Optimization and Algorithms for Wireless Networks: Enhancing Problem Solvability, Channel Bonding Under Demand Stochasticity, and Receiver Characteristic Awareness

Amr Nabil A. Abdelfattah

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Computer Engineering

Allen B. MacKenzie, Chair
Scott F. Midkiff
Sedki M. Riad
Hesham A. Rakha
Mustafa Y. ElNainay

December 5, 2017
Blacksburg, Virginia

Keywords: Wireless networks, problem solvability, channel bonding, stochastic optimization, receiver characteristic awareness

Optimization and Algorithms for Wireless Networks: Enhancing Problem Solvability, Channel Bonding Under Demand Stochasticity, and Receiver Characteristic Awareness

Amr Nabil A. Abdelfattah

ABSTRACT

5G networks appear on the horizon with distinguished Quality of Service (QoS) requirements such as aggregated data rate and latency. Managing such networks in either a distributed or centralized manner to best utilize the available scarce resources is still a big challenge. Better mechanisms are needed for resource allocation. In this dissertation, we discuss three distinct research problems related to this theme.

The first part addresses enhancing the solvability of network optimization problems. For the class of problems studied, we show that a traditionally-formulated model is insufficient from a problem-solving perspective. When the size of the problem increases, even state-of-the-art optimizers cannot obtain an optimal solution because of memory constraints. We show that augmenting the model with suitable additional constraints and structure enables the optimizer to derive optimal solutions, or significantly reduce the optimality gap.

The second problem is optimal channel bonding in wireless LANs under demand uncertainty. An access point (AP) can aggregate multiple contiguous channels to satisfy demand. We discuss how to optimally utilize available frequency bands under uncertainty in AP demand using two stochastic optimization frameworks: a *static* scheme which minimizes the total occupied bandwidth while satisfying the demand of each AP with probability at least β and an *adaptive* scheme that allows adaptability of the bandwidth allocation in response to the AP demand variations. Given its complexity, we propose a novel framework to solve the adaptive stochastic optimization prob-

lem efficiently.

The third problem is to allocate resources with receiver characteristic awareness in a multiple radio access technology environment. We propose a novel adjacent channel interference (ACI)-aware joint channel and power allocation framework that takes into account receiver imperfections arising due to (i) imperfect image frequency rejection and (ii) analog-to-digital converter aliasing. As the overall problem is in the form of Mixed-Integer-Linear-Programming (MILP) which is NP-hard, we develop an efficient algorithm to solve it.

Optimization and Algorithms for Wireless Networks: Enhancing Problem Solvability, Channel Bonding Under Demand Stochasticity, and Receiver Characteristic Awareness

Amr Nabil A. Abdelfattah

GENERAL AUDIENCE ABSTRACT

The applications of next generation wireless networks have distinct requirements such as high speed for video streaming, low delay for interactive applications, and scalability to manage huge numbers of wireless devices. Managing such networks is challenging given the scarcity of wireless resources. In this dissertation, we discuss three distinct research problems related to this theme. The first part addresses enhancing the solvability of network optimization problems. State-of-the-art commercial optimization tools are unable to solve these problems for reasonable network sizes. We propose multiple strategies that help the tool obtain optimal solutions quickly. The second part considers indoor wireless networks. For such a network, we propose a technique that matches the instantaneous resources allocated to each location in the network with the amount of data traffic currently at the location. The third part addresses a problem of a network with multiple wireless transmitters and receivers where each receiver suffers from interference from other transmitters differently. We develop an algorithm to allocate resources and adjust transmit power so that each pair can communicate while meeting a minimum required data rate. The three parts of the dissertation are useful in either saving resources and hence allowing more users to use the network, or providing higher service quality for wireless device users.

This material is based upon work supported by the National Science Foundation under Grant No. 1564148. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Dedication

To prophet Muhammad, peace be upon him, who taught me to be kind and beneficial to humanity.

Acknowledgments

All praise is due to Allah, I thank him and seek his help, guidance, and forgiveness. I would like to acknowledge all people who had a significant impact during my studies for the PhD degree.

First, I would like to express my sincere gratitude to Prof. Allen B. MacKenzie for his guidance, encouragement, flexibility, and open mindedness during my work on this dissertation. His comments and suggestions were extremely helpful in exploring the most promising research avenues.

I would like to thank Dr Mohammad J. Abdel-Rahman. I had the great opportunity to work closely to him on several research problems during which he helped me sharpening my research skills and widening my perspectives. I thank him for the countless hours spent on in-depth and fruitful technical discussions. Also, I would like to thank Aditya V. Padaki for the valuable discussions and help.

Special thanks are due to Prof. Hanif Sherali for his tremendous advice and support. I was honored to work closely to him and sharpen my research and technical writing skills through his invaluable comments and suggestions. Also, I would like thank Prof. Barbara M. P. Fraticelli for teaching me several courses on optimization. I would like to thank my advisory committee members, Prof. Scott F. Midkiff, Prof. Sedki M. Riad, Prof. Hesham Rakha, and Prof. Mustafa ElNainay for their support, guidance and valuable comments.

Last but not least, I would like to thank my mother, Mrs. Essmat Ahmed Ali for her patience, kindness and encouragement during the years of my PhD. Indeed, her influence on my life cannot

be expressed in words. Also, I would like to thank my wife, Amira Abdelhamid. Her impact on my personal life was a significant factor in being able to reach the final stage of my PhD on time. Finally, I would like to thank my father, Mr. Nabil (May Allah have mercy on him), my sister, Doaa, and her husband, Hesham Elhelw for all their love and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

5G networks appear on the horizon with strict Quality of Service (QoS) requirements for different applications [1]. Aggregated data rate (or area capacity) requirements are expected to increase by roughly $1000\times$ from 4G to 5G. Roundtrip latency requirements are anticipated to decrease by an order of magnitude (from 15 ms to 1 ms) to support specific applications such as mission-critical control and traffic safety [2]. Operating different networks with distinguished QoS requirements is generally difficult. It is particularly challenging to achieve this goal in the wireless part of the network given the scarcity of spectrum resources. Most of the recently proposed solutions are based on coexistence between multiple technologies (e.g. Long Term Evolution (LTE), Wi-Fi) [3–7]. Managing such networks in either a distributed or centralized manner to best utilize the available resources is still an open challenge. Even how to manage overlapping networks using similar access technologies is not fully understood. Consequently, better mechanisms for resource allocation are needed. In this chapter, we describe three research topics related to this theme, and then we summarize our contributions in this dissertation.

Optimal performance of wireless networks requires joint consideration and optimization across

multiple layers. Typically, these problems in the most complex form involve integer, binary, and continuous variables. In most cases, this leads to an overall formulation of a Mixed-Integer-Linear Program (MILP) [11, 16, 17]. Moreover, if non-linear terms appear in the formulation, advanced optimization techniques, such as the Reformulation-Linearization Technique (RLT) [12], enable the linearization of such terms, resulting in an equivalently reformulated MILP, which is more convenient for solution using powerful, robust available software. However, for large networks, the optimization tools are unable to handle the problem. Special techniques are then needed to tackle this challenge.

In the context of wireless local-area networks (WLANs) (IEEE 802.11 [9]), user association and dissociation from an access point (AP) are not deterministic. These processes were shown in the literature to follow specific stochastic distributions [10, 50]. Each AP is obligated to serve all connected users in a fair manner. Hence, it is important to consider uncertainty in AP demands when solving practical resource allocation problems. The naive, commonly adopted approaches for channel allocation under AP demand uncertainty are to optimize network performance for either *average* or *peak* demand values. Considering the average demand values does not provide any guarantees on satisfying a significant percentage of AP demands. This is because the ability to meet actual demands will depend on the probability that the actual demand will exceed its average value. On the other hand, considering the maximum demand values, although potentially meeting all AP demands, usually results in huge and unnecessary consumption of scarce channel resources. Consequently, a new channel allocation framework is needed in which the level of AP demand satisfaction is controllable. Stochastic optimization is a powerful tool to model and handle uncertainty in problem parameters [30].

In a general multi-radio-access-technology environment, Adjacent Channel Interference (ACI) plays a significant role in determining the performance of wireless networks. Strict overlay mask regulations and constant-width guard bands have been previously proposed to mitigate ACI. However, these strategies are insufficient when different technologies with different overlay mask regulations operate in the same band. In an experiment[1], the degradation in throughput in LTE and

[1]The experimental work has been done in WIRELESS@VT labs by Aditya V. Padaki.

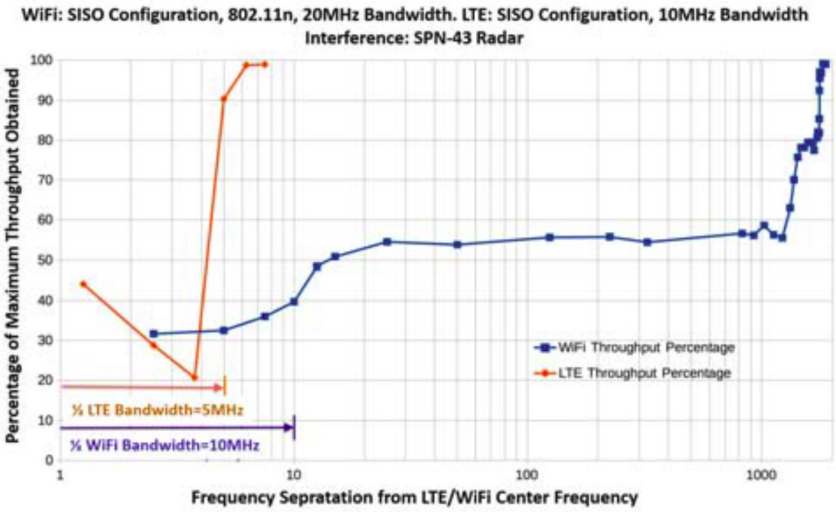


Figure 1.1: Experimental work shows the huge difference between LTE and WiFi receiver performance.

WiFi devices was measured when a SPN-43 radar signal was swept across frequency in the channels adjacent to the operating frequencies of these devices. Fig. 1.1 shows the susceptibility of LTE and WiFi receivers to the ACI coming from the radar signal. The figure demonstrates good out-of-band interference rejection capabilities for LTE receivers. However, WiFi receivers showed poor ACI tolerance, even when the adjacent channel signal is 100 MHz apart. This is mainly due to the typical inexpensive WiFi receivers, which have poor frequency selectivity and wide non-linear operating regions. Exploiting the knowledge of receiver characteristics will lead to better utilization of the available spectrum while minimizing ACI. This can be achieved through careful design of receiver characteristic-aware resource allocation strategies.

1.2 Summary of Contributions

We summarize the main contributions in this dissertation as follows.

- **Solvability of Network Optimization Problems.** Joint consideration across multiple layers is required to achieve optimal network performance. The general trend in solving these problems is to develop strong mathematical programming formulations that are capable

of providing near-optimal solutions to practical-sized problems. For the class of problems studied, we show that a traditionally formulated model turns out to be insufficient from a problem-solving perspective. When the size of the problem increases, even state-of-the-art optimizers cannot obtain an optimal solution because of insufficient memory. In Chapter 2, we show that augmenting the model with suitable additional constraints and structure enables the optimizer to derive optimal solutions, or significantly reduce the optimality gap, which were previously elusive given available memory restrictions.

- **Channel Bonding in Wireless LANs Under Demand Uncertainty.** Channel bonding is one promising approach to cope with rising WLAN data demand, given scarce spectrum resources. An access point (AP) can aggregate multiple contiguous channels to satisfy demand. In Chapter 3, we discuss how to optimally utilize available frequency bands under uncertainty in AP demands using two stochastic optimization frameworks: a *static* scheme which minimizes the total occupied bandwidth while satisfying the demand of each AP with probability at least β, and an *adaptive* scheme that allows adaptability of the bandwidth allocation in response to the AP demand variations. Given its complexity, we propose a novel framework to solve the adaptive stochastic optimization problem efficiently. The proposed framework exploits the special structure of the problem through decomposition into two subproblems. A particle swarm optimization (PSO)-based algorithm is tailored to the first-stage problem in order to obtain good solutions. The second-stage problem is further decomposed into several subproblems that can be solved independently in parallel. Our numerical results (i) demonstrate the advantages of stochastic compared to deterministic allocation, (ii) illustrate that the proposed framework reaches the optimal solution for the two-stage problem in few iterations, and (iii) explain the bandwidth–user satisfaction trade-off provided by the adaptive allocation approach.

- **Resource Allocation with Receiver Characteristic Awareness.** Next-generation wireless systems are expected to use multiple radio access technologies, with different receive and transmit characteristics, operating over the same band of spectrum in a spatial-temporal neighborhood. This will make the radio frequency (RF) front-ends susceptible to unprece-

dented adjacent-channel interference (ACI), which can jeopardize communication performance. In Chapter 4, we propose a novel ACI-aware joint channel and power allocation framework that takes into account receiver imperfections arising due to (i) imperfect image frequency rejection, and (ii) analog-to-digital converter aliasing. The objective is to minimize the number of allocated channels and the aggregate power transmitted while satisfying the rate demands of different links in a multi-RAT environment. As the overall problem is in the form of Mixed-Integer-Linear-Programming (MILP) which is NP-hard, we develop an efficient algorithm to solve it. The proposed algorithm decomposes the problem into two subproblems: channel allocation and power assignment. Then, it solves the decomposed problem in iteration using a two-phase structure. In Phase I, a PSO-based algorithm is tailored to obtain good solutions for the channel allocation subproblem. Phase II solves the power allocation subproblem optimally given the channel allocation solution obtained in Phase I. The algorithm alternates between the two phases until a stopping criterion is met. The results demonstrate (i) the criticality of receiver-characteristic awareness when designing resource allocation schemes for different types of networks, (ii) the ability of the proposed framework to control the trade-off between channel allocation and power assignment, and (iii) the superiority of the proposed algorithm in solving the overall problem efficiently and obtaining near-optimal solutions.

