_1	1.23274	Infeasible
ℓ_2	1.34481	Feasible
ℓ_3	1.23274	Infeasible
ℓ_4	1.34481	Feasible
ℓ_5	1.23274	Infeasible

The optimal solution obtained under the protocol model shows that all links may be active in the same time slot and their spectrum allocation is shown in Fig. 5.2(middle). Since links ℓ_1, ℓ_3 and ℓ_5 do not interfere with each other, they can be allocated the same spectrum block (0-40MHz) and the other two links (ℓ_2 and ℓ_4) are allocated the same spectrum (40-80MHz). Table 5.3 shows the SINR on all active links; we can see that links ℓ_1, ℓ_3 and ℓ_5 have SINR less than the threshold (β) and thus transmissions on these links are not feasible. Indeed, the protocol model underestimated the interference from other transmissions in the network. The cumulative effect of interference, when counted, affect the SINR on active links and therefore the transmission capacity of these links should be accordingly adjusted.

Table 5.4: Comparison between physical and protocol model

	Physical Model	Protocol Model
No. of Constraints	115,346	4,258
No. of Variables	29,248	2,592
CPU Time (sec)	441104.28	4.48
Sys. Act. time (sec)	0.910219	0.728175

Fig. 5.2(right) shows the optimal link scheduling and spectrum allocation under the SINR model. Clearly, the partitioning of the spectrum and the scheduling of the links are different from those in Fig. 5.2(middle) and the resulting system activation time is higher. The obtained solution is both optimal and feasible. Table 5.4 compares the complexity and performance of the protocol and physical models in terms of system activation time and the CPU time to obtain the optimal solution. We observe from the Table that the protocol model is quite fast (4.48s) in getting the solution whereas the physical model needed 5 days (for this small network) to return the optimal solution. This indeed shows that the SINR model is not attractive for any practical network size; the protocol model however must be tuned properly [97] to yield feasible solutions, as is explained in the next section.

5.5 A Simulated Annealing-based Pricing Subproblem

The previous section showed that the computational complexity when considering the SINR model is prohibitively expensive and that the solution obtained using the protocol model is not feasible. We observe that when decomposing the problem using CG, the master problem deals with non-integer variables (f_ℓ^m, λ_p) and usually is easy to solve (LP). The pricing, however deals with integer variables ($e_\ell, s_\ell, x_\ell, y_{\ell,\ell'}, C_{\ell\ell'}^i$) which is commonly more difficult to solve; further, to ensure optimality, the last

iteration of the Simplex method must be solved to optimality, which is usually quite difficult for larger networks. The number of constraints and variables resulting from the small network in the previous section is shown in Table 5.4; as the network size increases, even the protocol model becomes difficult to solve.

In this section, we present our two-fold approach for overcoming the issues with the protocol and the physical models. Our solution consists of developing a heuristic, based on simulated annealing (SA) [98], to solve the pricing subproblem. SA is a very efficient meta-heuristic technique, which tries to find better solutions (in our case configurations) by comparing the cost of current and candidate solutions. Unlike greedy heuristics, SA usually does not stuck in the local optima rather it tries to find the global optimal solution. Recall, pricing returns to the master a candidate configuration with negative reduced cost with the potential of improving the objective of the master LP. In our heuristic, we relax the SINR (physical) interference model and use a graph-based interference (protocol) model; all active links in each chosen configuration are mutually outside the interference range of each other. However, we consider among these configurations, only those that have their link SINR constraints satisfied:

Step 1: Generate an initial configuration $p_{initial}$ (configuration is defined in Section 5.2) from a set of randomly selected links without violating the interference constraints.

Step 2: Assign current configuration $p_{cur} = p_{initial}$.

Step 3: Compute $\overline{cost}_{cur}$ of p_{cur} using Eq. (5.23).

Step 4: Set initial temperature $T = \overline{cost}_{cur}$ and the temperature reduction factor RF to some constant. Initialize the outer-loop (OL), the maximum number of temperature reduction and inner-loop (IL), the maximum number of iteration with a particular temperature.

Step 5: Repeat *Step 6* to *Step 10* for OL times.

Step 6: Repeat *Step 7* to *Step 9* for IL times.

Step 7: Generate a candidate solution, p_{can}, through the function ***Schedule*** (which will be introduced in Section 5.5.1) and compute the cost $\overline{cost}_{can}$ using Eq. (5.23).

Step 8: If $(\overline{cost}_{can} \leq \overline{cost}_{cur})$, accept the configuration; $p_{cur} = p_{can}$; $\overline{cost}_{cur} = \overline{cost}_{can}$.

Step 9: If $(\overline{cost}_{can} > \overline{cost}_{cur})$, accept the configuration with a probability $Pr = e^{(\frac{\overline{cost}_{cur} - \overline{cost}_{can}}{T})}$ and set $p_{cur} = p_{can}$, $\overline{cost}_{cur} = \overline{cost}_{can}$.

Step 10: Reduce T $(T = T \times RF)$.

We terminate as soon as we obtain a negative reduced cost $(\overline{cost}_{can} < 0)$ and add the configuration (p_{can}) as a new column into the basis of the master. The alternation between master and pricing subproblem continues as long as the pricing subproblem provides a negative reduced cost.

5.5.1 Link Scheduling

The function ***Schedule***, introduced earlier, generates a candidate configuration and is described through Algorithm 5.1. The main idea of this function is to start from an already known configuration and generate a new candidate (neighboring) configuration which may be added to the basis of the master problem. The approach is to remove arbitrarily a link from the current configuration and add one (or more) link which either does not conflict with existing links in the configuration (i.e., which satisfies the radio and interference constraints). If the new link to be added interferes with existing links in the current configuration, then a spectrum block in the permissible band is determined for the new link which does not overlap with the spectrum of those links it interferes with. Otherwise, the link is not added. The details are

presented below.

Let F_{CR}^{ℓ} be a set that contains all links in the network which share a common radio with link ℓ. Let F_{CC}^{ℓ} be a set that contains all links in the network which fall inside the interference range of ℓ. Let A be a set of active links ℓ where $x_{\ell}^{p} = 1$ in current configuration (p_{cur}).

To construct a new candidate (neighboring) configuration, we remove at random a link ℓ (or more) from A (Line (11) in Algorithm 5.1) and search for links which may be added to the current configuration (and do not violate both interference and radio constraints). For example, after removing a link, another link ℓ' will be added to A if for every existing link $\ell \in A$, $\ell \notin F_{CR}^{\ell'}$ and $\ell \notin F_{CC}^{\ell'}$ (Lines (12-18) in Algorithm 5.1). If $\ell \in F_{CC}^{\ell'}$ (Lines (19-28) in Algorithm 5.1), then ℓ' will be added to A given that we can find a spectrum block to be assigned on ℓ' and which does not overlap with that of ℓ (Lines (20-21) in Algorithm 5.1). Otherwise, ℓ' cannot be added. To facilitate the search for such non overlapping spectrum block, we use a slotted representation of the permissible spectrum band.

For each link ℓ, let BA_{ℓ} denote the set of contiguous spectrum slots or fragments, of unit length each (e.g., 1MHz), and of total length B(MHz). Let t_i^{ℓ} $(1 \leq i \leq B)$ be a binary variable indicating whether spectrum fragment i is assigned ($t_i^{\ell} = 1$) to link ℓ or not ($t_i^{\ell} = 0$). Thus, BA_{ℓ} is a bitset of length B for each ℓ and is configured (according to t_i^{ℓ}) when ℓ is added to A. That is, BA_{ℓ} indicates the location of spectrum block b_{ℓ} in the permissible spectrum. Next, we illustrate the spectrum block assignment on link ℓ' ($\ell \in F_{CC}^{\ell'}$) that may be added to A. We perform a *bitwise OR* operation (Line (20) in Algorithm 5.1) on all bitsets BA_{ℓ}, $\forall \ell \in F_{CC}^{\ell'}$ and determine the resultant binary vector $K : K = \bigvee BA_{\ell}, \forall \ell \in F_{CC}^{\ell'}, \ell \in A$. We search through K for available contiguous spectrum fragments of size $b_{\ell'}$ to determine $s_{\ell'}$ and thus allocate the spectrum block to the link and add ℓ' to A (Line (21) in Algorithm

Algorithm 5.1 Link scheduling

F_{CR}^{ℓ}, F_{CC}^{ℓ}, A: Set of active links, E: Set of links.
s_ℓ, e_ℓ, b_ℓ: identify the location and size of spectrum allocated to link ℓ.
BA_ℓ: Spectrum availability vector for link ℓ
Schedule (p_{cur})
Initialize $BA_\ell = \emptyset \qquad (\ell \in E)$
Determine A (from current configuration, p_{cur})
Set BA_ℓ corresponding to the spectrum band (all fragments) of ℓ, $\forall \ell \in A$
Select at random $\ell (\in A)$; $A \leftarrow A - \{\ell\}$; $E \leftarrow E - \{\ell\}$
For each link $\ell \in A$, determine its BA_ℓ, where the values of those bits corresponding to the spectrum block is set to 1
while $E \neq \emptyset$ **do**
    Select at random $\ell' (\in E)$ and remove it from E
    if $F_{CR}^{\ell'} \cap A = \emptyset$ and $F_{CC}^{\ell'} \cap A = \emptyset$ **then**
        $s_{\ell'} = 0$; $e_{\ell'} = b_{\ell'}$
        $sinr\_check(A, \ell', s_{\ell'}, e_{\ell'})$
        if $sinr\_check$=true **then**
            $A \leftarrow A + \{\ell'\}$; Update $BA_{\ell'}$ (set status of slots $\{0, b_{\ell'}\}$ to 1)
        end if
    end if
    if $F_{CR}^{\ell'} \cap A = \emptyset$ and $F_{CC}^{\ell'} \cap A \neq \emptyset$ **then**
        Perform bit-wise OR operation on all BA_ℓ, $\forall \ell \in A$ and the result is in K
        $s_{\ell'} = F(K, b_{\ell'})$; $e_{\ell'} = s_{\ell'} + b_{\ell'}$
        if $s_{\ell'}$ is valid **then**
            $sinr\_check(A, \ell', s_{\ell'}, e_{\ell'})$
            if $sinr\_check$=true **then**
                $A \leftarrow A + \{\ell'\}$; Update $BA_{\ell'}$
            end if
        end if
    end if
end while
Set candidate configuration p_{can} from A

Algorithm 5.2 SINR-Check Function

sinr_check$(A, \ell', s_{\ell'}, e_{\ell'})$
Calculate all normalized I-factors ($I_{\ell,\ell'}$) using Table 5.2 (use BA_ℓ to obtain the start and end frequencies of each link $\ell \in A$)
Calculate SINR of all links ($SINR_\ell$) using Eq. 5.3
if all $SINR_\ell \geq \beta$ **then**
    Return ($sinr\_check$=true)
else
    Return ($sinr\_check$=false)
end if

5.1). We use a first fit allocation where the first spectrum block available (in K) is allocated to ℓ'. Otherwise, if a block could not be found, then ℓ' is not added to A. This method is shown in Line (21) in Algorithm 5.1 and makes use of the string matching algorithm of Knuth-Morris-Pratt (KMP) [99] to find the start frequency of the allocated spectrum block. This procedure is repeated for all links in the network, except those in A. Finally, we obtain a candidate configuration $p_{can} = A$.

Note that when Algorithm 5.1 generates a candidate configuration, it is guaranteed to be feasible under SINR constraint. The feasibility conformance check is performed through Algorithm 5.2; each time a new candidate link ℓ' is to be added to the set A, a test (lines (14) and 23 in Algorithm 5.1) is carried out to determine whether each link currently in A, including the new one, satisfies the SINR constraint. Algorithm 5.2 makes use of Eq. (5.3) which in turns considers the cumulative interference in determining the transmission quality (i.e., SINR).