1.3 Dissertation Outline

The remainder of this dissertation is organized as follows: Chapter 2 discusses how to augment optimization models with suitable additional constraints and structure to enable the optimizer to derive optimal solutions, or significantly reduce the optimality gap, which were previously elusive given available memory restrictions. In Chapter 3, we discuss how to optimally utilize available frequency bands under uncertainty in AP demands using two stochastic optimization frameworks: a *static* scheme which minimizes the total occupied bandwidth while satisfying the demand of each AP with probability at least β, and an *adaptive* scheme that allows adaptability of the bandwidth

allocation in response to the AP demand variations. Due to the complexity of the adaptive scheme, we propose a novel framework to solve it efficiently. In Chapter 4, we propose a novel ACI-aware joint channel and power allocation framework that takes into account the receiver imperfections. Then, we develop an efficient algorithm to solve it. Chapter 5 summarizes the dissertation and indicates directions for future research.

Chapter 2

Enhancing the Solvability of Network Optimization Problems Through Model Augmentation

2.1 Introduction

Optimal performance of wireless networks requires joint consideration and optimization across multiple layers. Typically, these problems in the most complex form involve integer, binary, and continuous variables. At the network layer, rates of data sessions can be represented using continuous variables. At the medium access control (MAC) layer, scheduling can be done in either frequency or time domain if the available spectrum/time frame is fragmented into number of small divisions. In either case, a binary variable is needed to model the link activity between two nodes on specific frequency band or time slot. At the physical layer, adopting discrete power levels within power control strategies, and exploiting different technologies such as Multiple-Input-Multiple-Output (MIMO) and Interference Alignment (IA) mandates the use of binary and integer variables to correctly model their behavior. In most cases, this leads to an overall formulation of a Mixed-Integer-Linear Program (MILP) [11, 16, 17]. Moreover, if non-linear terms appear in

the formulation, advanced optimization techniques, such as the Reformulation-Linearization Technique (RLT) [12], enable the linearization of such terms, resulting in an equivalently reformulated MILP, which is more convenient for solution using powerful, robust available software.

State-of-the-art optimizers (such as CPLEX [13]) implement a wide range of techniques and methods to solve MILPs. Most of these algorithms are based on the well-known Branch-and-Bound (B&B) method [14]. In this method, a search tree is constructed by fixing one or more binary variable to the value of zero or one. For reasonable network sizes, the number of binary variables is relatively large. Consequently, the search tree of the optimizer eventually explodes if the problem instance is sufficiently challenging so that an optimal solution is not found during the early steps. Maintaining a large search tree requires huge amount of memory, which can be beyond traditional desktop machine capabilities. In such cases, the optimizer runs out of memory and fails to obtain an optimal/near-optimal solution. On the other hand, most (if not all) optimizers are designed as general-purpose tools to tackle optimization problems. That is, they are not tailored to efficiently solve specifically structured problems such as the class of wireless network problems at hand. When the network size is small, any optimizer can easily provide the optimal solution for the problem under study. However, for large networks, the optimization tools are unable to handle the problem (e.g., in [15], the solver could only solve the relaxed version of the MILP problem while in [16] and [17] simulations were limited to network size of 20 and 25 nodes, respectively).

One way to tackle this problem is to use a distributed procedure. Newton's method [18], among other efficient algorithms, can be adopted to solve linear and, more generally, convex optimization problems in a distributed way. However, the integrality restrictions on some variables in the MILP problem preclude a straightforward extension of this algorithm. Optimization software does not understand the networking problem itself; it recognizes the problem as an objective function with a set of variables and constraints. Although it generates different kinds of generic cuts, these cuts do not fully exploit the inherent physical structure of the problem.

This chapter makes the following contributions: We demonstrate how the structure of the networking problem can be exploited to generate effective specialized cuts (constraints). The basic

idea behind these cuts is to associate flows with the link activity variables based on the inherent nature of the problem. Moreover, we develop an effective strategy of introducing auxiliary binary variables to induce a specialized disjunctive constraint-based branching process. In the following sections, a case study is considered for an MILP problem and different strategies are introduced to implement this idea. It is worth mentioning here that the proposed special cuts and strategies are also applicable to any multi-hop network problem formulated as (or reduced to) an MILP problem. The introduced cuts in Section 2.3.3 are applicable to any multi-hop network problem having minimum rate requirements on some of its data flows.

The remainder of this chapter is organized as follows. Section 2.2 reviews the basic components of mathematically modeling a multi-hop network. Section 2.3 introduces our specialized techniques in details. In Section 2.4, we introduce a case study concerning an MILP formulation of a multi-hop network-based problem. In Section 2.5, we present our results. Section 2.6 concludes our work and indicates directions for future research.

2.2 Basic Mathematical Model for a Multi-hop Network

In this section, we review how we can mathematically model data flow balance and enforce data rate requirements of sessions in multi-hop networks. Consider a multi-hop network where a set $\mathcal{N}$ of wireless nodes are placed randomly in bounded area. A node i is capable of directly transmit and relay signals to a subset of the surrounding nodes $\mathcal{T}_i$ in its transmission range. Also, a subset of nodes $\mathcal{I}_j$ can overhear (being interfered) by a transmit node i if they fall inside the latter's interference range. A link $(i, j) \in \mathcal{L}$ from i to j is defined if and only if $j \in \mathcal{T}_i$, where $\mathcal{L}$ is the set of all links in the network. We assume that the time frame is divided into finite number of equal sized time slots T. Link activity at any time slot t is represented using a binary variable $x_{ij}[t]$. That is, $x_{ij}[t] = 1$ if node i transmits to node j during time slot t, and $x_{ij}[t] = 0$, otherwise. We assume a static environment where the wireless channel remains constant. Also, the interference in the network follows the protocol model [19]. In this model, a transmission is considered successful

if the receive node is inside the transmit node's transmission range, and outside the interference ranges of other non-intended simultaneous transmit nodes. Therefore, two interfering links cannot be activated simultaneously. Given these two assumptions, we can simply consider that each link (i, j) has a constant capacity C_{ij} when activated. There is a set of end-to-end sessions $\mathcal{M}$ to transfer data through the network. A session $m \in \mathcal{M}$ is defined by its source-destination pair $(s(m), d(m))$, and its data rate $r(m)$. Data flows of all sessions in the network are assumed to be steady and infinite. Without loss of generality, we also assume that each node has an infinite buffer to temporarily store the relayed data traffic. Table 2.1 lists the relevant notation used in this chapter. The different constraints described below model the basic behavior for a wireless multi-

Symbol	Definition
$\mathcal{N}_s$	Set of source nodes in the network
$\mathcal{N}_m$	Set of intermediate nodes in the network
$\mathcal{N}_d$	Set of destination nodes in the network
$\mathcal{N}$	$\mathcal{N}_s \bigcup \mathcal{N}_m \bigcup \mathcal{N}_d$, the set of all nodes in the network
$\mathcal{L}$	Set of all links in the network
$\mathcal{M}$	Set of all sessions in the network
$\mathcal{T}_i$	Set of nodes within the transmission range of node i
$\mathcal{I}_i$	Set of nodes within the interference range of node i
T	Total number of available time slots
$r(m)$	Data rate of session $m \in \mathcal{M}$
$s(m), d(m)$	Source and destination nodes of session $m \in \mathcal{M}$
$x_{ij}[t]$	Link activity indicator for the link (i, j) in time slot t
$f_{ij}(m)$	Data rate attributed to session $m \in \mathcal{M}$ on link (i, j)
C_{ij}	Capacity of link (i, j)

Table 2.1: Notation.

hop network.

Avoiding self- and mutual-interference: At any time slot t, if node i transmits signal to node j, it cannot transmit to nor receive from any other node. This can be expressed as follows:

$$\sum_{j \in \mathcal{T}_i} x_{ij}[t] + \sum_{k \in \mathcal{T}_i} x_{ki}[t] \leq 1, \quad i \in \mathcal{N}, t \in \{1, 2, \cdots, T\}. \tag{2.1}$$

To avoid mutual interference, when a node j receives signal from i at a time slot t, every node $p \neq i$, where $j \in \mathcal{I}_p$ should not be transmitting in the same time. The following constraint models this behavior.

$$\sum_{i \in \mathcal{T}_j} x_{ij}[t] + \sum_{q \in \mathcal{T}_p} x_{pq}[t] \leq 1, \quad p : j \in \mathcal{I}_p, p \neq i, j \in \mathcal{N}, t \in \{1, 2, \cdots, T\}. \tag{2.2}$$

Maintaining network flow balance: Denote $f_{ij}(m)$ as the data rate that is attributed to session m on link (i, j). We assume that flow splitting is allowed inside the network. This means that a data flow can split or merge at any node inside the network at the bit level. Then, the flow balance at each node can be maintained using the following constraints:

$$\sum_{j \in \mathcal{T}_i}^{j \neq s(m)} f_{ij}(m) = \sum_{k \in \mathcal{T}_i}^{k \neq d(m)} f_{ki}(m), m \in \mathcal{M}, i \in \mathcal{N}, i \neq s(m), d(m), \tag{2.3}$$

$$\sum_{j \in \mathcal{T}_i} f_{ij}(m) = r(m), \quad m \in \mathcal{M}, i = s(m), \tag{2.4}$$

$$\sum_{j : i \in \mathcal{T}_j} f_{ji}(m) = r(m), \quad m \in \mathcal{M}, i = d(m). \tag{2.5}$$

It is easy to show that the third constraint above can be derived from the former two. When a session is data rate requirement-restricted, $r(m)$ becomes a predetermined constant.

Link capacity: The total amount of data rate of different flows on link (i, j) cannot exceed its capacity C_{ij}. This can be represented using the following constraint:

$$\sum_{m \in \mathcal{M}}^{j \neq s(m), i \neq d(m)} f_{ij}(m) \leq \frac{1}{T} \sum_{t=1}^{T} C_{ij}.x_{ij}[t], \quad i \in \mathcal{N}, j \in \mathcal{T}_i. \tag{2.6}$$

Objective functions: In designing multi-hop wireless networks, several objectives can be considered. Maximizing the total data flow rates of all sessions in the network is one example. Another example is to maximize the minimum data flow rate in order to achieve fairness between sessions and avoid starvation. We assume a general utility function U to be maximized/minimized in order to express the complete problem formulation.

Traditional formulation: A general formulation of a multi-hop wireless network can be expressed as follows:

General Formulation of a Multi-hop Wireless Network

OPT
max/min U
subject to:
Self- and mutual-interference constraints: $(2.1), (2.2)$;
Network flow balance constraints: $(2.3), (2.4)$;
Link capacity constraints: (2.6).

In this formulation, $f_{ij}(m)$ and $r(m)$ are continuous variables, $x_{ij}[t]$ are binary variables, T and C_{ij} are constants. The problem is in the form of MILP.

2.3 Proposed Network Structure-Based Techniques

In this section, we derive our specialized valid inequalities (VIs) by considering the particular inherent special structures of a multi-hop network. We first examine the relationship between the sets of incoming/outgoing links associated with each node in the network. Second, we tie up the activation of links associated with each source-destination pair in the network. The last set of constraints relates the sessions with data rate requirements by generating suitable lower bounds on the number of active links associated with their source and destination nodes.

2.3.1 Relationships between nodes' incoming and outgoing links

Each physical link in the network between two nodes can be active during one or more time slots as long as this link activation does not interfere with other active links. The number of time slots during which this link needs to be active (having corresponding x-variables set at one) depends

on its capacity and the amount of data flow passing through. If we consider every single node in the network, explicit constraints can be derived by simply exploiting the relationship between the number of time slots during which the incoming and outgoing links associated with this node, are active. These constraints are presented in two different ways in the following subsections.

2.3.1.1 VIs based on nodes' incoming and outgoing links

Let $\mathcal{N}_s$ the set of source nodes, $\mathcal{N}_m$ the set of intermediate nodes excluding any node that acts as a source or destination to any session, $\mathcal{N}_d$ the set of destination nodes. Note that the data flow on each session is infinite. To keep the network stable, all data accumulated at any intermediate node should be forwarded within one time frame. Therefore, to simplify our analysis, we look into one time frame because the link activation schedule in other frames will be the same. Also, the time slot indices ignore the order of incoming and outgoing link activation. For example, consider a time frame length of ten time slots. If the outgoing link is activated during time slot 5, the incoming link can be activated at any other time slot other than 5 including the slots from 6 to 10. This is because what actually happens is that the data may arrive the node at time slot 7 and it will be forwarded to the next node in time slot 5 of the next frame.

For any multi-hop network, the following are implied:

1. For all intermediate nodes (excluding source and destination nodes of other sessions), if one of the incoming links is active during one of the available time slots, at least one of the outgoing links must be active during at least one of the available time slots, and vise versa. This can be expressed as follows:

$$\sum_{i:j\in\mathcal{T}_i}\sum_{t=1}^{T} x_{ij}[t] \geq 1 \Leftrightarrow \sum_{k\in\mathcal{T}_j}\sum_{t=1}^{T} x_{jk}[t] \geq 1, \forall j \in \mathcal{N}_m, \tag{2.7}$$

and can be formulated as shown below.

For each $j \in \mathcal{N}_m$:

$$\sum_{i:j\in\mathcal{T}_i} x_{ij}[t] \leq \sum_{k\in\mathcal{T}_j} \sum_{\acute{t}=1}^{T} x_{jk}[\acute{t}], \forall t \in \{1, 2, \cdots, T\}, \tag{2.8}$$

$$\sum_{k\in\mathcal{T}_j} x_{jk}[t] \leq \sum_{i:j\in\mathcal{T}_i} \sum_{\acute{t}=1}^{T} x_{ij}[\acute{t}], \forall t \in \{1, 2, \cdots, T\}. \tag{2.9}$$

2. For all source nodes, if one of the incoming links is active during one of the available time slots, at least one of the outgoing links must be active during at least one of the available time slots, but not vice versa. In this case, the source node acts as an intermediate node for another session. This can be expressed as follows:

$$\sum_{i:s\in\mathcal{T}_i} \sum_{t=1}^{T} x_{is}[t] \geq 1 \Rightarrow \sum_{k\in\mathcal{T}_s} \sum_{t=1}^{T} x_{sk}[t] \geq 1, \forall s \in \mathcal{N}_s, \tag{2.10}$$

which leads to the formulation given below.