5.6 Numerical Results

Since the computation complexity of the SINR model is prohibitive for obtaining an optimal solution, we do not consider it in our evaluation. The models we consider are PI_v and PH_v, where the former refers to the method presented in section 5.3 with the protocol interference constraints of section 5.2.1 The second refers to the model in which the pricing subproblem is solved using simulated annealing of section 5.5. We consider two versions of PH_v, namely $PH1_v$ and $PH2_v$ where the former does not perform SINR feasibility check when generating configurations and the latter does. The comparison between $PH1_v$ and $PH2_v$ is necessary to show that generating schedules not satisfying the SINR constraint (i.e., underestimating interference) will yield to solutions which are not practically feasible. The comparison between PI_v and PH_v is to show both the better scalability of our SA method as well as to underline the

Table 5.5: System activation time and CPU time

	Sys. Activation Time (sec)			CPU Time (sec)		PI_v/ $PH1_v$	PI_v/ PI_{fb}
Nodes _Sess	PI_{fb}	PI_v	$PH1_v$	PI_v	$PH1_v$	O.G. (%)	Gain (%)
05_03	1.68	1.33	1.33	3.47	1.04	0	20
05_10	4.22	3.60	3.60	3.23	1.32	0	14
05_20	8.71	7.76	7.76	4.36	1.43	0	11
05_30	14.39	12.35	12.35	3.60	1.68	0	14
10_03	1.28	1.05	1.05	90.81	14.47	0	18
10_10	2.90	2.301	2.305	247.02	111.1	0.17	20
10_20	8.83	5.80	5.80	133.84	44.37	0	34
10_30	11.65	8.446	8.449	158.82	72.74	0.03	27
20_03	0.80	0.77	0.77	28398	760	0	3
20_10	2.59	1.49	1.49	31334	1323	0	42
20_20	3.97	2.44	2.51	321301	20936	2.8	38
20_30	6.02	3.71	3.71	204278	47320	0	38

effectiveness of the method to obtain close to optimal solutions. To study the benefits of variable spectrum assignment, we consider a model which assumes allocation of spectrum of preset widths and we take the spectrum that yields best results. This model (whose pricing subproblem is solved using ILP) is referred to as PI_{fb} in the discussion.

We consider randomly generated networks (5, 10, and 20 nodes) deployed over a 100m×100m area and different traffic instances (3-30 sessions). Each session m has a traffic demand randomly generated in the range of $0< R_m \leq$35Mbits. The values of the outer-loop (OL) and inner-loop (IL) of SA vary from 40 to 200. The cooling factor is assumed to be constant ($RF = 0.98$).

We start by first noting that $PH1_v$ and $PH2_v$ performed quite accurately on the network of Fig. 5.2 where $PH1_v$ and $PH2_v$ yielded the same results as that of Figs. 5.2(middle) and (right) respectively, which shows that SA achieved the same optimal solution obtained when solving the pricing subproblem through ILP. The results in

Table 5.6: $PH1_v$ vs.$PH2_v$ for a 10-node network

Nodes _Sess Nodes _Sess	Model Model	# of Config.	# of Active Links	# of Infeasible Links	Sys. Act. Time (sec.)
10_03	$PH1_v$	12	58	5	1.05
	$PH2_v$	7	35	0	1.05
10_10	$PH1_v$	17	78	9	2.305
	$PH2_v$	17	75	0	2.33
10_20	$PH1_v$	18	80	10	5.80
	$PH2_v$	17	76	0	5.80
10_30	$PH1_v$	21	92	13	8.449
	$PH2_v$	20	87	0	8.50

Table 5.5 compare both the accuracy and scalability of the SA method and underline the benefits of non-uniform spectrum allocation. Both PI_v and $PH1_v$ achieve very similar activation time, constantly outperforming PI_{fb}, especially at higher loads (relative to the network size) where the interference becomes more problematic. Thus, our method becomes more efficient (than PI_{fb}) in managing the interference by properly allocating spectrum of different widths to selected links. The improvement gain is shown in Table 5.5; smaller gains are observed when the load is lighter. This is because the interference is not a bottleneck when there is sufficient spectrum resources in the network and the traffic load is light and flows may be routed apart from each other to avoid interference. Solving the pricing subproblem using SA yields a solution that is optimal or very close to it (worst optimality gap we observed is below 3%) and the computation time is much smaller; for small networks, both models return the solution in few seconds or few minutes and for larger networks, e.g., 20 nodes, $PH1_v$ can be up 37 times faster.

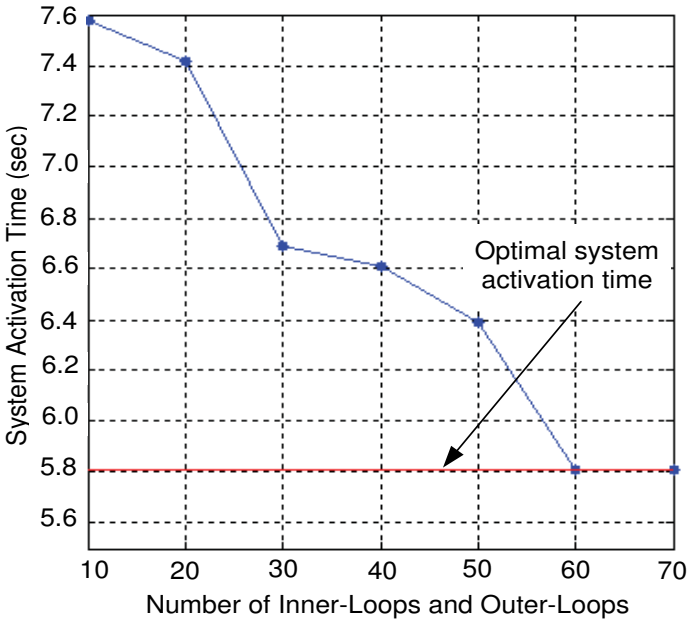
We compare both $PH1_v$ and $PH2_v$ in Table 5.6 using a network of 10 nodes. Both methods yield almost the same system activation time with the former yields slightly smaller activation times when the load is higher. $PH1_v$, however, does not check the feasibility of the generated configurations. Table 5.6 shows the number of

Table 5.7: $PH1_v$ vs. $PH2_v$, different networks

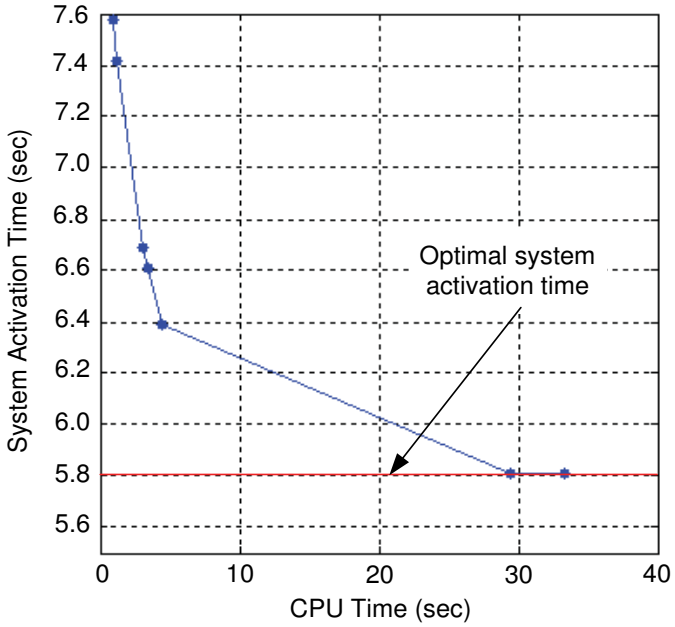
	Sys. Activation Time (sec)			CPU Time (sec)		$PH2_v$/ $PH2_{fb}$
Nodes_ Session	$PH2_{fb}$	$PH2_v$	$PH1_v$	$PH2_v$	$PH1_v$	Gain (%)
10_03	1.28	1.05	1.05	74.50	14.47	18
10_10	2.90	2.33	2.305	1371.8	111.11	19.6
10_20	8.83	5.80	5.80	145.29	44.37	34.3
10_30	11.65	8.50	8.449	907.99	72.74	27
20_03	0.80	0.77	0.77	4364.4	760.96	3
20_10	2.66	1.49	1.49	31155	1323.6	44
20_20	5.97	2.70	2.51	67675	20936	54.7
20_30	9.16	3.95	3.71	98395	47320	56.9

active links whose SINR is not satisfied, under $PH1_v$, but are nonetheless scheduled for transmission; the table shows that the number of active links whose SINR is not satisfied varies between 8% and 15% of all active links in the corresponding network instances. Transmissions on all links generated under $PH2_v$ are, however, all feasible, which is guaranteed by the feasibility check in Algorithm 5.2. This feasibility test comes however with additional computation complexity, as shown in Table 5.7 since the algorithm (Algorithm 5.1) will try different schedules to find a candidate feasible solution (section 5.5.1). The larger the size of the network, the larger the search space and thus the higher the computation time. We note, however, that $PH2_v$ remains more scalable than PI_v and produces practically feasible solutions that both PI_v and $PH1_v$ cannot generate.

Table 5.7 compares the performance of $PH2_v$ with a design method which uses predetermined spectrum widths, $PH2_{fb}$; in $PH2_{fb}$, the pricing is solved using SA and ensures that all candidate configurations are SINR-feasible. We show the results of the best spectrum width. The results in Table 5.7 confirm our reasoning that fixed spectrum allocation results in a poor performance due to the lack of flexibility in spectrum allocation and thus poor interference management capability. The gains of



(a) System activation time Vs. Number of Inner and Outer-loops



(b) System activation time Vs. CPU time

Figure 5.3: Effect of inner and outer loops on system activation time

$PH2_v$ are more pronounced at higher loads when the interference effect becomes more intense. Hence, flexible spectrum allocation provides an effective way for dealing with interference and results in better overall performance.

Finally, we note that the performance of SA largely depends on the selection of OL and IL parameters, whose values determine a tradeoff between the quality of the obtained solution and the scalability of the method. Fig. 5.3(a) shows the convergence of the solution of PH_v to the optimal solution as we vary the values of the configuration parameters. Fig. 5.3(b) shows that within 30 seconds the optimal solution is obtained and this corresponds to IL=OL=60, values beyond which the CPU time unnecessarily increases to yield the same optimal solution.

5.7 Conclusion

We mathematically formulated the joint problem of scheduling and spectrum allocation and decomposed it using CG, under the SINR interference constraints. Our investigation showed that the associated computation complexity is prohibitive from obtaining the optimal solution. We therefore presented a SA approach for solving the pricing and is based on a simplified interference model; we augment it however with an SINR check to make sure that only feasible configurations are used towards obtaining the optimal solutions. Our results indicate that the SA method yields solutions that are very close to optimal and the computation time is substantially reduced. Our results also showed that a flexible spectrum allocation is effective in managing the interference in the network and leads to improving the system performance.

Chapter 6

Resource Allocation in Cooperative Cellular Networks

In this chapter, we study the joint problem of optimal relay node selection and power allocation among the selected relays for cooperative networks with amplify and forward relaying. The objective of the joint optimization problem is to maximize the total network capacity. We consider both unicast and multicast traffic scenarios.

We start with modeling the problem considering unicast traffic as a mixed Boolean-convex problem and solve it to optimality using the Branch and Bound (B&B) technique [83]. As the complexity of the problem grows exponentially with the size of the network (number of users and relays), it becomes computationally infeasible to solve the problem optimally. Therefore, similar to [59], we derive a tight upper-bound for the problem and exploit the Karush-Kuhn-Tucker (KKT) conditions [76] to illustrate the tightness of the bound. We also confirm the tightness of the upper-bound solutions by comparing it with its optimal counterpart. A simple water-filling (WF) method is then presented, which solves the problem efficiently (in terms of computation time) and provides very near to optimal solution.

We then extend our unicast model for multicast traffic scenarios and also model

the problem as a mixed Boolean-convex one, similar to unicast scenarios we use B&B to solve the problem optimally. This model also suffers from the same problem of computational infeasibility for larger size networks. Unlike unicast scenarios, it does not appear to be straight forward to obtain a tight upper-bound as the KKT conditions for multicast scenarios do not demonstrate similar insights as in the unicast case. As a result of that, the WF technique, in this case, fails to provide a near optimal solution and we present a sequential fixing (SF) procedure where the Boolean variables are determined iteratively through solving a sequence of convex optimization problems [53]. Our results confirm that the SF technique substantially reduces the computation time and achieves near optimal solutions.

We finally introduce a joint relay and power allocation technique considering a more practical scenario while the relay allocates power from a set of discrete values. Although almost all previous works consider continuous relay power, which can be allocated to the data flows with any value as long as it does not violate the total power constraint, in reality, the power levels in a digital cellular system are quantized into discrete values. Therefore, in discrete power level scenarios, we present both optimal (based on B&B) and heuristic (based on SF) models to solve the optimization problem. We then vary the number of discrete power levels to compare the results with continuous power allocation and show that considering as small as four levels of discrete power yields solutions close to the continuous power allocation.

The rest of the chapter is organized as follows: The system model is described in Section 6.1. The problem formulations considering unicast traffic is discussed in Section 6.2. In Section 6.3 an algorithm based on water-filling technique is presented for unicast scenario. In Sections 6.4 and 6.5, we present problem model for multicast scenarios and describe a sequential fixing algorithm respectively. Joint relay selection and power allocation considering discrete power levels is presented in Section 6.6.

Numerical results and discussions are given in Section 6.7 and conclusion in Section 6.8.

6.1 System Model

We model a cellular network with a single base station (BS) s serving a set of users $\mathcal{D}$ ($|\mathcal{D}| = D$). The communication between the BS and the users is assisted by a set of $\mathcal{R}$ ($|\mathcal{R}| = R$) stationary relay nodes. In this work, we focus on the downlink transmissions. Each of the users in a unicast scenario is assigned an orthogonal channel (e.g., orthogonal frequency division multiple access system). Hence, there is no intra-cell interference. We also assume relays with AF capabilities; in such cooperative communications, a transmission between a s-d pair requires two time slots. In the first time slot, the BS transmits data to all users and the relay nodes overhear these transmissions. In the second time slot, one relay for each transmission, chosen a priori, amplifies the data and forwards them to their intended users. In this work we assume that one user cannot get assistance from more than one ('single best') relay .

On the other hand, for multicast traffic, users are divided into M multicast groups; let $\mathcal{S}^m$ denote the set of users belonging to multicast group m ($m = 1, 2, ..., M$). Note that, we assume users within a multicast group may be assisted by one or more relays, each however gets assistance from at most one. The users belonging to the same multicast group assisted by the same relay will be assigned the same channel. Therefore, use of DSTC (distributed space time codes) is not necessary, hence eliminating the symbol level synchronization problem. Additionally, we note that, users belonging to different multicast groups are assumed to use orthogonal channels. Hence, for multicast traffic also, there is no intra-cell interference. For both unicast and multicast traffic, we further do not consider any interference from

neighboring cells.

The capacity for a user d assisted by AF capable relay r can be written as follows:

$$C(d) = \frac{W}{2}\log_2(1 + SNR_{sd} + \frac{SNR_{sr} \cdot SNR_{rd}}{SNR_{sr} + SNR_{rd} + 1}) \tag{6.1}$$

It is important to note that, multicast transmission has some distinctive aspects opposed to unicast networks. The most important among them is the fact that, multicast capacity is dominated by the user with the weakest link in order to minimize outage and retransmission [100,101]. Taking this into account, the multicast capacity for a session m can be written as follows:

$$C(m) = \frac{W}{2} \arg\min_{d \in \mathcal{S}^m} \{\log_2(1 + SNR_{sd} + \frac{SNR_{sr} \cdot SNR_{rd}}{SNR_{sr} + SNR_{rd} + 1})\} \tag{6.2}$$

Here, W is the channel bandwidth and SNR_{sd}, SNR_{sr}, SNR_{rd} are the signal to noise ratio at the destination and relay nodes, defined as:

$$SNR_{sd} = \frac{P_s}{W\eta_0}|h_{sd}|^2, \quad SNR_{sr} = \frac{P_s}{W\eta_0}|h_{sr}|^2, \quad SNR_{rd} = \frac{P_r}{W\eta_0}|h_{rd}|^2,$$

where, P_r is the total power used by a relay and P_s is the transmission power used by the BS for each session. η_0 is the noise power. h_{ab} denotes the channel gain and captures the effect of path-loss and fading between transmitter a and receiver b. Similar to [59], we use the COST-231 model as recommended by the IEEE 802.16j working group [102] to model the channel between the BS, relay and users. While the BS and relays are usually placed at some height above the ground and the fading has a line-of-sight (LoS) component and modeled as a Rician random variable, BS to users and relays to users are modeled as Rayleigh random variables for non-LoS communications.

6.2 Unicast Problem Formulation

We start our problem formulation considering unicast traffic. Our objective is to maximize the performance of the network in terms of the overall network capacity by performing a joint relay selection and relay power allocation among the sessions going through these relays.

Let α_{rd} be the fraction of total power used by relay r to forward data to user d. The optimization problem then can be written as:

$$\text{Max} \sum_{d=1}^{D} C(d) = \text{Max} \sum_{d=1}^{D} \frac{W}{2} \log_2(1 + \frac{P_s}{W\eta_0}|h_{sd}|^2 + \sum_{r=1}^{R} \frac{\frac{P_s}{W\eta_0}|h_{sr}|^2 \frac{P_r}{W\eta_0}|h_{rd}|^2 \alpha_{rd}}{\frac{P_s}{W\eta_0}|h_{sr}|^2 + \frac{P_r}{W\eta_0}|h_{rd}|^2 \alpha_{rd} + 1}) \tag{6.3}$$

Subject to:

$$\alpha_{id} \times \alpha_{jd} = 0 \qquad i \neq j,\ i, j \in \mathcal{R},\ \forall d \in \mathcal{D} \tag{6.4}$$

$$\sum_{d=1}^{D} \alpha_{rd} \leq 1 \qquad \forall r \in \mathcal{R} \tag{6.5}$$

$$\alpha_{rd} \geq 0.$$

Eq. (6.3) is the objective function, which is a concave increasing function of the relay power. The objective function aims at maximizing the overall network capacity. If α_{rd} is greater than 0, relay r is selected to forward data to user d and the value of $P_r\alpha_{rd}$ is the power allocated to that data flow. As we have mentioned earlier, each user can only be assisted by at most one relay, which is enforced by Eq. (6.4). Eq. (6.5) ensures that the combined power allocated to all users going through a particular relay r does not exceed the total power budget of that relay.

The solution to this optimization problem ((6.3)-(6.5)) is made complicated by constraint (6.4). The usual gradient based methods can not be used to maximize the objective function of this problem. An exhaustive search to find a solution of

the problem involves testing all R^d possible relay combinations, which is impossible for realistic user and relay sizes. One way to solve this problem is to re-write it as a mixed Boolean-convex problem and use B&B to solve it optimally. To re-write the problem as a mixed Boolean-convex problem, we introduce a binary variable a_{rd} and represent Eq. (6.4) with the following two constraints:

$$a_{rd} \geq \alpha_{rd} \qquad \forall d \in \mathcal{D}, \forall r \in \mathcal{R} \tag{6.6}$$

$$\sum_{r=1}^{R} a_{rd} \leq 1 \qquad \forall d \in \mathcal{D} \tag{6.7}$$

Here, $0 \leq \alpha_{rd} \leq 1$ and $a_{rd} = \{0, 1\}$. Constraint (6.7) enforces that for each user d at most one a_{rd} can be 1 (others are 0); hence from Eq. (6.6), we ensure that for each user d, at most one α_{rd} can have non-zero value. Therefore the Mixed Boolean-convex problem can be formulated as follows.

Optimal Unicast Model:

$$\text{Max} \sum_{d=1}^{D} C(d) \tag{6.8}$$

Subject to:

$$a_{rd} \geq \alpha_{rd} \qquad \forall d \in \mathcal{D}, \forall r \in \mathcal{R} \tag{6.9}$$

$$\sum_{r=1}^{R} a_{rd} \leq 1 \qquad \forall d \in \mathcal{D} \tag{6.10}$$

$$0 \leq \alpha_{rd} \leq 1 \ ; \ a_{rd} = \{0, 1\}.$$

The Mixed Boolean-convex problem exploits the B&B technique to find the value of the Boolean variable (a_{rd}) and obtains the optimal solution for the problem. B&B is a nonheuristic method. It searches for a globally optimal solution of nonconvex problems with predefined precision of optimality [83]. The problem, however, with B&B is that the technique is often slow and in the worst case, the computation time increases exponentially with the problem size. As a result as we verify later, our B&B

approach for solving the problem does not work well (in terms of computation time) for large number of relays and users. To this end, we explore in the next section, a method which solves the problem efficiently and provides near optimal solutions.

6.3 A Near Optimal Algorithm for Unicast Based on Water-Filling Technique

We again consider the initial nonconvex optimization problem ((6.3)-(6.5)) and transform it into a convex one by ignoring Eq. (6.4) as in [59]. We will refer to this convex problem as the relaxed problem. Solving this relaxed problem enable us to get an upper-bound solution to the problem. However, ignoring Eq. (6.4) means that some users might get served by more than one relay, which results in conflicts. But these conflicts are limited, and despite that, the upper-bound is quite tight. The tightness of the upper-bound can be observed by deriving the KKT condition for the relaxed problem. Furthermore, we later introduce additional constraints to eliminate any conflicts that may arise.

Let us assume that there are only one user (d) and two relays ($r_1,\ r_2$) in the network. Then the Lagrangian for the relaxed problem can be written as:

$$\begin{aligned}\mathcal{L}(\{\alpha_{r_1d},\alpha_{r_2d}\};\{\lambda_d^{r_1}\},\{\lambda_d^{r_2}\},\nu_{r_1},\nu_{r_2}) = \log_2(1+z_{sd}+\frac{x_{sr_1}y_{r_1d}\alpha_{r_1d}}{x_{sr_1}+y_{r_1d}\alpha_{r_1d}+1}\\ +\frac{x_{sr_2}y_{r_2d}\alpha_{r_2d}}{x_{sr_2}+y_{r_2d}\alpha_{r_2d}+1}) + \lambda_d^{r_1}\alpha_{r_1d}+\lambda_d^{r_2}\alpha_{r_2d}-\nu_{r_1}(\alpha_{r_1d}-1)-\nu_{r_2}(\alpha_{r_2d}-1),\end{aligned} \tag{6.11}$$

where, λ_d^r and ν_r are non-negative Lagrangian multipliers. To make the equation brief, the signal to noise ratio (SNR) between source-relay, relay-destination, and source-destination are replaced by x_{sr}, y_{rd}, z_{sd}, respectively. Note that, for this

relaxed problem the strong duality holds, as the constraints satisfy the Slater's condition. Similar to [59] the Slater's condition for the relaxed problem can be trivially satisfied by considering any value of α_{rd} smaller than $\frac{1}{D}$. Now, The KKT conditions for the problem are:

$$\frac{g_{r_1 d}}{X} + \lambda_d^{r_1} = \nu_{r_1} \tag{6.12}$$

$$\lambda_d^{r_1} \alpha_{r_1 d} = 0 \tag{6.13}$$

$$\lambda_d^{r_1} \geq 0 \tag{6.14}$$

$$\frac{g_{r_2 d}}{X} + \lambda_d^{r_2} = \nu_{r_2} \tag{6.15}$$

$$\lambda_d^{r_2} \alpha_{r_2 d} = 0 \tag{6.16}$$

$$\lambda_d^{r_2} \geq 0 \tag{6.17}$$

where,

$$g_{r_i d} = \frac{(x_{sr_i} + y_{r_i d}\alpha_{r_i d} + 1)x_{sr_i} y_{r_i d} - x_{sr_i} y_{r_i d}^2 \alpha_{r_i d}}{(x_{sr_i} + y_{r_i d}\alpha_{r_i d} + 1)^2} \tag{6.18}$$

and

$$X = 1 + z_{sd} + \frac{x_{sr_1} y_{r_1 d}\alpha_{r_1 d}}{x_{sr_1} + y_{r_1 d}\alpha_{r_1 d} + 1} + \frac{x_{sr_2} y_{r_2 d}\alpha_{r_2 d}}{x_{sr_2} + y_{r_2 d}\alpha_{r_2 d} + 1} \quad . \tag{6.19}$$

Now let us assume that user d is getting assistance from both relays r_1 and r_2. Then both $\alpha_{r_1 d}$ and $\alpha_{r_2 d}$ will be nonzero and the KKT conditions in Eqs. (6.13) and (6.16) dictate that $\lambda_d^{r_1}$ and $\lambda_d^{r_2}$ are both zero. Replacing $\lambda_d^{r_1}$ and $\lambda_d^{r_2}$ with zero in Eqs. (6.12) and (6.15), respectively, we can write

$$\frac{\nu_{r_1}}{g_{r_1 d}} = \frac{\nu_{r_2}}{g_{r_2 d}} \quad . \tag{6.20}$$

For randomly generated networks, x_{sr}, y_{rd} and z_{sd} are independent continuous random variables. As a result, the probability of Eq. (6.20) being true is rather low

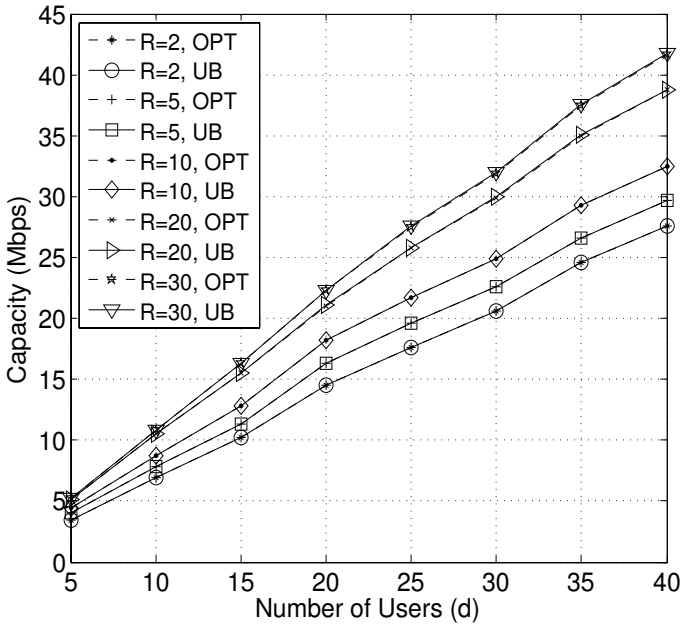


Figure 6.1: Capacity comparison between optimal and upper-bound solutions

and the upper-bound solution is quite tight. This insight remains true when we have multiple users and relay nodes in a network.