For each $s \in \mathcal{N}_s$:

$$\sum_{i:s\in\mathcal{T}_i} x_{is}[t] \leq \sum_{k\in\mathcal{T}_s} \sum_{\acute{t}=1}^{T} x_{sk}[\acute{t}], \forall t \in \{1, 2, \cdots, T\}. \tag{2.11}$$

3. For all destination nodes, if one of the outgoing links is active during one of the available time slots, at least one of the incoming links must be active during at least one of the available time slots, but not vice versa. In this case, the destination node acts as an intermediate node for another session. This can be expressed as follows:

$$\sum_{k\in\mathcal{T}_d} \sum_{t=1}^{T} x_{dk}[t] \geq 1 \Rightarrow \sum_{i:d\in\mathcal{T}_i} \sum_{t=1}^{T} x_{id}[t] \geq 1, \forall d \in \mathcal{N}_d, \tag{2.12}$$

and can be formulated as specified below.

For each $d \in \mathcal{N}_d$:

$$\sum_{k \in \mathcal{T}_d} x_{dk}[t] \leq \sum_{i:d \in \mathcal{T}_i} \sum_{\acute{t}=1}^{T} x_{id}[\acute{t}], \forall t \in \{1, 2, \cdots, T\}. \tag{2.13}$$

These constraints, although simple and can be derived from the original constraints, are very unlikely to be automatically generated by the standard, generic methods implemented within the optimization software. However, augmenting the original formulation with these constraints may cause a slight degradation in the performance since the number of added constraints is $(|\mathcal{N}_s| + 2|\mathcal{N}_m| + |\mathcal{N}_d|) * T$, where $|\mathcal{N}_i|$ is the total number of nodes in the set $\mathcal{N}_i$. As such, when they do not offer significant reduction in the search space of the problem's feasible region, they can cause an overhead on the optimizer and negatively affect its performance. However, this occurs only in a few cases.

2.3.1.2 Introducing additional binary variables to facilitate special branching strategies

Although the derived formulation in Section 2.3.1.1 offers a tight model formulation, it sometimes does not provide satisfactory performance improvement. We therefore propose a new strategy of introducing certain binary variables to induce a disjunctive constraint-based branching using the developed set of cuts, which affords improved performance for some difficult instances. We motivate this strategy below and provide results in Section 2.5 to demonstrate its utility.

First, consider the set of constraints (2.8). By introducing additional binary variables, a revised set of constraints can be modeled as follows:

1. For each $j \in \mathcal{N}_m$:

$$\begin{aligned} &\sum_{i:j \in \mathcal{T}_i} x_{ij}[t] \leq z_j, \qquad \forall t \in \{1, 2, \cdots, T\}, \\ &\sum_{k \in \mathcal{T}_j} \sum_{t=1}^{T} x_{jk}[t] - z_j \geq 0, \end{aligned} \tag{2.14}$$

where $z_j \in \{0, 1\}$. It is straightforward to check the validity of (2.14) by considering the cases of $z_j = 0, 1$. That is, when $z_j = 0$, these constraints reduces to $\sum_{i:j\in\mathcal{T}_i} x_{ij}[t] \leq 0, \forall t \in \{1, 2, \cdots, T\}$, and $\sum_{k\in\mathcal{T}_j} \sum_{t=1}^{T} x_{jk}[t] \geq 0$. The second set of constraints becomes redundant but the first set enforces all $x_{ij}[t], \forall t \in \{1, 2, \cdots, T\}$, to have the value of zero. On the other hand, when $z_j = 1$, the constraints reduce to $\sum_{i:j\in\mathcal{T}_i} x_{ij}[t] \leq 1, \forall t \in \{1, 2, \cdots, T\}$, and $\sum_{k\in\mathcal{T}_j} \sum_{t=1}^{T} x_{jk}[t] \geq 1$. Here, the first set is redundant but the second set enforces that at least one of $x_{jk}[t]$ is equal to one.

Note that the addition of such superfluous binary variables to a model is atypical from a modeling perspective. However, this strategy turns out to be advantageous when done in the proposed fashion because it affords the opportunity for the solver to branch on certain key constraints (as opposed to just branching on variables as in the standard branch-and-bound/cut procedure) by virtue of the usual branching on the auxiliary binary variable. Indeed, this is evident by examining the effect of the disjunctive constraints imposed by (2.14) when considering the cases of z_j equal to zero and one.

Similarly, we can modify (2.9) as follows:
For each $j \in \mathcal{N}_m$:

$$\begin{aligned} &\sum_{k\in\mathcal{T}_j} x_{jk}[t] \leq y_j, \qquad \forall t \in \{1, 2, \cdots, T\}, \\ &\sum_{i:j\in\mathcal{T}_i} \sum_{t=1}^{T} x_{ij}[t] - y_j \geq 0, \end{aligned} \tag{2.15}$$

where $y_j \in \{0, 1\}$.

Likewise, we can derive similar constraints for the source and destination nodes:

2. For each $s \in \mathcal{N}_s$:

$$\begin{aligned} &\sum_{i:s\in\mathcal{T}_i} x_{is}[t] \leq z_s, \qquad \forall t \in \{1, 2, \cdots, T\}, \\ &\sum_{k\in\mathcal{T}_s} \sum_{t=1}^{T} x_{sk}[t] - z_s \geq 0, \end{aligned} \tag{2.16}$$

where $z_s \in \{0, 1\}$.

3. For each $d \in \mathcal{N}_d$:

$$\begin{aligned} &\sum_{k \in \mathcal{T}_d} x_{dk}[t] \leq y_d, \qquad \forall t \in \{1, 2, \cdots, T\}, \\ &\sum_{i:d \in \mathcal{T}_i} \sum_{t=1}^{T} x_{id}[t] - y_d \geq 0, \end{aligned} \tag{2.17}$$

where $y_d \in \{0, 1\}$.

The proposed auxiliary binary variables help the optimizer improve the partitioning process in the search tree. However, if the optimization tool does not benefit from such branching opportunities due to its internal heuristics, the increased dimension of the problem might slightly negatively impact its performance. In our experience, this deterioration in performance for certain instances is outweighed by the improvement achieved for other challenging instances.

2.3.2 VIs based on links of source and destination nodes

In this section, we jointly consider the activation of links associated with each session's source-destination node pair in the network. This consideration is under two conditions. The first condition is that at least one of the outgoing links of the session's source node is active during any time slot. The second condition is that this session's source node is not an intermediate node for any other session. Then, at least one of the incoming links to the session's destination node must be active during at least one time slot, and vice versa. Then, for each session, if these two conditions on the links associated with the source node are met, we can derive a restriction on the incoming links associated with the destination node. This can be mathematically expressed as follows:

Defining (s, d) as the source-destination pair of the session under consideration:

$$\left\{ \sum_{k \in \mathcal{T}_s} \sum_{t=1}^{T} x_{sk}[t] \geq 1 \,\&\, \sum_{i:s \in \mathcal{T}_i} \sum_{t=1}^{T} x_{is}[t] \leq 0 \right\} \Rightarrow \left\{ \sum_{i:d \in \mathcal{T}_i} \sum_{t=1}^{T} x_{id}[t] \geq 1 \right\}. \tag{2.18}$$

$$\left\{\sum_{i:d\in\mathcal{T}_i}\sum_{t=1}^{T}x_{id}[t]\geq 1\&\sum_{k\in\mathcal{T}_d}\sum_{t=1}^{T}x_{dk}[t]\leq 0\right\}\Rightarrow\left\{\sum_{k\in\mathcal{T}_s}\sum_{t=1}^{T}x_{sk}[t]\geq 1\right\}. \tag{2.19}$$

Focusing on (2.18), since both expressions are linear, non-negative and integer valued, $\sum_{k\in\mathcal{T}_s}\sum_{t=1}^{T}x_{sk}[t]\leq T$, and $\sum_{i:s\in\mathcal{T}_i}\sum_{t=1}^{T}x_{is}[t]\leq T$, we get that (2.18) is equivalent to the following:

$$\left\{\sum_{k\in\mathcal{T}_s}\sum_{t=1}^{T}x_{sk}[t]>0\&\sum_{i:s\in\mathcal{T}_i}\sum_{t=1}^{T}x_{is}[t]<1\right\}\Rightarrow\left\{\sum_{i:d\in\mathcal{T}_i}\sum_{t=1}^{T}x_{id}[t]\geq 1\right\}.$$

This in turn is equivalent to:

$$\left\{\sum_{i:d\in\mathcal{T}_i}\sum_{t=1}^{T}x_{id}[t]\geq 1\right\}\text{OR}\left\{\sum_{k\in\mathcal{T}_s}\sum_{t=1}^{T}x_{sk}[t]\leq 0\right\}\text{OR}\left\{\sum_{i:s\in\mathcal{T}_i}\sum_{t=1}^{T}x_{is}[t]\geq 1\right\}$$

which can be modeled as follows:

$$\begin{aligned}
&h_1+h_2+h_3=1, h\in\{0,1\},\\
&\sum_{i:d\in\mathcal{T}_i}\sum_{t=1}^{T}x_{id}[t]\geq h_1,\\
&\sum_{k\in\mathcal{T}_s}\sum_{t=1}^{T}x_{sk}[t]\leq(1-h_2)*T,\\
&\sum_{i:s\in\mathcal{T}_i}\sum_{t=1}^{T}x_{is}[t]\geq h_3.
\end{aligned} \tag{2.20}$$

Similarly, (2.19) can be modeled as follows:

$$\begin{aligned}
&g_1+g_2+g_3=1, g\in\{0,1\},\\
&\sum_{k\in\mathcal{T}_s}\sum_{t=1}^{T}x_{sk}[t]\geq g_1,\\
&\sum_{i:d\in\mathcal{T}_i}\sum_{t=1}^{T}x_{id}[t]\leq(1-g_2)*T,\\
&\sum_{k\in\mathcal{T}_d}\sum_{t=1}^{T}x_{dk}[t]\geq g_3.
\end{aligned} \tag{2.21}$$

The benefit from these constraints occurs when the source and/or destination node of a data session are not participating in other data sessions as an intermediate node. Otherwise, the added constraints may result in overhead on the overall formulation and might cause a slight degradation in the performance.

2.3.3 VIs for data rate requirement-restricted sessions

A data session is usually defined by its source-destination nodes pair in the network. We focus here on a data session with a minimum data requirement. The source node of such session can be a source of one or more other data rate requirement-restricted sessions. Similarly, its destination node can be a destination of one or more other data rate requirement-restricted sessions. As a traditional node in the multi-hop network, these source/destination nodes may relay other sessions' traffic in the network. Consequently, the amount of data transmitted from a source node is lower-bounded by the summation of the rates of sessions for which this is the source node. Similarly, the amount of data received by a destination node is lower-bounded by the summation of the rates of sessions for which this is the destination node. We can exploit this simple fact to derive special cuts as explained below.

Denote $\hat{\mathcal{M}}$ as the set of data rate requirement-restricted sessions (note that $\hat{\mathcal{M}} \bigcap \mathcal{M} = \phi$). Consequently, denote $r(\hat{m})$, $s(\hat{m})$ and $d(\hat{m})$ as the data rate, source and destination nodes of session $\hat{m} \in \hat{\mathcal{M}}$, respectively. Also, denote $f_{ij}(\hat{m})$ as the data rate that is attributed to data rate requirement-restricted session $\hat{m}$ on link (i, j). The capacity constraint for any source node $s \in \mathcal{N}_s$ is given as follows:

$$\sum_{\hat{m} \in \hat{\mathcal{M}}}^{k \neq s(\hat{m}), s \neq d(\hat{m})} f_{sk}(\hat{m}) + \sum_{m \in \mathcal{M}}^{k \neq s(m), s \neq d(m)} f_{sk}(m) \leq \frac{1}{T} \sum_{t=1}^{T} C_{sk}.x_{sk}[t], \qquad (k \in \mathcal{T}_s).$$

By summing both sides over $k \in \mathcal{T}_s$,

$$\sum_{k\in\mathcal{T}_s}\left(\sum_{\hat{m}\in\hat{\mathcal{M}}}^{k\neq s(\hat{m}),s\neq d(\hat{m})} f_{sk}(\hat{m}) + \sum_{m\in\mathcal{M}}^{k\neq s(m),s\neq d(m)} f_{sk}(m)\right) \leq \frac{1}{T}\sum_{t=1}^{T}\sum_{k\in\mathcal{T}_s} C_{sk}.x_{sk}[t].$$

As mentioned earlier, a lower bound on the LHS of the last inequality is given by the sum of $r(\hat{m})$ over all sessions $\hat{m} \in \hat{\mathcal{M}}$ for which s is the source.
Denoting this lower bound by H_s, we have:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{k\in\mathcal{T}_s} C_{sk}.x_{sk}[t] \geq H_s = \sum_{\hat{m}\in\hat{\mathcal{M}}:s=s(\hat{m})} r(\hat{m}).$$

Multiplying both sides by T and dividing both sides by $C_s^{\max} = \max\{C_{sk} : k \in \mathcal{T}_s\}$, we get:

$$\sum_{t=1}^{T}\sum_{k\in\mathcal{T}_s} \frac{C_{sk}}{C_s^{\max}}.x_{sk}[t] \geq \frac{T * H_s}{C_s^{\max}}.$$

Because $\frac{C_{sk}}{C_s^{\max}} \leq 1$, a Chvatal inequality [14] is given as follows:

$$\sum_{t=1}^{T}\sum_{k\in\mathcal{T}_s} x_{sk}[t] \geq \left\lceil \frac{T * H_s}{C_s^{\max}} \right\rceil. \tag{2.22}$$

Similarly, for any destination node $d \in \mathcal{N}_d$, defining $H_d = \sum_{\hat{m}\in\hat{\mathcal{M}}:d=d(\hat{m})} r(\hat{m})$, and $C_d^{\max} = \max\{C_{kd} : d \in \mathcal{T}_k\}$, we derive

$$\sum_{t=1}^{T}\sum_{k:d\in\mathcal{T}_k} x_{kd}[t] \geq \left\lceil \frac{T * H_d}{C_d^{\max}} \right\rceil. \tag{2.23}$$

Note that these constraints apply only to sources and destinations of data rate requirement-restricted sessions. Particularly, in cases when the number of such sessions passing through the same source/destination node increases, the proposed constraint becomes tighter and can thereby assist in enhancing performance.

2.4 A Case Study

In this work, we consider a cognitive radio (CR) network as a case study to evaluate the effectiveness of the proposed strategies. CR enables efficient spectrum sharing in wireless networks [20]. That is, the nodes of a primary network usually do not fully utilize the available spectrum all the time. Hence, secondary CR nodes communicate by exploiting the available opportunities in time, frequency, and space domains. The prevailing paradigm is to have completely uncooperative primary and secondary networks. When the primary and secondary networks are co-located geographically, a more cooperative paradigm is to let the secondary nodes help relay the primary nodes' traffic but not vice versa [21]. Yuan *et al.* introduced the concept of transparent coexistence of primary and secondary multi-hop networks in [22]. In that work, primary and secondary networks are completely coordinating. That is, each node in both networks may relay data from any node that belongs to the other network. The data rate requirement-restricted sessions in the context of this chapter are the primary sessions. The objective in that work was to maximize the minimum rate of the secondary sessions while maintaining all data rate requirements of the primary sessions. For the details of the model and description of the constraints, see [22].

2.5 Performance Evaluation

In this section, we present the performance of CPLEX (v12.6) [13] in solving the cut-enhanced optimization problem (augmented with the proposed cuts discussed in Section 2.3) compared to its performance when solving the original problem. The set of test cases consists of 55 randomly generated instances (combinations of Maximin and Maxisum[1] versions of the original problem), with 11 instances each of 30, 35, 40, 45, and 50-node networks. Each network has four active sessions: two primary and two secondary sessions where the source and destination of each were randomly selected. We used a cluster at VirginiaTech, called BlueRidge [23], to run our exper-

[1]The Maxisum version is very similar to the Maximin version except that the objective function in the former is to maximize the sum of the flow rates of secondary sessions instead of maximizing the minimum flow rate as in the latter.

Form. Index	Description
1	OPT_Maxisum/OPT_Maximin
2	OPT_Maxisum/OPT_Maximin;2.8;2.9;2.11;2.13
3	OPT_Maxisum/OPT_Maximin;2.14;2.15;2.16;2.17
4	OPT_Maxisum/OPT_Maximin;2.20;2.21
5	OPT_Maxisum/OPT_Maximin;2.22;2.23
6	OPT_Maxisum/OPT_Maximin;2.8;2.9;2.11;2.13;2.20;2.21
7	OPT_Maxisum/OPT_Maximin;2.8;2.9;2.11;2.13;2.22;2.23
8	OPT_Maxisum/OPT_Maximin;2.8;2.9;2.11;2.13;2.20;2.21;2.22;2.23
9	OPT_Maxisum/OPT_Maximin;2.14;2.15;2.16;2.17;2.20;2.21
10	OPT_Maxisum/OPT_Maximin;2.14;2.15;2.16;2.17;2.22;2.23
11	OPT_Maxisum/OPT_Maximin;2.14;2.15;2.16;2.17;2.20;2.21;2.22;2.23

Table 2.2: A summary of formulations.

iments. Each experiment was executed on a single node of BlueRidge that has 16 processors (utilized by CPLEX when possible) and 64GB memory. This hardware configuration is very similar to a traditional desktop machine so that any practitioner can use the proposed algorithms to run similar experiments without the need of state-of-the-art cluster capabilities. Each experiment is terminated when its run-time reaches 144 hours (this limitation comes from the rules enforced by the BlueRidge administration), reaches optimal solution, or runs out of memory, whatever happens first. For the sake of clarity, Table 2.2 summarizes all problem formulations tested in our experiments. As shown in the table, each formulation represents either one of the two versions of the original problem or one of the versions augmented with one or more of the proposed sets of cuts described in Section 2.3.

2.5.1 Recognizing hard instances

We define "hard" instances as the ones that CPLEX could not solve to optimality within the enforced computational limits. Consequently, these instances were run using formulations augmented with different combinations of the proposed cuts to test their relative effectiveness and performance improvement. In order to distinguish the hard instances from others in the test set,

each of the instance's statistics (number of binary variables, number of constraints, etc.) was correlated with the level of the instance's hardness. However, we could not derive a clear relationship using these statistics. On the other hand, we noticed that CPLEX significantly reduces the number of binary variables for some instances during the preprocessing step. We found a high correlation between the "reduced" number of binary variables and the difficulty of the instance. That is, if the resulting reduced number of binary variables is above a certain threshold (2000 in the problem under consideration), CPLEX could not solve it to optimality for almost all cases because it runs out of memory. As a result, this serves as a good test for deciding whether an instance should be augmented with one or more of the proposed techniques, or not (before attempting to solve it). In the following sections, we will focus only on the instances that CPLEX could not solve to optimality when implementing the original formulation.

2.5.2 Potential of the proposed formulations

Augmenting the original problem with only one set of cuts in Section 2.3.2 or 2.3.3 (Formulations 4 and 5) did not result in considerable improvement over the original Formulation 1. Consequently, all the results presented below will focus on the comparison between the performance of different representations of the logical implications in Sections 2.3.1.1 or 2.3.1.2 augmented by one or more of the proposed sets of cuts in Sections 2.3.2 and 2.3.3. Due to space limitations, we show detailed results for a few key formulations followed by comparative results for all formulations. We define the optimality gap for any maximization problem as follows:

$$\text{Optimality gap} = \frac{\text{UB-LB}}{\text{LB}} * 100\%$$

where LB (lower bound) is determined by calculating the objective value of the best obtained solution, i.e., the incumbent solution, and UB is the value of the LP-relaxation, which is an "upper bound" for the optimal solution of the MILP problem. We consider improvement/degradation in the performance if the optimality gap is decreased/increased by at least 5%, respectively.

Instance	Formulation 1		Formulation 2		Formulation 3	
	opt. gap	time (h)	opt. gap	time (h)	opt. gap	time (h)
2.3-1	118.03%	6.41	101.77%	12.5	optimal	2.63
2.3-2	156.86%	3.39	optimal	82.65	optimal	7.8
2.3-3	26.54%	12.03	13.67%	16.96	optimal	33.86
2.3-4	62.63%	5.70	40.93%	13.23	21.07%	24.63
2.3-5	290.18%	24.35	127.68%	39.64	95.1%	40.1

Table 2.3: Effect of augmenting the original formulation with the VIs w/ and w/o superfluous binary variables.

2.5.3 Detailed results

We discuss here the effect of introducing additional binary variables to the constraints in Section 2.3.1.1 (see Section 2.3.1.2). Table 2.3 shows some of the instances of different network sizes where the auxiliary binary variables significantly enhanced the performance. Formulation 2 helped CPLEX significantly reduces the optimality gap (Instances 2.3-1, 2.3-3, 2.3-4, and 2.3-5) or even reach the optimal solution (Instance 2.3-2). When introducing binary variables to the added cuts (Formulation 3), CPLEX attained optimality for most instances (and further reduced the optimality gap for others). More interestingly, for Instance 2.3-2, CPLEX obtained the optimal solution in 82.65 hours using Formulation 2. With Formulation 3, it attained optimality much quicker (7.8 hours). As discussed in Section 2.3.1.2, this improved performance when using the extra binary variables is likely because of the different type of branching opportunities that are afforded by these extra variables. We also show the significance of adding the cuts in Section 2.3.2 to Formulations 3 (Formulation 9). As shown in Table 2.4, the added cuts caused the optimality gap to further shrink and/or reach the optimal solution in shorter time. Similar results were obtained when the cuts in Section 2.3.3 were used (Formulation 10). Note that longer running time does not always mean worse performance. It can also mean that CPLEX has more opportunity (before running out of memory) to further reduce or close the optimality gap.

Instance	Formulation 1		Formulation 3		Formulation 9	
	opt. gap	time (h)	opt. gap	time (h)	opt. gap	time (h)
2.4-1	77.98%	9.08	optimal	19.56	optimal	5.89
2.4-2	98.79%	7.57	70.8%	29.02	50.4%	23.34
2.4-3	25.76%	9.55	16.62%	17.98	optimal	6.00
2.4-4	19.84%	11.12	optimal	47.74	optimal	34.93
2.4-5	23.19%	11.28	12.96%	26.50	8.19%	20.36

Table 2.4: CPLEX's performance with Formulations 3 and 9.

2.5.4 Comparative results

Table 2.5 summarizes statistics for the overall performance of different formulations. Here, "p1", "p2" and "p3" in the table refer to the logical implications in Sections 2.3.1, 2.3.2 and 2.3.3, respectively. The "+ve effect"/"-ve effect" columns represent the percentage of instances for which the optimality gap was enhanced/degraded, respectively. The "no effect" column shows the percentage of instances where there was no noticeable effect on the optimality gap. Table 2.6 compares the positive effect performance of the proposed formulations for different network sizes (represented by node count). Figure 2.1 shows CPLEX's behavior under Formulation 11 for all instances. From these results, we can deduce the following:

- Overall, for all the proposed formulations, we obtained significant improvement between $50-65\%$ of the instances, slight degradation for about 20%, and no noticeable effect on the remaining set of instances.
- The proposed formulations were most effective for networks of size between 30-45 nodes. When the number of nodes is less than 30, the problem size is small enough so that every instance can be solved to optimality using the original formulation. When the number of nodes exceeds 45, the increased problem difficulty suggests that further model or algorithmic enhancements are needed in this case. For networks of this size, the branch-and-bound tree gets relatively huge when approaching small values of optimality gap despite the tightening effect of the proposed cuts.

Form. Index	Description			+ve effect	-ve effect	no effect
	p1	p2	p3			
2	VIs			65.45%	21.82%	12.73%
6		✓		60%	23.64%	16.36%
7			✓	61.82%	21.82%	16.36%
8		✓	✓	60%	21.82%	18.18%
3	VIs w/ binary vars			58.18%	12.73%	29.09%
9		✓		60%	21.82%	18.18%
10			✓	60%	20%	20%
11		✓	✓	50.91%	18.18%	30.91%

Table 2.5: Overall performance of the formulations.

Form. Index	Node count				
	30	35	40	45	50
2	45.45%	90.91%	72.73%	72.73%	45.45%
6	72.73%	54.55%	81.82%	45.45%	45.45%
7	63.64%	90.91%	81.82%	45.45%	27.27%
8	45.45%	63.64%	72.73%	72.73%	45.45%
3	72.73%	63.64%	63.64%	54.55%	36.36%
9	45.45%	72.73%	63.64%	63.64%	54.55%
10	54.55%	54.55%	72.73%	45.45%	72.73%
11	63.64%	54.55%	72.73%	36.36%	27.27%

Table 2.6: Percentage of instances exhibiting enhanced performance.

In general, adding cuts helps by tightening relaxations, but also influences branching strategies and the performance of the solver's internal heuristics, which can have unpredictable effects if the augmented model representation is not aligned with the software's built-in algorithmic strategies. Likewise, introducing auxiliary binary variables to the formulation does enhance the performance by providing alternative constraint-based branching options, but adds some difficulty to the overall problem and causes "improvement degradation" in a few instances (where the imposed extra variables burden the formulation more that they assist it). However, overall, the proposed combination of strategies provide a significant impetus to resolving challenging problem instances that were otherwise unsolvable.

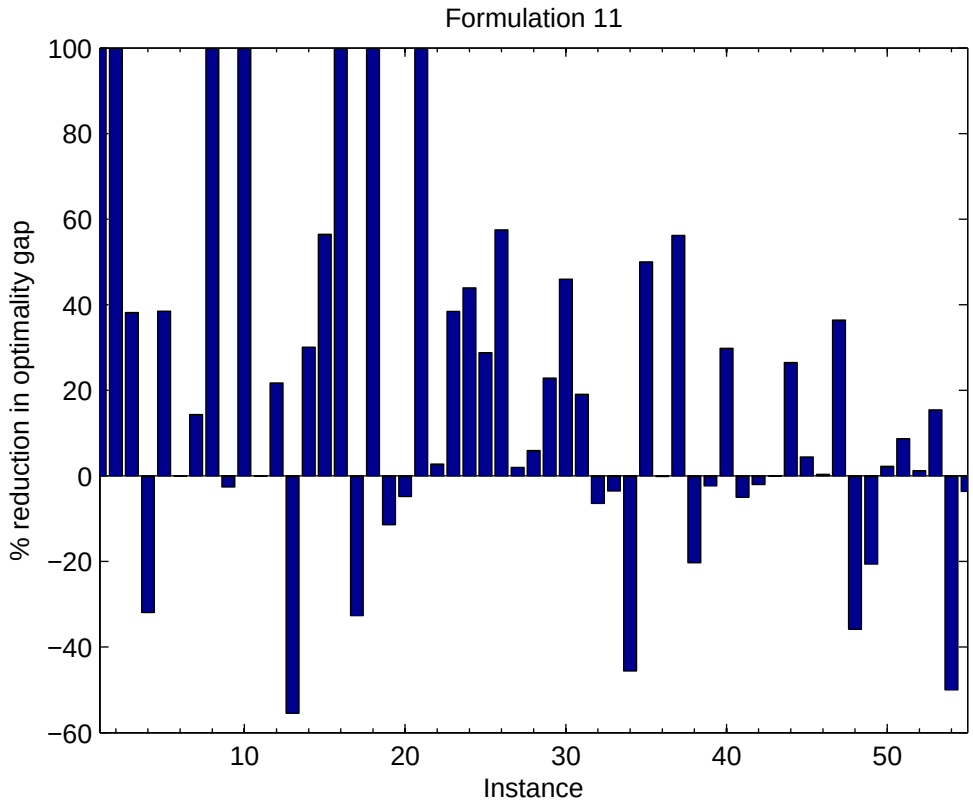


Figure 2.1: CPLEX behavior under Formulation 11.