Fig. 6.1 shows the comparison between the optimal (OPT)[1] and the upper-bound (UB) solutions. The figure shows that, for smaller number of relays (e.g., $R = 2, 5, 10$), upper-bound solutions are almost identical to the optimal solutions. As the number of relays increases (e.g., $R = 20, 30$), the upper-bound solutions only very slightly outperform the optimal solutions. This tightness in upper-bound solutions suggest that only few conflicts in the solutions occur. To remove these conflicts and make the upper-bound solution feasible, for each user d we can then select relay r which allots the maximum power and assign that relay to that user and force all the other relays not to allocate any power to the data flow forwarded to d.

[1]The optimal results shown in Fig. 6.1 are obtained by solving the optimal unicast model discussed in Section 6.2.

Algorithm 6.1 Water-filling Algorithm

Step 1: Solve optimization problem ((6.3)-(6.5)) by ignoring Eq. (6.4); that will achieve an upper-bound solution.
Step 2: For each user d find the relay r which allots the maximum power $\max_{r \in R}\{\alpha_{rd}\}$.
Step 3: Re-formulate the optimization problem ((6.3)-(6.5)) by adding Eqs. (6.21) and (6.22) instead of Eq. (6.4).
Step 4: Solve the new convex optimization problem to achieve near optimal solution.

This can be ensured by the following two constraints.

$$\alpha_{rd} > 0 \qquad \text{if} \quad \alpha_{rd} = \arg\max_{j \in R}\{\alpha_{jd}\} \tag{6.21}$$

$$\alpha_{id} = 0 \qquad i \neq r, \quad \forall i \in \mathcal{R} \quad . \tag{6.22}$$

These two constraints make sure that not more than one relay is selected for a particular data flow. After this relay assignment is done for all the users, R water-filling problems can be solved for the power distribution at each of the relays. The complete algorithm is explained in Algorithm 6.1. This method has proven to be very fast and accurate (more on this in the results section).

6.4 Multicast Problem Formulation

Given a set of $\mathcal{M}$ ($|\mathcal{M}| = M$) multicast sessions, our objective is to maximize the performance of the network in terms of the overall multicast network capacity. Let B_r^m be the fraction of the total relay power used by relay r for a multicast session m and α_{rd}, like before, be the fraction of the total relay power used by relay r to forward data to user d. The optimization problem then can be written as

$$\text{Max} \sum_{m=1}^{M} C(m) \tag{6.23}$$

Subject to:

$$C(m) \leq C(d) \qquad \forall d \in \mathcal{S}^m, \forall m \in \mathcal{M} \tag{6.24}$$

$$B_r^m \geq \alpha_{rd} \qquad \forall d \in \mathcal{S}^m, \forall r \in \mathcal{R}, \forall m \in \mathcal{M} \tag{6.25}$$

$$\sum_{m=1}^{M} B_r^m \leq 1 \qquad \forall r \in \mathcal{R} \tag{6.26}$$

$$\alpha_{id} \times \alpha_{jd} = 0 \qquad i \neq j,\ i, j \in \mathcal{R},\ \forall d \in \mathcal{D} \tag{6.27}$$

$$0 \leq \alpha_{rd} \leq 1\ ;\ B_r^m \geq 0 \quad .$$

Eq. (6.23) is the objective function which aims to maximize the network capacity. The definition of Eq. (6.2) is modeled using constraint (6.24). The other difference between the unicast model and the multicast model is the introduction of variable B_r^m and two constraints (6.25) and (6.26). Since more than one user belonging to a particular multicast session m (i.e., $d \in \mathcal{S}^m$) may get assistance from relay r, Eq. (6.25) will make sure that none of the data flow using that relay and forwarded to that particular multicast group can exceed the power allocated to that multicast session. Eq. (6.26) makes sure that the combined power allocated to all multicast sessions going through a particular relay r does not exceed the total power budget of that relay. Again, the solution to this optimization problem ((6.23)-(6.27)) is made complicated by constraint (6.27). As an exhaustive search to find a solution for the problem involves testing all $R^{\sum_{m=1}^{M} |\mathcal{S}^m|}$ possible relay combinations, which is impossible for realistic user and relay sizes, solving the problem optimally involves removing constraint (6.27) and transforming the model into an optimal mixed Boolean-convex optimization problem. Therefore, like before, we introduce binary relay selection variable a_{rd} and replace Eq. (6.27) with Eqs. (6.6) and (6.7) and the multicast mixed Boolean-convex problem can be formulated as follows.

Optimal Multicast Model:

$$\text{Max} \sum_{m=1}^{M} C(m) \tag{6.28}$$

Subject to:

$$C(m) \leq C(d) \qquad \forall d \in \mathcal{S}^m, \forall m \in \mathcal{M} \tag{6.29}$$

$$B_r^m \geq \alpha_{rd} \qquad \forall d \in \mathcal{S}^m, \forall r \in \mathcal{R}, \forall m \in \mathcal{M} \tag{6.30}$$

$$\sum_{m=1}^{M} B_r^m \leq 1 \qquad \forall r \in \mathcal{R} \tag{6.31}$$

$$a_{rd} \geq \alpha_{rd} \qquad \forall d \in \mathcal{D}, \forall r \in \mathcal{R} \tag{6.32}$$

$$\sum_{r=1}^{R} a_{rd} \leq 1 \qquad \forall d \in \mathcal{D} \tag{6.33}$$

$$0 \leq \alpha_{rd} \leq 1 \ ; \ B_r^m \geq 0 \ ; \ a_{rd} = \{0, 1\} \ .$$

Similar to the unicast case, the mixed Boolean-convex problem for multicast traffic scenarios remains slow (solution involves B&B) for larger sized networks. In the next section, we discuss an algorithm that yields near optimal solutions for the problem relatively fast.

6.5 A Sequential Fixing Method for Multicast

One way to handle the issue of computation time in multicast traffic scenarios is by solving the problem using the water-filling technique described for the unicast problem. However, unlike unicast scenarios, it is not that straight forward to obtain a tight upper-bound as the KKT conditions in multicast scenarios do not demonstrate a similar kind of insight as the unicast scenarios. As a result, the number of conflicts in multicast traffic scenarios are much higher than the unicast case and selecting the best relay for each user based on the basis of maximum power allocation does

Table 6.1: Optimal vs. Water-filling technique in Multicast scenarios

M	Users(d)	Relays(R)	Opt(Mbps)	WF(Mbps)	Gap(%)
2	10	5	1.2732	1.0899	16.8181
2	20	15	1.5227	1.0727	41.9502
3	15	10	2.0387	1.6881	20.7689
3	30	20	3.0819	1.6073	91.7439
4	20	15	3.3311	2.3149	43.8982
4	36	20	3.9695	2.1666	83.2133

Algorithm 6.2 Sequential Fixing Algorithm

Step 1: Relax all Boolean variables a_{rd} to $0 \leq a_{rd} \leq 1$, which transform the problem into a convex optimization problem.
Step 2: Solve the convex problem; where solution of each a_{rd} being a value between 0 and 1.
Step 3: Among all the values of a_{rd} find the one with maximum value and fix it to 1.
Step 4: Re-formulate the convex problem with fixed a_{rd} value(s) and solve the problem. Note that, after fixing a variable $a_{rd} = 1$, Eq. (6.33) will make sure that no other relays will allocate any power to the data flow forwarded to d (i.e., $\alpha_{id} = 0;\ i \neq r,\ \forall i = 1, .., R$)
Step 5: Repeat Steps 2-4 unless all the Boolean variables a_{rd} are fixed. Note that after the first iteration, when Step 3 tries to find the variable with maximum value, the algorithm will not consider the variables which have already been fixed.
Step 6: Formulate and solve the convex problem based on all fixed "a"-values.

not work well in solving the problem. Table 6.1 shows the comparison between the optimal solutions (Opt) and solutions obtained using the water-filling (WF) technique in multicast scenarios. From Table 6.1, it is clear that, for multicast traffic, the results provided by WF are far from the optimal.

Indeed, the computation time issue arises from the fact that the Mixed Boolean-convex problem exploits the B&B technique (which is often quite slow) to find the value of the Boolean variables (a_{rd}). To overcome the problem, in this section, we propose an algorithm based on sequential fixing of the Boolean variables and obtain a near optimal solution. The main idea behind this algorithm is to fix the values of binary variables a_{rd} iteratively through solving a sequence of convex optimization

problems. In each iteration of the algorithm, we fix one binary variable from a_{rd}, where $r \in \mathcal{R}$ and $d \in \mathcal{D}$. Before the start of the first iteration, we relax all the binary variables a_{rd} to $0 \leq a_{rd} \leq 1$. As a result, the mixed Boolean-convex problem becomes a simple convex problem which can be easily solved. In each iteration, we solve the convex problem and find a_{rd} with the maximum value and set it to 1. After each iteration, we re-formulate the convex problem with fixed values of a_{rd}. Once a variable a_{rd} is fixed to 1, constraint (6.33) of the convex problem makes sure that no other relays will allocate any power to the data flow forwarded to d (i.e., $\alpha_{id} = 0; \quad i \neq r, \ \forall i \in \mathcal{R}$) in the next iteration. The solution of the problem will be obtained once we fix the values of all a_{rd}. The complete sequential fixing algorithm is presented in Algorithm 6.2. For our problem, the number of iterations through the sequence of convex optimization problems are equal to the number of users in the network.

6.6 Power Allocation with Discrete Levels

We assumed so far that the relay power can be shared among the sessions traversing through it and the relay can freely assign any power value to the traversing session. The only constraint is that the maximum power of the relay can not be exceeded. However, in practical cellular systems the power levels are quantized into discrete values. For example, in GSM (Global System for Mobile Communications), the transmission power (both uplink and downlink) usually varies from 5 to 33dBm with an equal spacing of 2dBm [103].

The objective of this section is to rewrite the joint relay selection and power allocation problem using the quantized power levels and solve the resulting optimization problem. Let $\mathcal{V}$ be the set of all power levels $\{P_i\}$, where $P_i = P_r \times \mu_i$, $0 \leq \mu_i \leq 1$ is the discrete fraction of relay power and $i = 1, 2, ..., |\mathcal{V}|$. We also define δ as the

power step size, where $\delta = \frac{1}{|\mathcal{V}|}$. Therefore, μ_i can be written as follows:

$$\mu_i = \mu_{i-1} + \delta \qquad \text{where } \mu_0 = 0 \text{ and } \mu_{|\mathcal{V}|} = 1 \ .$$

We introduce a new binary decision variable a^i_{rd}, where $a^i_{rd} = 1$, if relay r is selected to forward data to user d using power level μ_i and 0 otherwise. Hence, rewriting the objective in 6.3 yields

$$\sum_{d=1}^{D} CD(d) = \sum_{d=1}^{D} \frac{W}{2} \log_2(1 + \frac{P_s}{W\eta_0}|h_{sd}|^2 + \sum_{r=1}^{R} \frac{\frac{P_s}{W\eta_0}|h_{sr}|^2 \frac{P_r}{W\eta_0}|h_{rd}|^2 \sum_{i=1}^{|\mathcal{V}|} \mu_i a^i_{rd}}{\frac{P_s}{W\eta_0}|h_{sr}|^2 + \frac{P_r}{W\eta_0}|h_{rd}|^2 \sum_{i=1}^{|\mathcal{V}|} \mu_i a^i_{rd} + 1}) \ . \tag{6.34}$$

where, $CD(d)$ is the capacity of user d using discrete fraction of relay power. In the case of unicast traffic, the corresponding constraints (6.9) and (6.10) will be replaced by

$$\sum_{r=1}^{R} \sum_{i=1}^{|\mathcal{V}|} a^i_{rd} \leq 1 \qquad \forall d \in \mathcal{D} \tag{6.35}$$

$$\sum_{d=1}^{D} \sum_{i=1}^{|\mathcal{V}|} \mu_i a^i_{rd} \leq 1 \qquad \forall r \in \mathcal{R} \ . \tag{6.36}$$

Constraint (6.35) ensures that only one power level is used by the selected relay and that a user d can use at most one relay. Constraint (6.36) makes sure that the power allocation to all users going through a particular relay does not exceed the total power budget of that relay.