2.6 Conclusions

In this chapter, we discussed different approaches to tackle the problem of excessive memory consumption when solving MILP problems. Generic formulations are often not sufficiently attractive from the problem-solving perspective. We demonstrated that generating special cuts through exploiting the structure of the problem offers a better strategy. In most cases, combining different kinds of special cuts outperformed the performance of the formulations that use these cuts individually (or not at all). Moreover, introducing auxiliary binary variables to provide partitioning opportunities based on these cuts, when applicable, significantly enhanced the performance for some instances. Overall, this work demonstrates how the use of proper combinations of model enhancement techniques can help optimize (or further reduce the optimality gap) for challenging

instances that were unsolvable using traditional formulations.

Chapter 3

A Stochastic Optimization Framework for Channel Bonding in Wireless LANs Under Demand Uncertainty

3.1 Introduction

Data transfer over wireless networks is dramatically increasing [25]. According to [26], by 2020 more than 55% of wireless cellular data will be offloaded to WLANs and small cells. In the context of WLANs, a mobile user is connected to the Internet through an access point (AP). Usually, an AP uses a single channel to communicate with its associated users. To improve the peak data rate of the AP when extra channels are available, a technique called "channel bonding" has been introduced.

Channel bonding (CB)– Channel bonding is a technique in which multiple contiguous non-overlapping channels can be combined into one wide channel. This can significantly increase the channel capacity. IEEE standards have facilitated this approach by introducing channel bonding in IEEE 802.11n and IEEE 802.11ac [27]. Initially, an AP is allocated a 20 MHz channel, called the

primary channel. The AP can then expand its bandwidth by seizing consecutive 20 MHz channels, called secondary channels, within the 5 GHz industrial, scientific, and medical (ISM) band, up to an aggregated bandwidth of 160 MHz. However, this requires these consecutive channels to be interference-free. Due to the nature of WLAN deployment, interference cannot be avoided between APs operating near each other on the same channel. Therefore, when allocating multiple consecutive channels to an AP, these channels cannot be utilized at the same time by a nearby AP. We use the term "interfering APs" to represent a subset of APs that can interfere with a chosen AP if operated on the same frequency channel.

Addressing AP demand uncertainty– In [28], the authors suggested using channel bonding to better utilize the spectrum and satisfy the deterministic demands of a set of APs. In this chapter, we consider uncertainty in the AP demands. That is, the nature of user association with APs and their data rate demands are stochastic. Therefore, the aggregated demand at each AP changes over time. This raises a question of how to efficiently allocate channels to each AP as needed to satisfy its demand while avoiding interference from other APs. Our objective is to minimize the total bandwidth used by the network. The smaller the bandwidth seized by our WLAN, the more space is facilitated for other networks to operate without interference. That is, our solution builds toward better coexistence between the controlled WLANs and other networks (e.g. other WLANs as well as networks based on other access technologies (e.g. LTE [29]).

A deterministic approach for channel allocation under AP demand uncertainty is to optimize network performance for either *average* or *peak* value. Considering the average demand does not provide guarantees on satisfying AP demands, as the ability to meet actual demands will depend on the probability that the actual demand will be less than or equal to its average value. On the other hand, considering the maximum demand value, although potentially meeting all AP demands, usually results in unnecessary consumption of scarce channel resources. In this work, we first propose a novel channel allocation framework, based on chance-constrained stochastic programming [30], in which the level of AP demand satisfaction is controllable. That is, we can guarantee a "tunable" threshold on the minimum proportion of AP demands to be met at the cost of the number of reserved channels for the WLAN. The higher this threshold is, the larger the number of allocated

channels will be. We refer to this chance-constrained channel allocation framework as CCCA. Although CCCA ensures a predetermined probabilistic Quality of Service (QoS) level, it provides a *static* channel allocation that does not adapt to AP demand variations. To better utilize the allocated resources, we then propose an adaptive channel allocation framework, based on two-stage stochastic optimization [30]. Instead of fixing the number of allocated channels for each AP, the framework adapts the number of allocated channels according to the AP's demand. This will save extra resources that would have been allocated under CCCA. Consequently, the total number of reserved channels for the whole network can be minimized. We refer to this adaptive channel allocation framework as ACA. Although ACA potentially achieves better spectrum utilization, it does not ensure a specific probabilistic QoS level (unlike CCCA).

As mentioned above, we use stochastic optimization to realize the proposed static and adaptive channel allocation frameworks. Stochastic optimization provides a powerful mathematical tool to handle optimization under uncertainty [30]. It has been recently exploited to optimize resource allocation in various types of wireless networks operating under uncertainties (e.g. [31–33, 47]).

We discuss the superiority of stochastic channel bonding over deterministic schemes, which consider the mean or maximum AP's demand values, through the following example.

Motivational example– Consider a network with two interfering APs as shown in Fig. 3.1(a). The demands of APs 1 and 2 are stochastic, as shown in Fig. 3.1(b). For illustration purposes, we assume that one channel is sufficient to satisfy one unit of AP demand. From here on, we assume that the deterministic approach considers the average value of each AP demand. As shown in Fig. 3.1(c), this approach results in allocating three channels for each AP all the time. Consequently, the AP demands will not be satisfied when they take the value of five. Besides, when an AP demand is one, the AP needs only one channel but this approach unnecessarily allocates three channels to it.

In contrast to the deterministic approach, the stochastic static approach provides the ability to balance the level of AP demand satisfaction and the total number of utilized channels through a predefined threshold β. Under this approach, the primary channel and bandwidth are determined

once such that the given AP demands are met with probability $\geq \beta$. Fig. 3.1(d) and Fig. 3.1(e) show the channel utilization and demand satisfaction when β is set to 0.8 and 1, respectively. Note that $\beta = 1$ setting is equivalent to deterministic naive approach that allocates for peak demand. It is easy to see that fewer channels will be utilized if β is set to a lower value but at the cost of satisfying the demand with a lower probability. Although this approach has an advantage over the deterministic approach, the channel allocation here is static and does not adapt to the demand variations.

Following the adaptive stochastic approach, the two AP demands can be fully satisfied under all scenarios while the total number of utilized channels is eight (as opposed to ten in the static case when $\beta = 1$), as shown in Fig. 3.1(f).

In IEEE Standard 802.11ac, an AP can extend its primary channel by seizing more contiguous channels to form one wide channel. If the allocated primary channel changes, the AP has to re-synchronize with its associated users. To eliminate this overhead, we keep the primary channel fixed in our schemes and allow the AP to extend/shrink its bandwidth around the same primary channel under different traffic demands.

Main contributions– Our contributions can be summarized as follows:

- Given the AP demand variability, we mathematically formulate our problem under the stochastic optimization framework using two different approaches. The first approach (CCCA) allows controlling the probability of demand satisfaction through a parameter β. The second approach (ACA) considers the adaptability in the allocated bandwidth according to the instantaneous demand. It also balances the level of AP demand satisfaction and the total number of channels allocated in the network.

- We develop an equivalent deterministic mixed-integer-linear-programming (MILP) formulation for each stochastic optimization problem in order to solve it using Branch-and Bound (B&B)-based methods [14].

- We propose a novel framework to solve the ACA problem efficiently. In this framework,

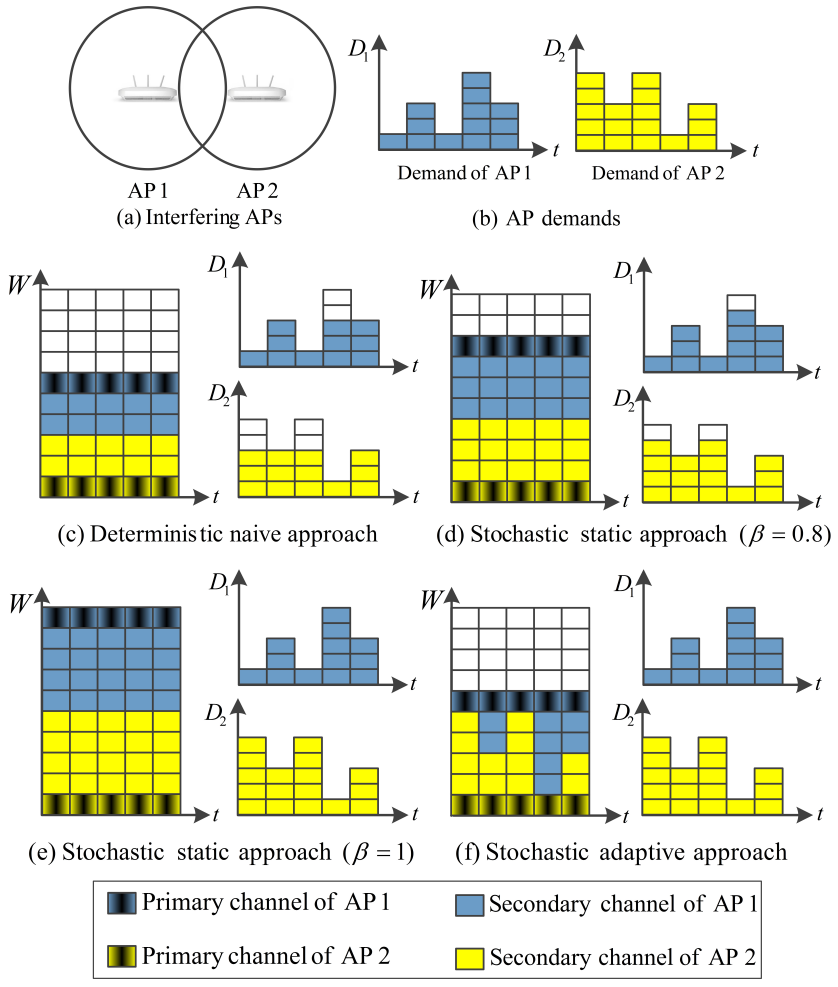


Figure 3.1: An illustration of static and adaptive channel bonding.

the special structure of the ACA is exploited through decomposition into two subproblems. A particle swarm optimization (PSO)-inspired algorithm is developed to solve the first subproblem. Besides, the framework facilitates the decomposition of the second subproblem into several subproblems that can be solved in parallel.

- We discuss the advantages of the proposed stochastic allocation approaches over the deterministic approach.
- We conduct extensive simulations to show the strength of the proposed stochastic optimization schemes and the PSO-inspired algorithm for different network sizes.
- We discuss insights on how to set different parameters under each stochastic allocation approach to best utilize the available spectrum.

Chapter organization–The rest of the chapter is organized as follows. We discuss the related work in the literature in Section 3.2. In Section 3.3, we introduce the system model and state our problem. In Section 3.4, we formulate our problem when the AP demands are deterministic. We mathematically formulate and tackle our problem under AP demand uncertainty through two different approaches in Sections 3.5 and 3.6. Given its complexity, we propose a novel PSO-inspired algorithm to solve the the ACA problem efficiently in Section 3.7. Section 3.8 demonstrates the superiority of the stochastic optimization framework and gives insights on the performance of the proposed solution approaches. In Section 3.9, we conclude our work and indicate directions for future research.

3.2 Related Work

Channel bonding (CB) was shown in the literature to be an effective technique for optimizing WLAN performance [34, 35]. However, these research efforts did not address the uncertainty in network parameters. In our work, we consider uncertainty in AP data rate demand as an important factor. If ignored, the overall performance of the network is significantly degraded.

In [28], the problem of band allocation with adaptive width in WLANs was addressed. They formulated the problem as an integer linear programming (ILP) and provided an NP-completeness proof. Then, they developed approximation algorithms where the interfering APs were assigned non-overlapping channels. In [36], Zarinni and Das proposed a dynamic spectrum distribution technique in which interfering APs are allowed to use overlapping channels if the links to their users are not interfering. In [37], an algorithm was proposed to dynamically select the channel center frequency and switch between 20 and 40 MHz channel widths in order to maximize the throughput. In [34], the channel bonding behavior in 802.11n networks was characterized and its impact on the network performance was studied. In [38], the problem of jointly allocating channel center frequencies and bandwidths for home WLANs was solved through a decentralized algorithm. The authors considered the trade-off between interference mitigation and the extra capacity offered by allocating more bandwidth to each AP. In [39], a game-theoretic approach was adopted to consider a scenario of interfering APs where each AP selects multi-bandwidth channels to maximize its own throughout. In [35], an analytical framework was proposed to study the benefits of CB in opportunistic spectrum access networks. The authors concluded that CB can enhance the network throughput in low occupancy of primary radios but it can reduce the throughput when the density of primary radio nodes is high. In [40], a performance comparison between CB and another channel aggregation technique (where the allocated frequency channels can be separate) in Carrier Sense Multiple Access (CSMA)-based networks was provided. The conclusion is that CB increases the throughput while the multi-channel approach minimizes the traffic congestion. In [41], the challenges raised by using carrier aggregations and CB in shared access and unlicensed bands were discussed. In [42], the capabilities of MIMO technology were exploited to intelligently use CB in order to overcome its main drawback, namely, decreased transmission range. In [43], the interaction between overlapping WLANs that use CB was studied and the effect on the network throughput was reported. The authors concluded that CB can have huge overall performance gains but also can result in unfairness in CSMA/Collision Avoidance (CSMA/CA)-based networks when some of the APs starve because they do not receive enough transmission opportunities. In our work, we manage AP channel usage through a central entity that ensures fairness and prevent

starvation. For a more comprehensive survey on channel bonding for wireless networks, the reader is referred to [44].

Recently, uncertainty in network parameters has been in the focus of several research efforts [45–49]. However, these efforts did not consider the uncertainty in user demands in CB-capable WLANs. For example, Abdel-Rahman and Krunz in [46] jointly optimized channel bonding and guard band allocation to maximize spectrum efficiency assuming deterministic and stochastic channel rates, respectively.