For multicast traffic, the mixed Boolean-convex problem will be as following:

$$\text{Max} \sum_{m=1}^{M} CD(m) \tag{6.37}$$

Subject to:

$$CD(m) \leq CD(d) \qquad \forall d \in \mathcal{S}^m, \forall m \in \mathcal{M} \tag{6.38}$$

$$B_r^m \geq \sum_{i=1}^{|\mathcal{V}|} \mu_i a_{rd}^i \qquad \forall d \in \mathcal{S}^m, \forall r \in \mathcal{R}, \forall m \in \mathcal{M} \tag{6.39}$$

$$\sum_{m=1}^{M} B_r^m \leq 1 \qquad \forall r \in \mathcal{R} \tag{6.40}$$

$$\sum_{r=1}^{R} \sum_{i=1}^{|\mathcal{V}|} a_{rd}^i \leq 1 \qquad \forall d \in \mathcal{D} \tag{6.41}$$

$$B_r^m \geq 0 \ ; \ a_{rd}^i = \{0, 1\} \ .$$

Eqs. (6.37 - 6.40) serve similar purpose as Eqs. (6.28 - 6.31) in the continuous power multicast model.

Clearly, the newly obtained mixed Boolean-convex optimization problems are much harder to solve using a B&B technique given the many fold increase ($|\mathcal{V}|$ times) in the number of binary variables a_{rd}^i. One way to solve this problem is by relaxing the values of a_{rd}^i to $0 \leq a_{rd}^i \leq 1$, and then solving the convex problem. From the obtained values of a_{rd}^i, select the highest and set it to 1. Successively, repeat this procedure until all values of a_{rd}^i are fixed to either 1 or 0. Note, however, that this technique is rather inefficient for the following reasons. First, note that not only does a_{rd}^i perform relay selection, but it also does power allocation. For example, when we consider 10 power levels, that is $|\mathcal{V}| = 10$, $\mu_i \in \mathcal{V}$, $\mu_1 = 0.1$, $\mu_{10} = 1$ and $\delta = 0.1$, $a_{rd}^i = 1$ means that the relay allocates power $P_r \times \mu_i$ for destination d. Suppose after solving the convex problem, $a_{rd}^{10} = 0.5$ is the largest among a_{rd}^is. Then, setting $a_{rd}^{10} = 1$ forces the relay r to allocate its maximum power to the session

whose destination is d, whereas only $0.5P_r$ is required for this session. Indeed, setting $a_{rd}^5 = 1$ would have been more efficient. In addition, there might be only half of the relay power left to be allocated (the other half might already been allocated in an earlier iteration) and forcing $a_{rd}^{10} = 1$ in this case will make the optimization problem infeasible. This flaw can be mitigated by observing that a_{rd}^i should not be treated as a fraction of relay power allocation, but rather $\mu_i a_{rd}^i$, which in turn should be quantized. That is, when $\mu_j < \mu_i a_{rd}^i < \mu_{j+1}, \quad \mu_j \in \mathcal{V}$, round down $\mu_i a_{rd}^i$ to the closest power level μ_j and fix $a_{rd}^j = 1$. Note that, we always round $\mu_i a_{rd}^i$ down to the closest quantization level as rounding up or rounding it to the nearest quantization level may raise the same infeasibility problem as we have discussed earlier. Hence, to solve the problem efficiently, only *Step 3* of the continuous power sequential fixing algorithm (Algorithm. 6.2) needs to be modified with the following four steps:

Step 3(a): Find the maximum value among all the values of a_{rd}^i.

Step 3(b): Multiply the obtained relaxed value of a_{rd}^i with μ_i to obtain the actual fraction of relay power allocated to user d.

Step 3(c): Using the obtained actual fraction of power, find the nearest (round down) power level μ_j.

Step 3(d): Set $a_{rd}^j = 1$.

6.7 Numerical Results

We present numerical results to evaluate the optimal design problems presented in Sections 6.2 and 6.4. In this section we also compare the optimal results with the near optimal solution provided by the algorithms presented in Sections 6.3 and 6.5. In our evaluation, we assume a source power per session P_s =100 mWatt. The total relay power at each relay is P_r =100 mWatt; the noise power is η_0 =-174dbm/Hz and fixed channel bandwidth W =200kHz. In our evaluation, we consider a circular

Table 6.2: Parameter used in COST-231 Model

Parameter	Value
BS Height	50m
Rooftop Height	30m
Relay Height	50m
User Height	1.5m
Frequency	1GHz
Road Orientation	90 degree
Building Spacing	50m
Street Width	12m

cell, centered at a BS, of radius one kilometer. The position of the users and relays are generated randomly inside the cell. All the other parameters to calculate channel response h_{ab} in COST-231 are presented in Table 6.2. The values of the parameters are taken from [59]. The convex optimization problems are implemented in MATLAB and solved using the TOMLAB (version 7.7) optimization solvers.

We first start with the unicast scenarios and compare the results of the WF with the optimal (OPT) one. We consider a total number of users d varying from 5 to 40 and we have evaluated the results by varying the number of relays R from 2 to 30. Fig. 6.2 shows the overall capacity comparison between the optimal and WF solutions and it becomes clear from the figure that WF provides very near to optimal solutions. For smaller number of relays, WF almost always provides optimal solutions. However, as the number of relays increases we observe a slight increase in the optimality gap but still very close to optimal. In our experiment, the maximum optimality gap we have observed is less than 2%. Next we compare the optimal mixed Boolean-convex model with WF in terms of computation time. Fig. 6.3 shows as the number of users and relays increases, the CPU time of optimal B&B technique increases substantially where the CPU time for WF remains quite low. While the optimal B&B technique needs hours of computation time to solve a problem of a larger size, it takes the WF technique only a few seconds.

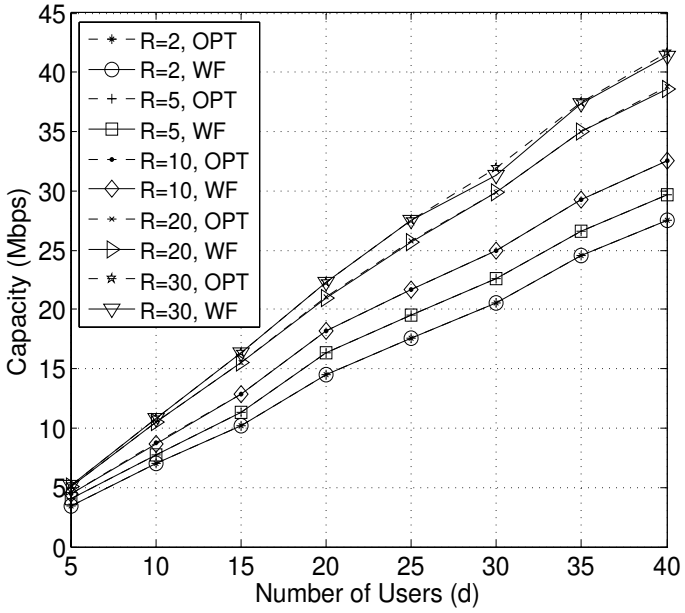


Figure 6.2: Capacity comparison between optimal and water-filling for unicast traffic scenarios

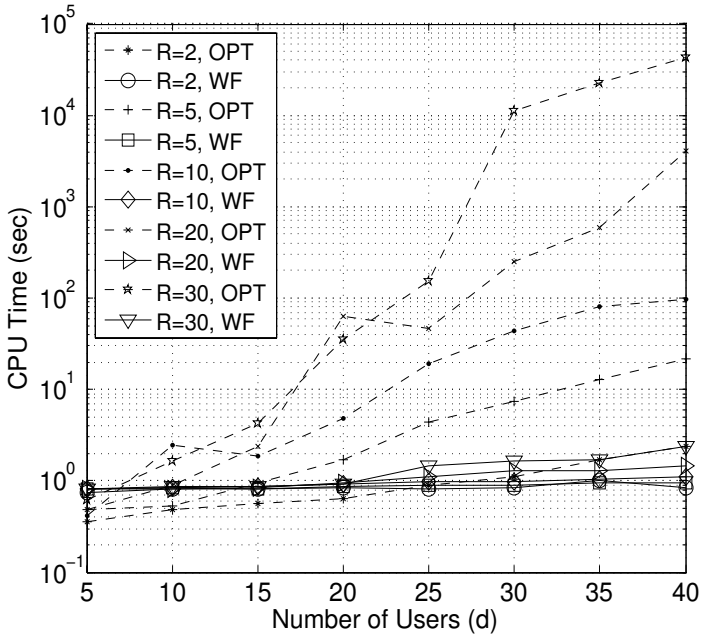


Figure 6.3: CPU time comparison between optimal and water-filling for unicast traffic scenarios

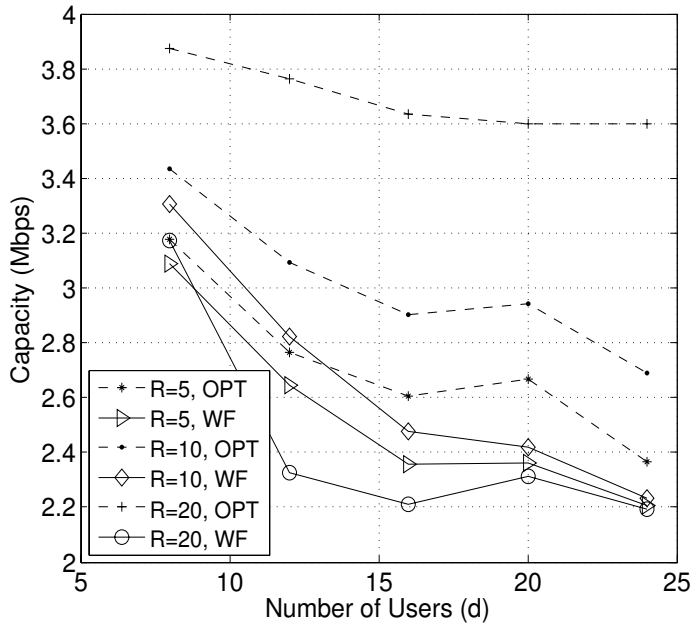
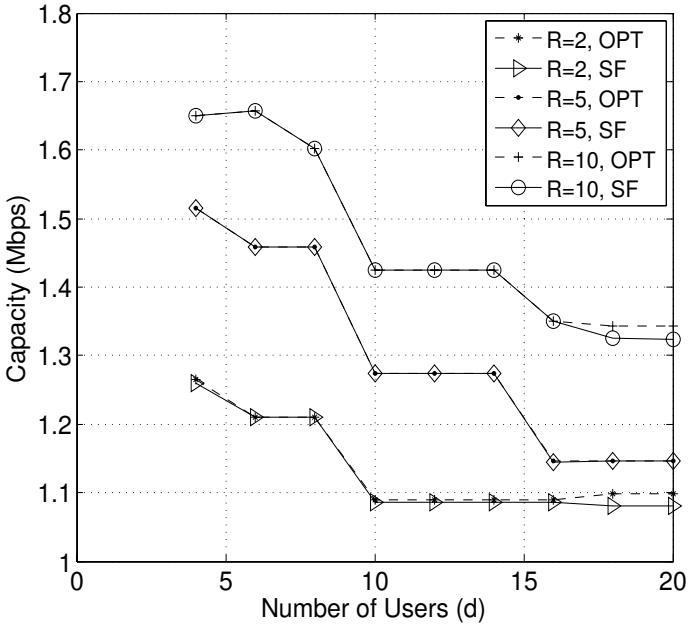


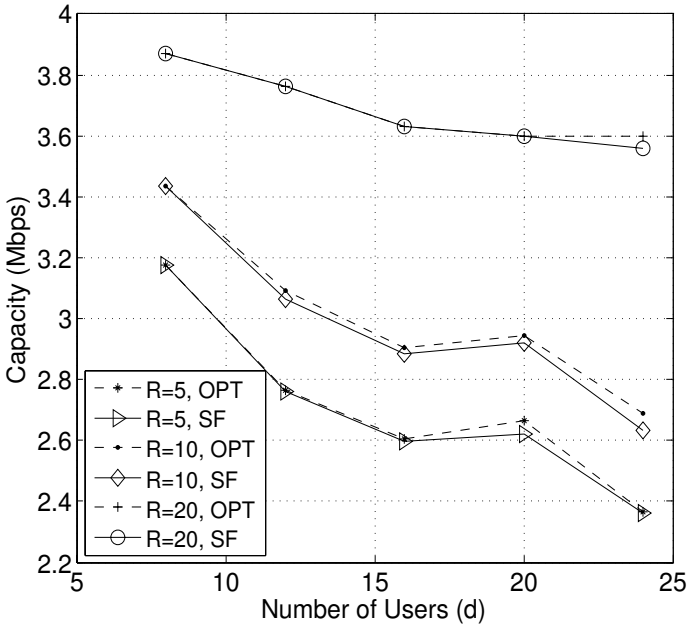
Figure 6.4: Capacity comparison between optimal and water-filling for multicast traffic scenario (M=4)

Next, we consider a multicast scenario and compare the results of the WF technique with the optimal solutions. We consider a multicast scenario with 4 multicast groups (i.e., M=4). In this scenario each multicast group contains a minimum of 2 to a maximum of 6 users with total number of users d varying from 8 to 24. We have evaluated the results by varying the number of relays R from 5 to 20. Fig. 6.4 shows the comparison between optimal and WF in terms of total network capacity. The figure confirms that in multicast unlike unicast traffic, the WF technique performs poorly. In this case the WF solutions remain far from optimal and as the number of relays increase the optimality gap of the WF solutions increases as well.

We now compare the optimal solution with its SF counterpart. We consider two multicast scenarios, first one with 2 multicast groups (Fig. 6.5(a)) and the other one with 4 multicast groups (Fig. 6.5(b)). Figs. 6.5(a) and 6.5(b) show that the SF algorithm provides very close to optimal results with highest optimality gap of less than 2.5%. From Fig. 6.5 we can also observe that, as the number of users increases



(a) Multicast group M=2



(b) Multicast group M=4

Figure 6.5: Capacity comparison between optimal and sequential fixing for Multicast traffic scenarios

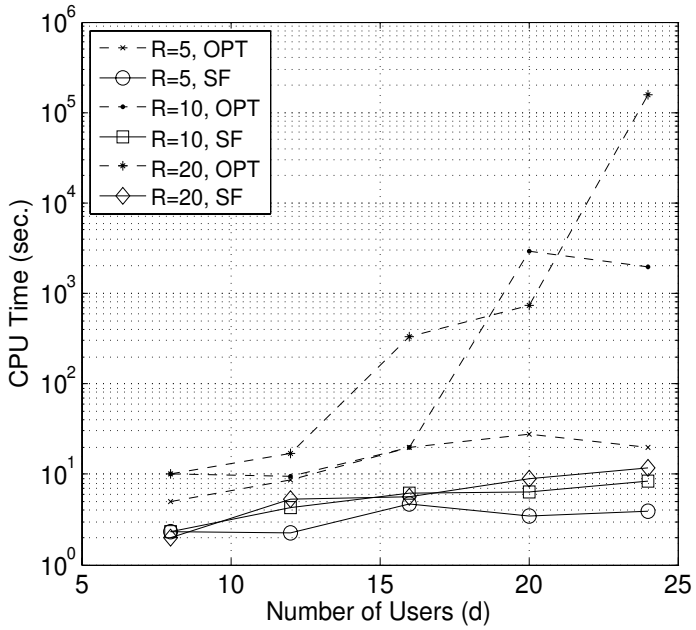
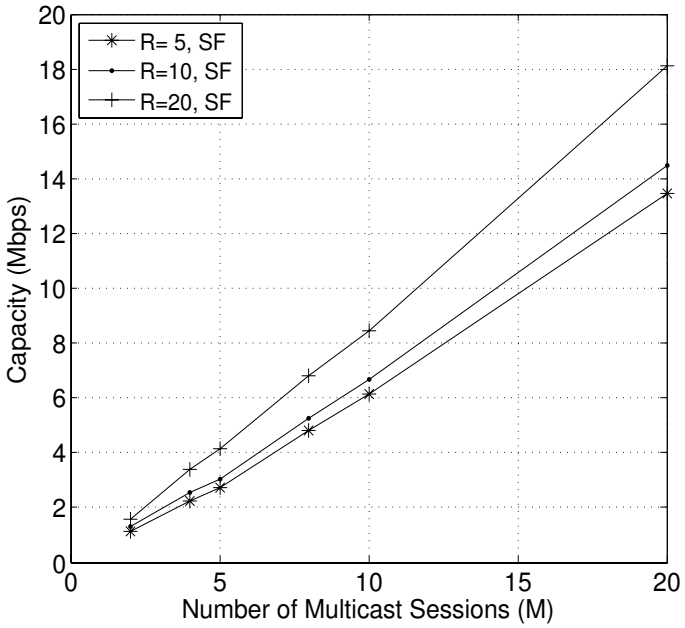


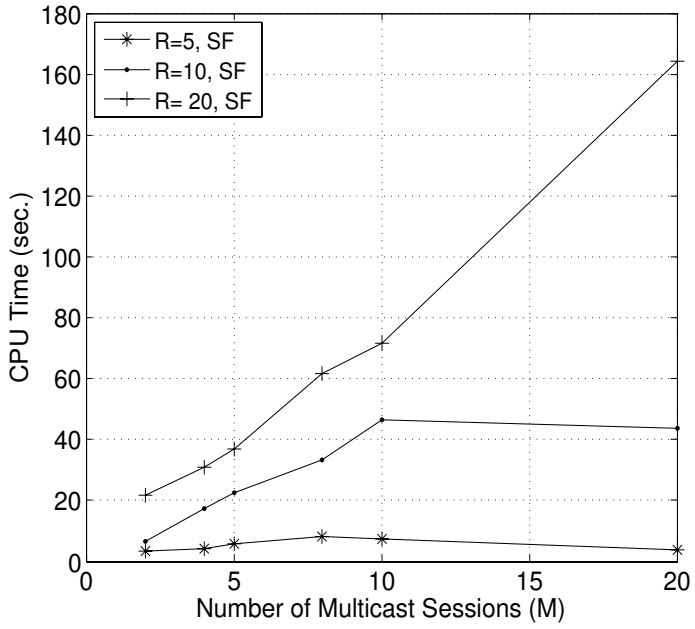
Figure 6.6: CPU time comparison between optimal and sequential fixing for multicast group (M=4)

the network capacity decreases. Multicast capacity is limited by the user with the weakest link and as a result of that more users in a multicast session decrease the session capacity as well as the network capacity. On the other hand, the running time of the optimal B&B model increases exponentially as the B&B technique slows down considerably with the increase of number of users and relays. Fig. 6.6 shows that while the running time of the mixed Boolean-convex model increases exponentially, sequential fixing finds near optimal solution very fast.

We then vary the number of multicast sessions from 2 to 20 while keeping the number of users fixed to 40. As we vary the number of relay nodes between 5 and 20, Figs. 6.7(a) and 6.7(b) show the network capacity and CPU time respectively. Note that, we use SF algorithm presented in Section 6.5 to obtain the results. We observe from the figures that as the number of session increases the total network capacity increases and for large number of multicast sessions and relay nodes, SF algorithm finds solutions in quite reasonable time.



(a) Capacity



(b) CPU time

Figure 6.7: Capacity and CPU time for multicast traffic

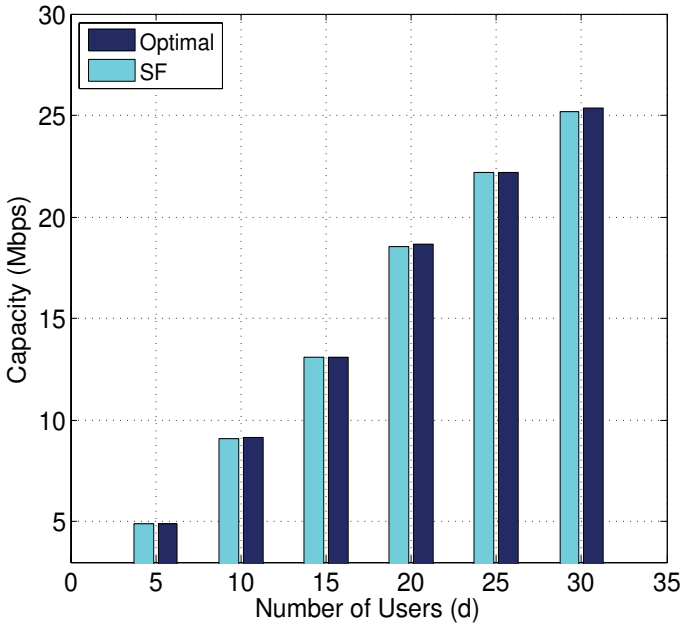
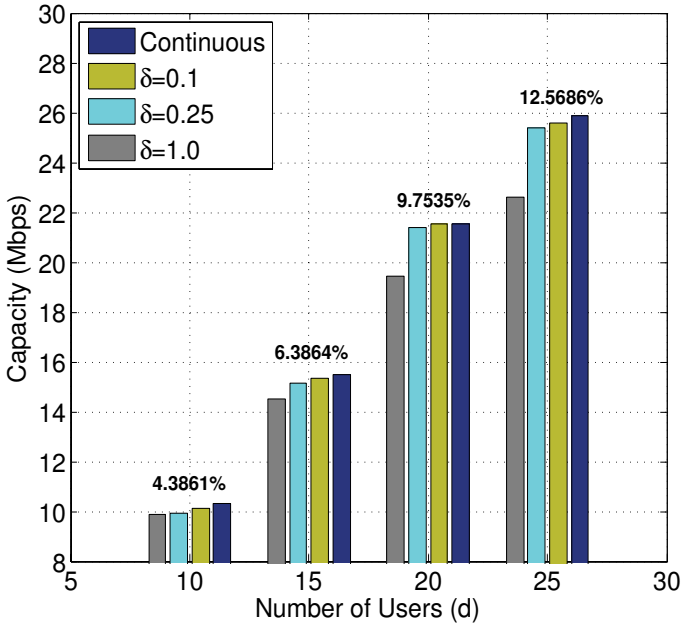


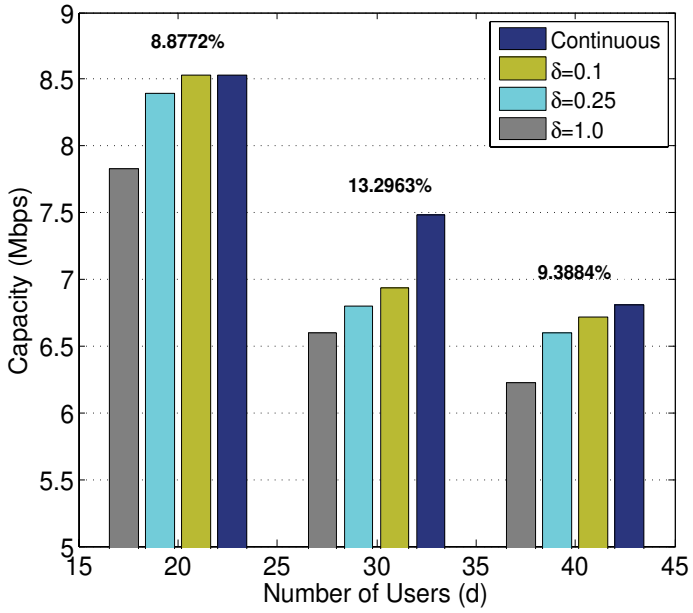
Figure 6.8: Capacity comparison between optimal and SF for discrete power levels

Next, we compare the optimal and the suboptimal model presented in Section 6.6. Fig. 6.8 shows the result for the unicast traffic scenarios with 5 relays ($R = 5$) and each relay has a maximum power P_r =500mWatt. We consider 4 discrete power levels with step size $\delta = 0.25$. We vary the number of users form 5 to 30. We notice from the figure that the suboptimal method based on SF provides a very close to optimal results. In this case, the gap between the optimal solution and solution achieved by the SF algorithm is always less than 1%.

Finally, we evaluate the network performance as we quantize the relay power allocation. We consider 3 different schemes by varying the total power levels and step sizes. The three scenarios are: $\delta = 0.1$ (10 power levels), $\delta = 0.25$ (4 power levels) and $\delta = 1$ (single power level). We consider both unicast traffic scenarios with 10 relays (Fig. 6.9(a)) and multicast traffic scenarios with 10 multicast groups ($M = 10$) and 5 relays (Fig. 6.9(b)). We can see from the figures that the single power level does not perform well against the continuous power allocation technique and



(a) Unicast traffic scenario



(b) Multicast traffic scenario

Figure 6.9: Capacity comparison between continuous and discrete power level allocation

their performance gaps for both unicast and multicast scenarios are presented in Figs. 6.9(a) and 6.9(b), respectively. Interestingly, both 4 and 10 level scenarios provide solutions closer to the continuous power allocation with 10 level scenario slightly outperforming the 4 level one. Note that, all the results provided for the discrete power allocation in Figs. 6.9(a) and 6.9(b) are obtained by solving the optimization problem with the SF algorithm presented in Section 6.6. The continuous power allocation for unicast traffic (Fig. 6.9(a)) is solved using the WF technique presented in Section 6.3 and the continuous power allocation for multicast traffic (Fig. 6.9(b)) is solved using the SF technique presented in Section 6.5.