3.3 System Model and Problem Statement

Consider a WLAN where a set $\mathcal{P} = \{1, 2, \ldots, P\}$ of IEEE 802.11ac APs is deployed. We assume a central controller in the back-end linked to all APs with a high speed connection. This network architecture is suitable for enterprise and campus environments where all APs are managed through one entity [28,50]. The controller module can also be implemented inside one of the deployed APs. All control signals are collected at the central controller via the APs. Each AP has a stochastic data rate demand that collectively represents the demand of its associated users[1]. Fig. 3.2 shows two snapshots of the network for different user association at each AP. Once changed, an AP communicates its instantaneous user demand to the cental controller. The central controller makes the needed decisions regarding channel assignment and bandwidth allocation then conveys these decisions to the APs.

The transmission range of each AP is adjusted through controlling its power level so that a minimum signal strength is achieved at all associated users. The transmissions of an AP p might be overheard by a subset of other APs or their associated users. We call this subset "interfering APs" and is denoted as $\mathcal{T}_p$. Let $\mathcal{C} = \{1, 2, \ldots, C\}$ be the set of all available consecutive channels. Each AP is capable of extending its bandwidth by seizing a subset of these channels when needed. To avoid interference, interfering APs cannot operate simultaneously on the same frequency channel.

[1] How each AP schedules its users to meet their data rate requirements is not within the scope of this work.

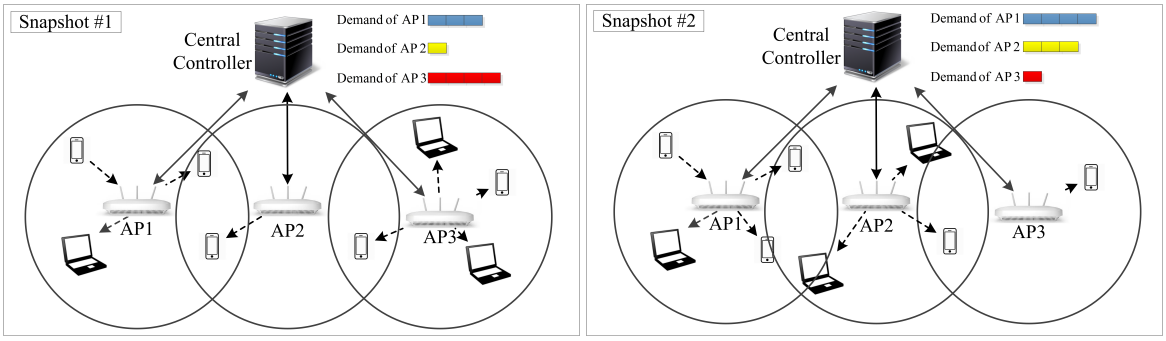


Figure 3.2: Network architecture and dynamics.

Consider a network of APs where different subsets of APs interfere with each other, and each AP has a stochastic demand that is described using a probability distribution function. Each AP is to be assigned a primary channel and allocated contiguous frequency channels to meet its data rate requirement while avoiding interference between interfering APs. The objective is to minimize the total bandwidth utilized by the network. Table 3.1 summarizes the notation used in this chapter.

3.4 Spectrum Distribution Under Deterministic AP Demands

In this section, we assume the user demand of each AP is fixed. The available spectrum is divided into b_{tot} small channels. Each AP will be allocated a contiguous band from these channels. We denote $b_{p\uparrow}$ and $b_{p\downarrow}$ as the indices of the top and bottom-frequency channels allocated to AP p, respectively. Each AP has a demand denoted by $D_p, p \in \mathcal{P}$. Let R_p be the achievable rate on a single channel between AP p and the associated users. We assume that R_p is fixed for each $p \in \mathcal{P}$.

Table 3.1: Notation.

General Sets, Parameters and Variables:

Symbol	Definition
$\mathcal{P}$	Set of APs in the network
$\mathcal{T}_p$	Set of APs interfering with AP p
D_p	Deterministic demand of AP $p \in \mathcal{P}$
$\widetilde{D}_p$	Stochastic demand of AP $p \in \mathcal{P}$
R_p	Achievable data rate for AP $p \in \mathcal{P}$ on a single channel
f_p	Index of the primary channel allocated to AP p
$b_{p\uparrow}$	Index of the top-frequency channel allocated to AP p
$b_{p\downarrow}$	Index of the bottom-frequency channel allocated to AP p
b_{tot}	Total number of available channels in the network
y_{pk}	Binary decision variable to indicate whether or not $b_{p\uparrow}$ is greater than $b_{k\downarrow}$
W	Total number of channels allocated to the network
β	Guaranteed probability of each AP demand is satisfaction under the static allocation scheme
α	Design coefficient to control the trade-off between the total number of utilized channels and the APs demand satisfaction under the adaptive allocation scheme

Algorithm-related Sets, Parameters and Variables:

Symbol	Definition
w_g	Global weighting factor in the PSO algorithm
w_l	Local weighting factor in the PSO algorithm
$\mathcal{I}$	Set of initial feasible solutions
$\mathcal{S}$	Set of all obtained feasible solutions
S_i	Current solution for particle i
$\bar{S}_i$	Best obtained solution for particle i
$\hat{S}$	Best obtained solution across all particles (the incumbent solution)
obj_i^{1st}	Current objective value of first stage problem for particle i
obj_i^{2nd}	Current objective value of second stage problem for particle i
obj_i	Current overall objective value for particle i
$f_{i,p}$	Current f_p of particle i
$v_{i,p}$	Current velocity of particle i in the p^{th} dimention
$\bar{obj}_i$	Best obtained overall objective value for particle i
$\hat{obj}$	Best obtained overall objective value across all particles (the incumbent objective value)

Evaluation-related Definitions:

Symbol	Definition
R/d	Ratio between the rate per channel and demand per user when each is the same for all APs
ξ	Average probability of demand satisfaction across all APs
χ	Average AP demand dissatisfaction

3.4.1 Optimization constraints

The achievable rate across the allocated band for any AP needs to meet its demand. This constraint can be formulated as follows:

$$R_p \left(b_{p\uparrow} - b_{p\downarrow} + 1\right) \geq D_p, \quad \forall p \in \mathcal{P}. \tag{3.1}$$

To ensure that two interfering APs will not be allocated the same channel, the following two sets of constraints are needed [28]:

$$b_{p\uparrow} - b_{k\downarrow} < b_{\text{tot}}\, y_{pk}, \qquad \forall p \in \mathcal{P}, k \in \mathcal{T}_p, \tag{3.2}$$

$$y_{pk} + y_{kp} \leq 1, \qquad \forall p \in \mathcal{P}, k \in \mathcal{T}_p, \tag{3.3}$$

where y_{pk} is a binary variable to indicate whether or not the upper-frequency index of AP p, given by $(b_{p\uparrow})$, is greater than the lower-frequency index of AP k, given by $(b_{k\downarrow})$. If both y_{pk} and y_{kp} are set to one, the bands allocated to APs p and k are overlapping. The second constraint prevents such overlap in the allocated bands if the transmission ranges of APs p and k are overlapping.

3.4.2 Optimization objective

As mentioned earlier, the goal of our problem is to minimize the total bandwidth utilized by the network while meeting all AP demands. Minimizing the total bandwidth can be achieved by minimizing the maximum upper-frequency index allocated to any AP. That is,

$$\min \max_{p \in \mathcal{P}} \left\{b_{p\uparrow}\right\}. \tag{3.4}$$

This objective function can be linearized by introducing an auxiliary decision variable W

which represents the total number of utilized channels, and adding the following set of constraints:

$$b_{p\uparrow} \leq W, \quad \forall p \in \mathcal{P}. \tag{3.5}$$

This set of constraints ensures that the upper-frequency index allocated to any AP cannot exceed the total number of channels allocated to the network. Let $\mathbf{b}_{\uparrow} = \{b_{p\uparrow} : p \in \mathcal{P}\}$. Then, the original objective function can be represented as follows:

$$\begin{aligned} &\underset{\{W,\mathbf{b}_{\uparrow}\}}{\text{minimize}} \quad W \\ &\text{subject to:} \\ &\qquad b_{p\uparrow} \leq W, \quad \forall p \in \mathcal{P}. \end{aligned}$$

3.4.3 Optimization formulation

Let $\mathbf{y} = \{y_{pk} : p \in \mathcal{P}, k \in \mathcal{T}_p\}$ and $\mathbf{b}_{\downarrow} = \{b_{p\downarrow} : p \in \mathcal{P}\}$. The frequency band allocation problem under deterministic AP demands can be formulated as an integer linear program (ILP) as follows:

Problem 1: Static Channel Allocation Under Deterministic AP Demands

$$\underset{\{W,\mathbf{b}_{\uparrow},\mathbf{b}_{\downarrow},\mathbf{y}\}}{\text{minimize}} \quad W$$

subject to:

Deterministic demand constraint: (3.1);

Interference avoidance constraints: $(3.2) - (3.3)$;

Top-frequency channels constraint: (3.5);

$$y_{pk} \in \{0, 1\}, \qquad \forall p \in \mathcal{P}, k \in \mathcal{T}_p; \tag{3.6}$$

$$b_{p\downarrow}, b_{p\uparrow} \in \{1, \ldots, b_{\text{tot}}\}, \qquad \forall p \in \mathcal{P}; \tag{3.7}$$

$$W \in \{1, \ldots, b_{\text{tot}}\}. \tag{3.8}$$

In the following two sections, we consider uncertainty in the AP demands. The demand of AP $p \in \mathcal{P}$ is represented as a stochastic variable $\widetilde{D}_p$. First, we propose a *static* stochastic frequency

band allocation scheme. Then, we develop a two-stage stochastic primary channel and *adaptive* bandwidth allocation scheme.

3.5 Spectrum Distribution Under Stochastic AP Demands: A Static Approach

In this section, we mathematically formulate the static single-stage stochastic frequency band allocation problem and develop its deterministic equivalent problem (DEP).

3.5.1 Problem formulation

We use chance-constrained stochastic programming [30] to formulate our problem. The objective is to minimize the amount of spectrum that is collectively utilized by all APs. At the same time, we want to guarantee that each AP demand is satisfied with probability $\geq \beta$, where $\beta \in [0, 1]$.

The probabilistic satisfaction of AP demands can be formulated using the following "chance constraint":

$$\Pr\left\{R_p\ (b_{p\uparrow} - b_{p\downarrow} + 1) \geq \widetilde{D}_p\right\} \geq \beta, \quad \forall p \in \mathcal{P}. \tag{3.9}$$

The overall problem can be formulated as follows:

CCCA: Static Channel Allocation Under Stochastic AP Demands

$$\underset{\{W,\mathbf{b}_\uparrow,\mathbf{b}_\downarrow,\mathbf{y}\}}{\text{minimize}} \; W$$

subject to:

Interference avoidance constraints: $(3.2) - (3.3)$;
Top-frequency channels constraint: (3.5);
Probabilistic demand constraint: (3.9);
Variable limits constraints: $(3.6), (3.7), (3.8)$.

3.5.2 Problem reformulation and solution procedure

To convert our chance-constrained program to a deterministic program, the following set provides an exact and equivalent reformulation for the original set of chance constraints:

$$R_p \left(b_{p\uparrow} - b_{p\downarrow} + 1\right) \geq F^{-1}_{\widetilde{D}_p}(\beta), \quad \forall p \in \mathcal{P}, \tag{3.10}$$

where $F^{-1}_{\widetilde{D}_p}(\beta)$ is the β-quantile function of $\widetilde{D}_p$ (equivalently, the inverse Cumulative Distribution Function (CDF) of $\widetilde{D}_p$ evaluated at β) and can be obtained numerically using the distribution of $\widetilde{D}_p$. Consequently, the DEP of CCCA problem can be stated as follows:

DEP of the CCCA Problem

$$\underset{\{W,\mathbf{b}_\uparrow,\mathbf{b}_\downarrow,\mathbf{y}\}}{\text{minimize}} \; W$$

subject to:

Interference avoidance constraints: $(3.2) - (3.3)$;
Top-frequency channels constraint: (3.5);
Equivalent demand constraint: (3.10);
Variable limits constraints: $(3.6), (3.7), (3.8)$.

DEP of CCCA problem is in the form of an MILP. B&B-based methods [14] can solve MILPs efficiently and provide the optimal solution for reasonable network sizes.

3.6 Spectrum Distribution Under Stochastic AP Demands: An Adaptive Approach

In this section, we mathematically formulate the two-stage stochastic primary channel and adaptive bandwidth allocation problem and develop an equivalent DEP.