6.8 Conclusion

We studied the joint problem of relay selection and power allocation in both wireless unicast and multicast cellular networks in terms of total network capacity. Our investigation showed that although the water-filling technique provides a near optimal solution for unicast traffic scenarios, it does not yield accurate solutions in multicast traffic scenarios. For both unicast and multicast traffic scenarios, we first modeled these combinatorially complex problems as mixed Boolean-convex optimization problems to maximize overall network capacity and solved them using the branch and bound technique. In both cases, this technique was proven to be computationally infeasible for large number of users and relays. We, therefore, presented a water-filling based technique for unicast and a sequential fixing technique for multicast and obtained near optimal solutions quite fast. We also proposed both optimal and sub-optimal problem formulations for joint relay and discrete power allocation techniques considering both unicast and multicast traffic scenarios. We then compared the results with continuous power allocation and showed that considering as small as four levels of discrete power yielded solutions close to the continuous power allocation.

Chapter 7

Conclusion and Future Directions

7.1 Conclusions

This thesis addresses two very important design aspects of wireless networks, namely, interference management and control through optimal cross-layer design and channel fading mitigation through relay-assisted cooperative communications.

We first presented a cross-layer formulation for the joint routing, scheduling and spectrum allocation problem in multi-channel multi-rate wireless mesh networks. We imposed optimal partitioning of the available spectrum band and divided it into a set of non-overlapping variable width channels. While narrower spectrum widths divide the total available spectrum into more non-overlapping channels allowing more parallel concurrent transmissions, wider bands have the effect, to either increase the transport capacity per link or reduce the SINR requirement to achieve the same link capacity. We presented two different formulations for solving the same problem; in the first one, we assume the link capacity as a function of channel bandwidth and in the second one, we assume fixed link capacity where the SINR threshold is a function of the channel bandwidth. We used column generation decomposition technique to exactly solve these combinatorially complex optimization problems. However, for

larger network sizes finding the exact solution is rather difficult as the computational complexity of the problem increased exponentially. We therefore proposed a greedy heuristic to solve the pricing subproblem and therefore reduced the complexity of the problem. We showed that the reduced complexity problem can be solved efficiently and provides near-optimal solutions.

Next, we investigated a more flexible spectrum access technique in multihop wireless networks with software defined radios. Unlike the previous work, we did not impose optimal partitioning of the available spectrum band into a set of non-overlapping channels rather we let the cross-layer design decide on the channel bandwidth positions. In this way a more flexible allocation of bandwidth was possible, since the transmissions used overlapping channels as well. We mathematically formulated the joint problem of scheduling and spectrum allocation and decomposed the problem using column generation technique, under the SINR interference constraints. Our investigation showed that the associated computation complexity is prohibitive from obtaining the optimal solution. We therefore presented a simulated annealing based approach for solving the pricing and is based on a simplified interference model; we augmented it however with an SINR check to make sure that only feasible configurations are used towards obtaining the optimal solutions. Our results indicated that the simulated annealing method yields solutions that are very close to optimal and the computation time is substantially reduced. Our results also showed that this flexible spectrum access technique effectively managed the interference in the network and improved the system performance.

Finally, we studied the joint problem of relay selection and power allocation in both wireless unicast and multicast cellular networks in terms of total network capacity. Our investigation showed that although the water-filling technique provides

a near optimal solution for unicast traffic scenarios, it does not yield accurate solutions in multicast traffic scenarios. For both unicast and multicast traffic scenarios, we first modeled these combinatorially complex problems as mixed Boolean-convex optimization problems to maximize overall network capacity and solved them using the branch and bound technique. In both cases, this technique was proven to be computationally infeasible for large number of users and relays. We, therefore, presented a water-filling based technique for unicast and a sequential fixing technique for multicast and obtained very near to optimal solutions quite fast. We also proposed both optimal and suboptimal problem formulations for joint relay and discrete power allocation techniques considering both unicast and multicast traffic scenarios. We then compared the results with continuous power allocation.

7.2 Future Work

The work presented in this thesis provided considerable performance enhancements of wireless multi-hop and cooperative cellular networks. However, there remain several future research directions of immense interest, especially in the domain of cooperative cellular networks.

The resource allocation problem in cooperative cellular network in this thesis considered only single cell and co-channel interference were not taken into account. The conventional non-cooperative approach to interference, via partitioning spatial reuse, prevents the reuse of any spectral resource within a certain cluster of cells. However, current designs do allow for full frequency reuse in each cell (typically for Code Division Multiple Access (CDMA) or frequency hopping spread spectrum systems) but this results in very severe interference conditions at the cell edge, causing a significant data rate drop at the terminals and a strong lack of fairness across cell users.

Inter-cell interference, which is a fundamental limiting factor in wireless cellular networks can be mitigated in future cellular networks by coordinating multiple BSs, where BSs only share channel state information (CSI). This has been proposed as a major technique to mitigate co-channel interference, since it shifts the signal processing burden to the BSs [104].

Although there has been some work done in resource allocation for BS coordinated multi-cell OFDMA system, almost all the existing works solved the problem sub-optimally using heuristics or dividing the problem into multiple subproblems and solving them iteratively. There are still many open research problems and the existing literature is still far from providing satisfactory solutions for many of them. We have identified the following problems as further research directions:

- Addressing the problem of coordinated relay selection, power and subcarrier allocation scheme in multi-cell OFDMA system across multiple BSs.

- Inclusion of effective time slot allocation strategy into the problem formulation to alleviate the co-channel interference and maximize the average weighted system throughput (bits/s/Hz/base station) by optimally allocating the network resources.

- Extend the idea of coordinated relay selection, power and sub-carrier allocation of multi-cell OFDMA system for multicast traffic. Since multicast capacity is different than unicast capacity, a multicast model and throughput metric which capture all these different aspects need to be developed.

Bibliography

[1] FCC, "Connecting america: The national broadband plan," March 16 2005, `http://www.broadband.gov/plan/#read-the-plan`.

[2] Y. Yang, H. Hu, J. Xu, and G. Mao, "Relay technologies for WiMax and LTE-advanced mobile systems," *IEEE Communications Magazine*, vol. 47, no. 10, pp. 100 –105, october 2009.

[3] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, December 2004.

[4] Q. Zhao and B. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 79–89, May 2007.

[5] B. Johansson, P. Soldati, and M. Johansson, "Mathematical decomposition techniques for distributed cross-layer optimization of data networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1535 –1547, August 2006.

[6] M. Chiang, S. Low, A. Calderbank, and J. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 255 –312, January 2007.

[7] V. Srivastava and M. Motani, "Cross-layer design: a survey and the road ahead," *IEEE Communications Magazine*, vol. 43, no. 12, pp. 112–119, December. 2005.

[8] V. Corvino, L. Giupponi, A. Perez Neira, V. Tralli, and R. Verdone, "Cross-layer radio resource allocation: The journey so far and the road ahead," in *Second International Workshop on Cross Layer Design (IWCLD '09)*, 2009, pp. 1–6.

[9] U. Kozat, I. Koutsopoulos, and L. Tassiulas, "A framework for cross-layer design of energy-efficient communication with QoS provisioning in multi-hop wireless networks," in *Proc. IEEE International Conference on Computer Communications (INFOCOM'04)*, vol. 2, 2004, pp. 1446 – 1456.

[10] I. Akyildiz and X. Wang, "A survey on wireless mesh networks," *IEEE Communications Magazine*, vol. 43, no. 9, pp. S23 – S30, September 2005.

[11] A. Mishra, E. Rozner, S. Banerjee, and W. Arbaugh, "Exploiting partially overlapping channels in wireless networks: Turning a peril into an advantage," in *Proc. Internet Measurment Conference (IMC '05)*, 2005, pp. 311–316.

[12] E. Rozner, Y. Mehta, A. Akella, and Q. L., "Traffic-aware channel assignment in enterprise wireless lans," in *Proc. IEEE International Conference on Network Protocols (ICNP '07)*, 2007, pp. 133–143.

[13] A. Mohsenian-Rad and V. Wong, "Partially overlapped channel assignment for multi-channel wireless mesh networks," in *Proc. IEEE International Conference on Communications (ICC '07)*, 2007, pp. 3770 –3775.

[14] R. Chandra, R. Mahajan, T. Moscibroda, R. Raghavendra, and P. Bahl, "A case for adapting channel width in wireless networks," in *Proc. ACM SIGCOMM '08*, 2008, pp. 135– 146.

[15] T. Moscibroda, R. Chandra, Y. Wu, S. Sengupta, P. Bahl, and Y. Yuan, "Load-aware spectrum distribution in wireless lans," in *Proc. IEEE International Conference on Network Protocols (ICNP '08)*, 2008, pp. 137–146.

[16] R.Gummadi, R. Patra, H. Balakrishnan, and E. Brewer, "Interference avoidance and control," in *Proc. 7th ACM Workshop on Hot Topics in Networks (Hotnets-VII)*, 2008, pp. 1–6.

[17] L. Cao, L. Yang, and H. Zheng, "The impact of frequency agility on dynamic spectrum sharing," in *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum (DySPAN'10)*, 2010, pp. 1–12.

[18] Z. Feng and Y. Yang, "Joint transport, routing and spectrum sharing optimization for wireless networks with frequency-agile radios," in *Proc. IEEE International Conference on Computer Communications (INFOCOM '09)*, 2009, pp. 1665–1673.

[19] T. Shu and M. Krunz, "Coordinated channel access in cognitive radio networks: A multi-level spectrum opportunity perspective," in *Proc. IEEE International Conference on Computer Communications (INFOCOM'09)*, 2009, pp. 2976 – 2980.

[20] Y. T. Hou, Y. Shi, and H. Sherali, "Optimal spectrum sharing for multihop software defined radio networks," in *Proc. IEEE International Conference on Computer Communications (INFOCOM'07)*, 2007, pp. 1–9.

[21] J. She, F. Hou, P. Ho, and L. Xie, "IPTV over WiMAX: Key success factors, challenges, and solutions [advances in mobile multimedia]," *IEEE Communications Magazine*, vol. 45, no. 8, pp. 87–93, August 2007.

[22] S. Kota, Y. Qian, E. Hossain, and R. Ganesh, "Advances in mobile multimedia networking and QoS [guest editorial]," *IEEE Communications Magazine*, vol. 45, no. 8, pp. 52–53, August 2007.

[23] H. Zhao and W. Su, "Cooperative wireless multicast: performance analysis and power/location optimization," *IEEE Transactions on Wireless Communications*, vol. 9, no. 6, pp. 2088–2100, June 2010.

[24] K. Vardhe, D. Reynolds, and B. Woerner, "Joint power allocation and relay selection for multiuser cooperative communication," *IEEE Transactions on Wireless Communications*, vol. 9, no. 4, pp. 1255–1260, April 2010.

[25] T. Wang and L. Vandendorpe, "Sum rate maximized resource allocation in multiple DF relays aided OFDM transmission," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 8, pp. 1559–1571, September 2011.

[26] ——, "WSR maximized resource allocation in multiple DF relays aided OFDMA downlink transmission," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3964–3976, August 2011.

[27] J. Laneman and G. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2415– 2425, October 2003.

[28] S. Kadloor, "Cooperative relaying in cellular networks," Master's thesis, University of Toronto, 2009.

[29] A. Bletsas, A. Khisti, D. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 659–672, March 2006.

[30] E. Beres and R. Adve, "Selection cooperation in multi-source cooperative networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 1, pp. 118–127, January 2008.

[31] Y. Zhao, R. Adve, and T. Lim, "Improving amplify-and-forward relay networks: optimal power allocation versus selection," *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, pp. 3114–3123, August 2007.

[32] S. M. Das, "A multi-layer approach towards high-performance wireless mesh networks," Ph.D. dissertation, Purdue University, 2007.

[33] "MIT roofnet," `http://www.pdos.lcs.mit.edu/roofnet`.

[34] "Southampton wireless network," `http://www.sown.org.uk`.

[35] "Wireless leiden," `http://www.wirelessleiden.nl`.

[36] "Seattle wireless," `http://www.seattlewireless.net`.

[37] S. Gollakota, S. D. Perli, and D. Katabi, "Interference alignment and cancellation," in *Proc. ACM SIGCOMM'09*, 2009, pp. 159–170.