3.6.1 Problem formulation

Using two-stage stochastic programming, we formulate our spectrum distribution problem. In contrast to the static approach, here the bandwidth assignment adapts to the variations in the AP demands. The goal of the first-stage problem is to optimally determine the primary channel for each AP, knowing the distribution of each AP demand. The first-stage problem decision is static and is taken before knowing the realization of each AP demand. In the second-stage problem, the bandwidth is optimized for each AP under each realization of AP demands. The target is to minimize the maximum deficit in AP demand satisfaction, i.e.

$$\max_{p\in\mathcal{P}}\left\{\left(\widetilde{D}_p - R_p\left(b_{p\uparrow} - b_{p\downarrow} + 1\right)\right)^+\right\}.$$

We denote f_p as the index of the primary channel assigned to AP p. For each AP, the primary channel is to fall within the allocated band. This can be modeled using the following two sets of simple constraints:

$$b_{p\uparrow} \geq f_p, \quad \forall p \in \mathcal{P}, \tag{3.11}$$

$$b_{p\downarrow} \leq f_p, \quad \forall p \in \mathcal{P}. \tag{3.12}$$

The first set of constraints ensures that the top-frequency channel does not fall below the primary channel in the allocated band for each AP. The second set of constraints ensures that the bottom-frequency channel does not fall beyond the primary channel of the allocated band for each AP. Let $\mathbf{f} = \{f_p : p \in \mathcal{P}\}$ and $\widetilde{\mathbf{D}} = \{\widetilde{D}_p : p \in \mathcal{P}\}$. Then, the two-stage stochastic optimization problem can be formulated as follows:

ACA: Adaptive Channel Allocation Under Stochastic AP Demands

$$\underset{\{W,\mathbf{f}\}}{\text{minimize}} \left\{ W + \alpha\, \mathbb{E}\left[\psi\left(\mathbf{y}, \mathbf{b}_{\uparrow}, \mathbf{b}_{\downarrow}, \widetilde{\mathbf{D}} \right) \right] \right\}$$

subject to:

$$f_p \in \{0, 1, \ldots, b_{\text{tot}} - 1\}, \qquad \forall p \in \mathcal{P}; \tag{3.13}$$

Variable limits constraints: (3.8),

where $\psi\left(\mathbf{y}, \mathbf{b}_{\uparrow}, \mathbf{b}_{\downarrow}, \widetilde{\mathbf{D}} \right)$ is the optimal value of the second-stage problem, which is given by:

$$\underset{\{\mathbf{b}_{\uparrow}, \mathbf{b}_{\downarrow}, \mathbf{y}\}}{\text{minimize}} \;\; \max_{p \in \mathcal{P}} \left\{ \left(\widetilde{D}_p - R_p \left(b_{p\uparrow} - b_{p\downarrow} + 1 \right) \right)^+ \right\}$$

subject to:

Interference avoidance constraints: $(3.2) - (3.3)$;
Top-frequency channels constraint: (3.5);
Primary channel constraints: $(3.11), (3.12)$;
Variable limits constraints: $(3.6), (3.7)$,

where α is a design coefficient to control the trade-off between the total number of utilized channels and the AP demand satisfaction.

3.6.2 Equivalent MILP reformulation

The objective function of the ACA second-stage problem is not linear. Denote S as the maximum deficit in AP demand satisfaction. To have an equivalent linear formulation, we add the

following constraints:

$$\widetilde{D}_p - R_p\,(b_{p\uparrow} - b_{p\downarrow} + 1) \leq S, \quad \forall p \in \mathcal{P}, \tag{3.14}$$

$$S \geq 0. \tag{3.15}$$

Note that a solution may provide excess data rate for the APs. We prevent rewarding extra data rate in the objective function by restricting S to be non-negative. Here, we need to represent each scenario which corresponds to a specific realization of AP demands in the new formulation. Denote Ω as the set of "scenarios", or all possible demand realizations, and $\omega \in \Omega$ is a specific realization. Let $p^{(\omega)}$ be the probability of scenario $\omega \in \Omega$, and $\mathbf{S} = \{S^{(\omega)} : \omega \in \Omega\}$. Also, we redefine $\mathbf{y}, \mathbf{b}_\uparrow, \mathbf{b}_\downarrow$ as follows. Let $\mathbf{y} = \{y_{pk}^{(\omega)} : p \in \mathcal{P}, k \in \mathcal{T}_p, \omega \in \Omega\}$, $\mathbf{b}_\uparrow = \{b_{p\uparrow}^{(\omega)} : p \in \mathcal{P}, \omega \in \Omega\}$ and $\mathbf{b}_\downarrow = \{b_{p\downarrow}^{(\omega)} : p \in \mathcal{P}, \omega \in \Omega\}$. Then, the second stage problem can reformulated as follows:

Reformulation of the Second Stage of the ACA Problem

$$\underset{\{\mathbf{b}_\uparrow, \mathbf{b}_\downarrow, \mathbf{y}, \mathbf{S}\}}{\text{minimize}} \sum_{\omega \in \Omega} p^{(\omega)}\ S^{(\omega)}$$

subject to:

$$D_p^{(\omega)} - R_p\,(b_{p\uparrow}^{(\omega)} - b_{p\downarrow}^{(\omega)} + 1) \leq S^{(\omega)}, \qquad \forall p \in \mathcal{P}, \omega \in \Omega; \quad (3.16)$$

$$b_{p\uparrow}^{(\omega)} - b_{k\downarrow}^{(\omega)} < b_{\text{tot}}\ y_{pk}^{(\omega)}, \qquad \forall p \in \mathcal{P}, k \in \mathcal{T}_p, \omega \in \Omega; \quad (3.17)$$

$$y_{pk}^{(\omega)} + y_{kp}^{(\omega)} \leq 1, \qquad \forall p \in \mathcal{P}, k \in \mathcal{T}_p, \omega \in \Omega; \quad (3.18)$$

$$b_{p\uparrow}^{(\omega)} \leq W, \qquad \forall p \in \mathcal{P}, \omega \in \Omega; \quad (3.19)$$

$$b_{p\uparrow}^{(\omega)} \geq f_p, \qquad \forall p \in \mathcal{P}, \omega \in \Omega; \quad (3.20)$$

$$b_{p\downarrow}^{(\omega)} \leq f_p, \qquad \forall p \in \mathcal{P}, \omega \in \Omega; \quad (3.21)$$

$$S^{(\omega)} \geq 0, \qquad \forall \omega \in \Omega; \quad (3.22)$$

$$y_{pk}^{(\omega)} \in \{0, 1\}, \qquad \forall p \in \mathcal{P}, k \in \mathcal{T}_p, \omega \in \Omega; \quad (3.23)$$

$$b_{p\downarrow}^{(\omega)}, b_{p\uparrow}^{(\omega)} \in \{1, \ldots, b_{\text{tot}}\}, \qquad \forall p \in \mathcal{P}, \omega \in \Omega. \quad (3.24)$$

Note that W and $f_p, \forall p \in \mathcal{P}$ are here constants and represent a solution to the first stage of the ACA problem.

In order to solve the ACA problem using a (B&B)-based method, the two stages of the problem must be combined into one linear programming formulation. This can be achieved using the DEP as follows:

DEP of ACA Problem

$$\underset{\{W,\mathbf{f},\mathbf{b}_{\uparrow},\mathbf{b}_{\downarrow},\mathbf{y},\mathbf{S}\}}{\text{minimize}} \left\{ W + \alpha \sum_{\omega \in \Omega} p^{(\omega)} \; S^{(\omega)} \right\}$$

subject to:

Demand deficit constraints: (3.16);
Interference avoidance constraints: $(3.17), (3.18)$;
Top-frequency channels constraint: (3.19);
Primary channel constraints: $(3.20), (3.21)$;
Variable limits constraints: $(3.8), (3.13), (3.22), (3.23), (3.24)$.

The DEP of the ACA problem is in the form of MILP which size grows exponentially with the number of APs. This MILP formulation can be solved to optimality using (B&B)-based methods for small network instances. However, the most recent derivatives of these methods fail to provide even a single feasible solution for reasonably large network sizes. In the sequel, we develop a novel framework to solve the ACA problem efficiently.

3.7 A Sub-optimal Algorithm for Adaptive Channel Bonding

In this section, we exploit the special structure of the ACA problem and propose a suboptimal algorithm to solve it efficiently.

3.7.1 Special structure of the ACA problem

The DEP of the ACA problem is in the form of an MILP and the number of scenarios (and hence, the number of variables) grows exponentially with the number of APs in the network. However, it has a special structure that can be exploited so that it can be solved efficiently. Constraints (3.19)-(3.21) are called coupling constraints between scenario-dependent variables ($b_{p\downarrow}^{(\omega)}$ and $b_{p\uparrow}^{(\omega)}$). The key point is that when the first-stage variables (i.e. W and $\mathbf{f}$) are fixed, these constraints become no longer coupling. Then, the second stage problem can be completely decomposed into finite number of subproblems. Each subproblem corresponds to one scenario and can be solved independently and in parallel to other subproblems. Let $\mathbf{y}^{(\omega)} = \{y_{pk}^{(\omega)} : p \in \mathcal{P}, k \in \mathcal{T}_p\}$, $\mathbf{b}_{\uparrow}^{(\omega)} = \{b_{p\uparrow}^{(\omega)} : p \in \mathcal{P}\}$ and $\mathbf{b}_{\downarrow}^{(\omega)} = \{b_{p\downarrow}^{(\omega)} : p \in \mathcal{P}\}$. For each scenario $\omega \in \Omega$, the corresponding subproblem can be written as follows:

Subproblem of Scenario $\omega \in \Omega$

$$\underset{\{\mathbf{b}_{\uparrow}^{(\omega)}, \mathbf{b}_{\downarrow}^{(\omega)}, \mathbf{y}^{(\omega)}, S^{(\omega)}\}}{\text{minimize}} \quad S^{(\omega)}$$

subject to:

$$D_p^{(\omega)} - R_p \left(b_{p\uparrow}^{(\omega)} - b_{p\downarrow}^{(\omega)} + 1\right) \leq S^{(\omega)}, \qquad \forall p \in \mathcal{P}; \quad (3.25)$$

$$b_{p\uparrow}^{(\omega)} - b_{k\downarrow}^{(\omega)} < b_{\text{tot}}\, y_{pk}^{(\omega)}, \qquad \forall p \in \mathcal{P}, k \in \mathcal{T}_p; \quad (3.26)$$

$$y_{pk}^{(\omega)} + y_{kp}^{(\omega)} \leq 1, \qquad \forall p \in \mathcal{P}, k \in \mathcal{T}_p; \quad (3.27)$$

$$b_{p\uparrow}^{(\omega)} \leq W, \qquad \forall p \in \mathcal{P}; \quad (3.28)$$

$$b_{p\uparrow}^{(\omega)} \geq f_p, \qquad \forall p \in \mathcal{P}; \quad (3.29)$$

$$b_{p\downarrow}^{(\omega)} \leq f_p, \qquad \forall p \in \mathcal{P}; \quad (3.30)$$

$$S^{(\omega)} \geq 0; \quad (3.31)$$

$$y_{pk}^{(\omega)} \in \{0, 1\}, \qquad \forall p \in \mathcal{P}, k \in \mathcal{T}_p; \quad (3.32)$$

$$b_{p\downarrow}^{(\omega)}, b_{p\uparrow}^{(\omega)} \in \{1, \ldots, b_{\text{tot}}\}, \qquad \forall p \in \mathcal{P}. \quad (3.33)$$

Fig. 3.3 demonstrates the idea of decomposing the DEP of the ACA problem.

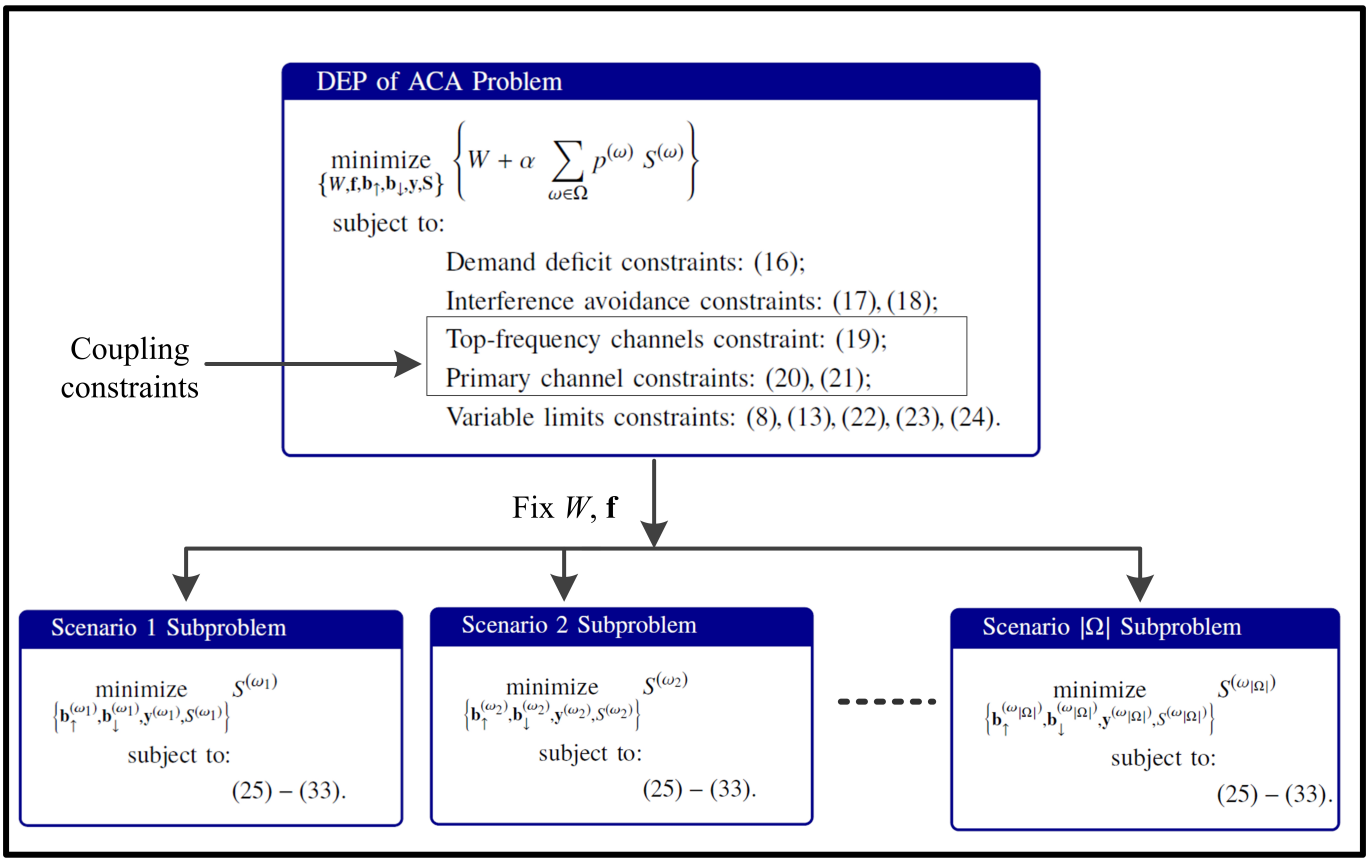


Figure 3.3: Decomposability of the ACA problem's DEP.