[38] A. Raniwala, K. Gopalan, and T. Chiueh, "Centralized channel assignment and routing algorithms for multi-channel wireless mesh networks," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 8, no. 2, pp. 50–65, April 2004.

[39] M. Alicherry, R. Bhatia, and L. E. Li, "Joint channel assignment and routing for throughput optimization in multiradio wireless mesh networks," *IEEE Journal*

on Selected Areas in Communications, vol. 24, no. 11, pp. 1960–1971, November 2006.

[40] A. Mohsenian-Rad and V. Wong, "Joint logical topology design, interface assignment,channel allocation, and routing for multi-channel wireless mesh networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 12, pp. 4432–4440, December 2007.

[41] J. Zhang, H. Wu, Q. Zhang, and B. Li, "Joint routing and scheduling in multi-radio multi-channel multi-hop wireless networks," in *Proc. 2nd International Conference on Broadband Networks (BroadNets '05)*, 2005, pp. 631– 640.

[42] S. Merlin, N. Vaidya, and M. Zorzi, "Resource allocation in multi-radio multi-channel multi-hop wireless networks," in *Proc. IEEE International Conference on Computer Communications (INFOCOM '08)*, 2008, pp. 610–618.

[43] V. Bhandari and N. Vaidya, "Capacity of multi-channel wireless networks with random (c, f) assignment," in *Proc. ACM International Symposium on Mobile Adhoc Networking and Computing (MobiHoc '07)*, 2007, pp. 229 – 238.

[44] M. Kodialam and T. Nandagopal, "Characterizing the capacity region in multi-radio multi-channel wireless mesh networks," in *Proc. ACM International Conference on Mobile Computing and Networking (MobiCom '05)*, 2005, pp. 73 – 87.

[45] L. Chen, Q. Zhang, M. Li, and W. Jia, "Joint topology control and routing in ieee 802.11-based multiradio multichannel mesh networks," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 5, pp. 3123–3136, September 2000.

[46] J. Tang, G. Xue, and W. Zhang, "Interference-aware topology control and QoS routing in multi-channel wireless mesh networks," in *Proc. ACM International*

Symposium on Mobile Adhoc Networking and Computing (MobiHoc '05), 2005, pp. 68 – 77.

[47] W. Z. J. Tang, G. Xue, "Cross-layer optimization for end-to-end rate allocation in multi-radio wireless mesh networks," *ACM Wireless Networks*, vol. 15, no. 1, pp. 53–64, January 2009.

[48] A. Mishra, V. Shrivastava, S. Banerjee, and W. A. Arbaugh, "Partially overlapped channels not considered harmful," in *Proc. ACM SIGMETRICS '06*, 2006, pp. 63–74.

[49] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network," in *Proc. ACM International Symposium on Mobile Adhoc Networking and Computing (MobiHoc '04)*, 2004, pp. 187 – 201.

[50] L. Yang, L. Cao, H. Zheng, and E. Belding, "Traffic-aware dynamic spectrum access," in *Proc. 4th International Conference on Wireless Internet (WICON '08)*, 2008, pp. 1–9.

[51] I. F. Akyildiz, W. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, September 2006.

[52] Y. Yuan, P. Bahl, R. Chandra, T. Moscibroda, and Y. Wu, "Allocating dynamic time-spectrum blocks in cognitive radio networks," in *Proc. ACM International Symposium on Mobile Adhoc Networking and Computing (MobiHoc '07)*, 2007, pp. 130–139.

[53] Y. Hou, Y. Shi, and H. Sherali, "Spectrum sharing for multi-hop networking with cognitive radios," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 146–155, January 2008.

[54] P. Soldati and M. Johansson, "Network-wide resource optimization of wireless OFDMA mesh networks with multiple radios," in *Proc. IEEE International Conference on Communications (ICC '07)*, 2007, pp. 4979–4984.

[55] N. Jindal, J. G. Andrews, and S. Weber, "Bandwidth partitioning in decentralized wireless networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 5408–5419, December 2008.

[56] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. Part I and Part II," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927 – 1948, November 2003.

[57] S. Sharma, Y. Shi, Y. Hou, and S. Kompella, "An optimal algorithm for relay node assignment in cooperative Adhoc networks," *IEEE/ACM Transactions on Networking*, vol. 19, no. 3, pp. 879 –892, June 2011.

[58] Y. Shi, S. Sharma, Y. T. Hou, and S. Kompella, "Optimal relay assignment for cooperative communications," in *Proc. ACM international symposium on Mobile Adhoc networking and computing (MobiHoc '08)*, 2008, pp. 3–12.

[59] S. Kadloor and R. Adve, "Relay selection and power allocation in cooperative cellular networks," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 1676–1685, May 2010.

[60] D. Michalopoulos and G. Karagiannidis, "Performance analysis of single relay selection in rayleigh fading," *IEEE Transactions on Wireless Communications*, vol. 7, no. 10, pp. 3718 –3724, October 2008.

[61] J. Chu, R. Adve, and A. Eckford, "Using the bhattacharyya parameter for design and analysis of cooperative wireless systems," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1384 –1395, March 2009.

[62] S. Sharma, Y. Shi, Y. Hou, H. Sherali, and S. Kompella, "Cooperative communications in multi-hop wireless networks: Joint flow routing and relay node assignment," in *IEEE International Conference on Computer Communications (INFOCOM '10)*, 2010, pp. 1–9.

[63] X. Zhang, M. Hasna, and A. Ghrayeb, "Performance analysis of relay assignment schemes for cooperative networks with multiple source-destination pairs," *IEEE Transactions on Wireless Communications*, Accepted, October 2011.

[64] X. Zhang, A. Ghrayeb, and M. Hasna, "On relay assignment for network-coded cooperative systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 3, pp. 868–876, March 2011.

[65] J. Cai, X. Shen, J. Mark, and A. Alfa, "Semi-distributed user relaying algorithm for amplify-and-forward wireless relay networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 4, pp. 1348–1357, April 2008.

[66] K. Phan, D. Nguyen, and T. Le-Ngoc, "Joint power allocation and relay selection in cooperative networks," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM '09)*, 2009, pp. 1–5.

[67] T. Chiu-Yam Ng and W. Yu, "Joint optimization of relay strategies and resource allocations in cooperative cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 328 –339, February 2007.

[68] W. Dang, M. Tao, H. Mu, and J. Huang, "Subcarrier-pair based resource allocation for cooperative multi-relay OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 9, no. 5, pp. 1640–1649, May 2010.

[69] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Transactions on Communications*, vol. 54, no. 7, pp. 1310–1322, July 2006.

[70] J. Yuan and Q. Wang, "Adaptive resource allocation schemes for multiuser OFDMA nonregenerative relay networks," in *Proc. IEEE International Conference on Communications (ICC '10)*, 2010, pp. 1–5.

[71] Z. Hasan, E. Hossain, and V. Bhargava, "Resource allocation for multiuser OFDMA-based amplify-and-forward relay networks with selective relaying," in *Proc. IEEE International Conference on Communications (ICC '11)*, 2011, pp. 1–6.

[72] L. Le and E. Hossain, "Cross-layer optimization frameworks for multihop wireless networks using cooperative diversity," *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, pp. 2592 –2602, July 2008.

[73] I. Maric and R. Yates, "Cooperative multihop broadcast for wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1080–1088, August 2004.

[74] ——, "Cooperative multicast for maximum network lifetime," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 127–135, January 2005.

[75] Y. Hong and A. Scaglione, "Energy-efficient broadcasting with cooperative transmissions in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 5, no. 10, pp. 2844–2855, October 2006.

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

[77] G. Dantzig, *Linear Programming and Extensions.* Princeton University Press, 1998.

[78] M. S. Bazaraa, J. J. Jarvis, and H. D. Sherali, *Linear programming and network flows.* Wiley India Pvt. Ltd., 2008.

[79] V. Chvatal, *Linear Programming.* Freeman, 1983.

[80] G. L. Nemhauser and L. A. Wolsey, *Integer and Combinatorial Optimization.* Wiley, 1999.

[81] A. Schrijver, *Integer Programming.* John Wiley & Sons Inc, 1998.

[82] L. A. Wolsey, *Theory of linear and integer programming.* New York: Wiley, 1998.

[83] S. Boyd and J. Mattingley, "Branch and bound methods," notes for EE364b, Stanford University, available online at http://www.stanford.edu/class/ee364b/notes/bb_notes.pdf.

[84] J. Yuan, Z. Li, W. Yu, and B. Li, "A cross-layer optimization framework for multihop multicast in wireless mesh networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 11, pp. 2092 –2103, November 2006.

[85] P. Gilmore and R. Gomory, "A linear programming approach to the cutting stock problem," *Operations Research*, vol. 9, no. 6, pp. 849–859, December 1961.

[86] ——, "A linear programming approach to the cutting stock problem Part II," *Operations Research*, vol. 11, no. 6, pp. 94–120, December 1963.

[87] P. Bjorklund, P. Varbrand, and D. Yuan, "Resource optimization of spatial TDMA in ad hoc radio networks: a column generation approach," in *Proc.*

IEEE International Conference on Computer Communications (INFOCOM '03), 2003, pp. 818– 824.

[88] A. Capone, J. Elias, and F. Martignon, "Joint routing and scheduling optimization in wireless mesh networks with directional antennas," in *Proc. IEEE International Conference on Communications (ICC '08)*, 2008, pp. 2951–2957.

[89] S. Kompella, J. Wieselthier, A. Ephremides, and H. Sherali, "A cross-layer approach to end-to-end routing and SINR-based scheduling in multi-hop wireless networks," in *Proc. 6th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks and Workshops (WiOPT '08)*, 2008, pp. 261–266.

[90] J. Luo, C. Rosenberg, and A. Girard, "Engineering wireless mesh networks: Joint scheduling, routing, power control, and rate adaptation," *IEEE/ACM Transactions on Networking*, vol. 18, no. 5, pp. 1387 –1400, October 2010.

[91] M. Johansson and L. Xiao, "Cross-layer optimization of wireless networks using nonlinear column generation," *IEEE Transactions on Wireless Communications*, vol. 5, no. 2, pp. 435 – 445, February 2006.

[92] J. Desrosiers and M. Lbbecke, "A primer in column generation," in *Column Generation*, G. Desaulniers, J. Desrosiers, and M. Solomon, Eds. Springer, 2005, pp. 1–32.

[93] C. Joo, X. Lin, and N.B.Shroff, "Understanding the capacity region of the greedy maximal scheduling algorithm in multi-hop wireless networks," in *Proc. IEEE International Conference on Computer Communications (INFOCOM '08)*, 2008, pp. 1103 –1111.

[94] T. Lin and J. Hou, "Interplay of spatial reuse and SINR-determined data rates in CSMA/CA-based, multi-hop, multi-rate wireless networks," in *Proc. IEEE International Conference on Computer Communications (INFOCOM '07)*, 2007, pp. 803–811.

[95] "Cplex. using the cplex callable library (v. 9.1.3)," in *CPLEX Optmization Inc.*, 2005.

[96] V. Shrivastava, S. Rayanchu, J. Yoon, and S. Banerjee, "802.11n under the microscope," in *Proc. ACM SIGCOMM '08*, 2008, pp. 105–110.

[97] Y. Shi, Y. T. Hou, J. Liu, and S. Kompella, "How to correclty use the protocol interference model for multi-hop wireless networks," in *Proc. ACM International Symposium on Mobile Adhoc Networking and Computing (MobiHoc '09)*, 2009, pp. 239–248.

[98] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, May 1983.

[99] D. E. Knuth, J. H. Morris, and V. R. Pratt, "Fast pattern matching in strings," *SIAM Journal on Computing*, vol. 6, no. 2, pp. 323–350, 1977.

[100] C. Liu and J. Andrews, "Multicast outage probability and transmission capacity of multihop wireless networks," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4344–4358, July 2011.

[101] C. Gao, Y. Shi, Y. Hou, H. Sherali, and H. Zhou, "Multicast communications in multi-hop cognitive radio networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, pp. 784 –793, April 2011.

[102] "Multi-hop relay system evaluation methodology," available online at http://ieee802.org/16/relay/docs/80216j-06_013r3.pdf.

[103] Y. Xing and R. Chandramouli, "Stochastic learning solution for distributed discrete power control game in wireless data networks," *IEEE/ACM Transactions on Networking*, vol. 16, no. 4, pp. 932–944, August 2008.

[104] O. Simeone, O. Somekh, Y. Bar-Ness, and U. Spagnolini, "Uplink throughput of TDMA cellular systems with multicell processing and amplify-and-forward cooperation between mobiles," *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, pp. 2942 –2951, August 2007.