After solving all subproblems, the overall value of the second stage problem objective function is a simple weighted sum of the subproblem objective function optimum values (i.e. $\sum_{\omega\in\Omega} p^{(\omega)}\ S^{(\omega)}$). The structure of the ACA problem is known in the literature as the *L-shaped* structure. As the first stage problem constrains non-continuous variables, the overall problem needs to be solved within a B&B framework.

In the basic form of the B&B method, a search tree is constructed by fixing one or more binary decision variables at the value of zero or one. Many derivatives of this method have been proposed in the literature. The main idea in all these derivatives is that during the search process, special constraints are generated in order to limit the B&B search tree so that the optimal solution can be found quickly. As an example, the integer L-Shaped method was first introduced in [51] and was further improved in [52] to solve two-stage stochastic optimization problems with complete recourse. However, the generated special constraints are assumed to be tight enough in order to eliminate a sufficient portion of the B&B search tree and reach the optimal solution in short time. Our preliminary experiments revealed that these constraints are actually loose. Therefore, the B&B search tree explodes, consuming a large amount of memory, and the algorithm runs for a long time (hours) without reaching even a single feasible solution. This also means that the optimality gap cannot be tracked during the B&B searching process. As a result, this approach is inefficient to solve the ACA problem.

In the sequel, we develop an efficient algorithm to overcome this challenge. The idea is to replace the B&B search procedure with a particle swarm optimization (PSO)-based algorithm to find solutions for the first stage problem while still solving the second stage problem optimally.

3.7.2 A PSO-inspired method

As explained in the previous section, the integer L-shaped method theoretically can eventually obtain the optimal solution for the overall problem but it takes a long time and consumes a large amount of memory. Here, we introduce an efficient algorithm based on PSO which can solve the

problem in a much shorter time. PSO is a global search algorithm inspired by the social behavior of schooling fish, herding animals, and flocking birds in which the groups search cooperatively for food. Compared to other search algorithms (such as genetic algorithms and ant colony optimization), PSO is easier to implement and has fewer parameters to control. Over the last decade, PSO has been identified as an effective way to find good solutions for NP-hard problems [53–56]. Here, we use a PSO-based algorithm only to obtain solutions for the first stage of the ACA problem while solving the second stage problem to optimality. The goal of limiting the use of PSO to find solutions to the first stage variables (as opposed to the variables of the overall problem) is twofold. First, possibly sacrificing the first stage problem's optimality when using PSO is compensated by solving the second stage problem to optimality. Second, the number of first stage variables is small compared to those of the overall problem. This leads to limiting the PSO-based algorithm's search space which leads to obtaining good solutions within a few iterations.

The basic idea of the algorithm is to find candidate solutions for AP primary channels using PSO. When the primary channels are fixed in each solution, it is easy to compute the bandwidth for each AP by solving the second stage problem. A set of candidate solutions for the first-stage problem $\mathcal{I}$ is maintained, and each candidate solution is called a particle. Recall that P is the number of APs in the network. The coordinates of each particle in the P-dimensional search space represents the indices of the assigned primary channels to the network APs. These coordinates are modified in each iteration of the PSO procedure in order to find better values for the objective function of the overall two-stage stochastic optimization problem. Initially, each particle $i \in \mathcal{I}$ is located randomly (i.e. each AP is assigned a random f_p) and moves in steps (integer number of channels) in each direction throughout the iterations of the algorithm. The movement of each particle is influenced by its current location, the best location it has ever been positioned at, and the best location that has been found by all particles. At termination, the algorithm reports the best found location among all particles throughout all iterations. Now, we describe the details of the proposed algorithm. The overall algorithm is described in Fig. 3.4. It shows the two main functional blocks. The first functional block obtains an initial feasible solution to the overall problem. The second functional block exploits the solutions obtained so far in order to generate better first

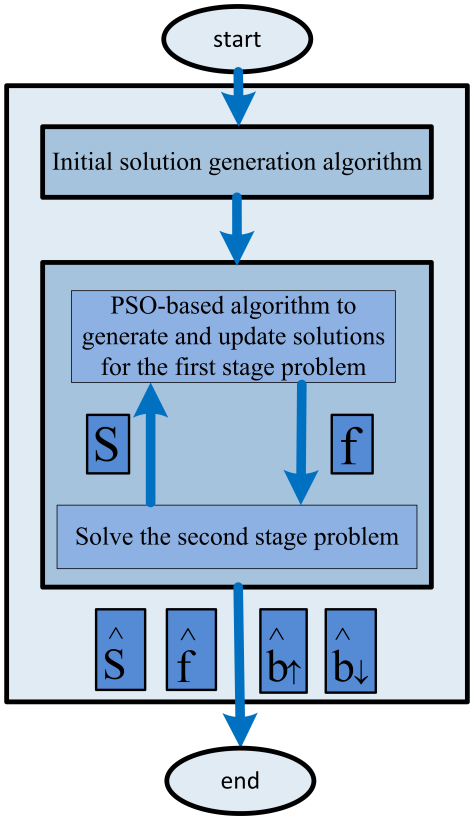


Figure 3.4: Proposed framework for solving the ACA problem.

stage solutions through iteration.

3.7.2.1 Initial solution generation

Algorithm 1 shows the pseudocode for generating an initial feasible solution. An AP $p \in \mathcal{P}$ is chosen randomly and assigned the first primary channel (i.e. $f_p = f = 0$) (lines 3-8). Denote $\mathcal{G}_f$ as the set of APs which has been assigned the primary channel f. Initially, $\mathcal{G}_f$ contains AP p (line 9). The next step is to assign the same primary channel f to as many APs as possible while avoiding interference. This is achieved by assigning f to each non-overlapping AP with p (say k) if none of the later's interfering APs ($\mathcal{T}_k$) is in $\mathcal{G}_f$ (lines 10-16). A new primary channel (to be assigned to the next group of APs) is chosen at a random number, say between two chosen predefined numbers a and b, of channels apart from the first chosen one (line 17). Another AP is randomly chosen

among the remaining APs (which have not yet been assigned a primary channel) and assigned the new primary channel. The procedure continues until each AP is assigned a primary channel. The algorithm then calculates the objective function value of the first stage problem for each of the obtained initial solution $i \in \mathcal{I}$ (denoted as obj_i^{1st}) (line 19). The second-stage subproblem is then solved efficiently using a state-of-the-art solver (CPLEX [13], which implements B&B methods) and the objective function value (denoted as obj_i^{2nd}) is obtained given the associated initial solution (line 20). The overall objective function value (denoted as obj_i) is a simple weighted sum of the first and second objective function values (line 21). After obtaining all $|\mathcal{I}|$ feasible solution, the "initial" best (incumbent) objective value $\hat{obj}$ is the minimum among all values obtained from the set of obtained solutions.

Algorithm 1 Procedure of initial solution generation

1: Set $\mathcal{H} = \mathcal{P}$; $f = 0$;
2: **while** $\mathcal{H} \neq \emptyset$ **do**
3: Pick an AP p randomly;
4: **while** $p \notin \mathcal{H}$ **do**
5: Pick another AP p randomly;
6: **end while**
7: $f_p = f$;
8: $\mathcal{H} := \mathcal{H} \setminus p$;
9: $\mathcal{G}_f = p$
10: **for** $k \in \mathcal{T}_p$ **do**
11: **if** $(k \in \mathcal{H})$ && $(\mathcal{T}_k \cap \mathcal{G}_f = \emptyset)$ **then**
12: $f_k = f$;
13: $\mathcal{H} := \mathcal{H} \setminus k$;
14: $\mathcal{G}_f := \mathcal{G}_f \cup k$
15: **end if**
16: **end for**
17: $f := f + rand(a, b)$;
18: **end while**
19: Determine obj_i^{1st};
20: Obtain obj_i^{2nd} using CPLEX;
21: $obj_i = obj_i^{1st} + \alpha \cdot obj_i^{2nd}$

3.7.2.2 Core algorithm

Now, we describe the core part of the algorithm. Algorithm 2 shows the pseudocode for the core proposed PSO-inspired algorithm. Denote $\mathcal{S}$ as the set of all obtained feasible solutions. Initially, $\mathcal{S}$ is identical to $\mathcal{I}$ (line 1). The algorithm starts each iteration by determining the coordinates of particle (the solution vector) $i \in \mathcal{S}$ through the following two equations (lines 5-8):

$$v_{i,p} := v_{i,p} + w_g \cdot (\hat{S}(p) - f_{i,p}) + w_l \cdot (\bar{S}_i(p) - f_{i,p}), \tag{3.34}$$

$$f_{i,p} := \max(0, \min(f_{i,p} + \text{round}(v_{i,p}), b_{tot} - 1)), \tag{3.35}$$

where $v_{i,p}$ is the velocity of particle i in the p^{th} dimension, w_g and w_l are the global and local weighting factors, respectively, $f_{i,p}, \bar{S}_i(p)$ and $\hat{S}(p)$ are the p^{th} entries in current solution vector i, best found solution for particle i, and the best solution among all particles, respectively. The global weighting factor determines how much the difference between a particle's current solution and the best (incumbent) solution ($\hat{S}$) affects its next movement of the particle. Similarly, the local weighting factor determines how much the difference between a particle's current solution and its best obtained solution ($\bar{S}_i$) affects its next movement. Note that in the last expression, the updated value of $v_{i,p}$ is rounded because the particle move in any direction is restricted to be an integer number of steps since it represents a channel index. Also, the updating process avoids moving the particle outside the search space in any of the P directions by adding a simple correction in the location update equation. Each updated solution $i \in \mathcal{I}$ is then checked for feasibility (line 9). If the solution is feasible, it is checked against the entries in the list of previously obtained solutions (i.e. $\mathcal{S}$) (line 10). If it is not found, obj_i^{1st} is calculated, the second-stage problem is solved using CPLEX to determine obj_i^{2nd}, and the overall objective function value is:

$$obj_i = obj_i^{1st} + \alpha \cdot obj_i^{2nd}, \tag{3.36}$$

and a new entry for this solution is created in $\mathcal{S}$ (lines 11-14). If the solution has already been obtained before, the values of its first-stage objective (obj_i^{1st}), second-stage objective (obj_i^{2nd})and overall objective (obj_i) are updated with the values in corresponding entry of $\mathcal{S}$ (line 16). If the solution is not feasible, obj_i is set to high value (line 19). At the end of the iteration, the best achieved position (solution) for each particle ($\bar{obj}_i$) and the overall best (incumbent) solution ($\hat{obj}$) are updated (lines 21-26). The algorithm terminates when either the maximum number of iterations is reached.

Algorithm 2 PSO-inspired algorithm to solve the ACA problem

1: Find a set of $\mathcal{I}$ initial feasible solutions using Algorithm 1;
2: Set $k = 0$; $\mathcal{S} = \mathcal{I}$;
3: **while** $k \leq$ #Iterations **do**
4: **for** $i \in \mathcal{I}$ **do**
5: **for** $p \in \mathcal{P}$ **do**
6: $v_{i,p} := v_{i,p} + w_g \cdot (\hat{S}(p) - f_{i,p}) + w_l \cdot (\bar{S}_i(p) - f_{i,p})$;
7: $f_{i,p} := \max(0, \min(f_{i,p} + \text{round}(v_{i,p}), b_{tot} - 1))$;
8: **end for**
9: **if** S_i is feasible **then**
10: **if** $S_i \notin \mathcal{S}$ **then**
11: Determine obj_i^{1st};
12: Use CPLEX to obtain obj_i^{2nd};
13: $obj_i := obj_i^{1st} + \alpha \cdot obj_i^{2nd}$;
14: $\mathcal{S} := \mathcal{S} \cup S_i$;
15: **else**
16: Get $obj_i, obj_i^{1st}, obj_i^{2nd}$ from the corresponding entry in $\mathcal{S}$;
17: **end if**
18: **else**
19: Set $obj_i := obj_i^{1st} := obj_i^{2nd} := \infty$;
20: **end if**
21: **if** $\bar{obj}_i > obj_i$ **then**
22: $\bar{obj}_i := obj_i$;
23: **end if**
24: **if** $\hat{obj} > \bar{obj}_i$ **then**
25: $\hat{obj} := \bar{obj}_i$;
26: **end if**
27: **end for**
28: $k := k + 1$;
29: **end while**

3.8 Performance Evaluation

In this section, we evaluate our stochastic allocation schemes, and compare them to the existing approaches. Moreover, we study the performance of the PSO-inspired algorithm for solving the two-stage stochastic problem.

3.8.1 Distribution of the AP demands

To generate sensible instances of AP demands, we use the model proposed by Chen, Kurose, and Towsley in [50]. The authors developed a mixed queueing network model to capture user mobility in a campus network. Two classes of users were considered: open and closed. The open class consists of users who visit the network for a short period of time then leave. The marginal occupancy distribution at AP p of the open-class users has been expressed as a Poisson distribution with parameter $\rho_{p_o} = \frac{\lambda_p}{\mu_p}$, where ρ_{p_o} is the load of AP p generated from the open-class users, λ_p is the aggregate arrival rate, and $1/\mu_p$ is the expected time the user stays in the network. The closed class consists of users who stay in the campus for long period and switch between multiple APs. The marginal occupancy distribution at AP p of the closed-class users has been approximated by a Poisson distribution with parameter $\rho_{p_c} = N\ \nu_p$, where N is the total number of users in the closed class, and ν_p is the fraction of time during which a closed-class user is associated with AP p. As the distributions of the two classes are independent, the overall marginal occupancy distribution of AP p is the convolution of two Poisson distributions, which is also Poisson with parameter ρ_p, where $\rho_p = \rho_{p_o} + \rho_{p_c}$. We assume that the demand per user (denote by d) is the same for all APs and users. Then, AP demand has a Poisson distribution with parameter ρ_p, and its instantaneous value can be calculated through multiplying the number of currently associated users by d.

Table 3.2: System parameter values.

Parameter	Value
$\lambda_p, \forall p \in \mathcal{P}$	1-5 users/min.
$1/\mu_p, \forall p \in \mathcal{P}$	1-5 min.
N	50
$\nu_p, \forall p \in \mathcal{P}$	0.001-0.005

Table 3.3: Algorithm parameter values.

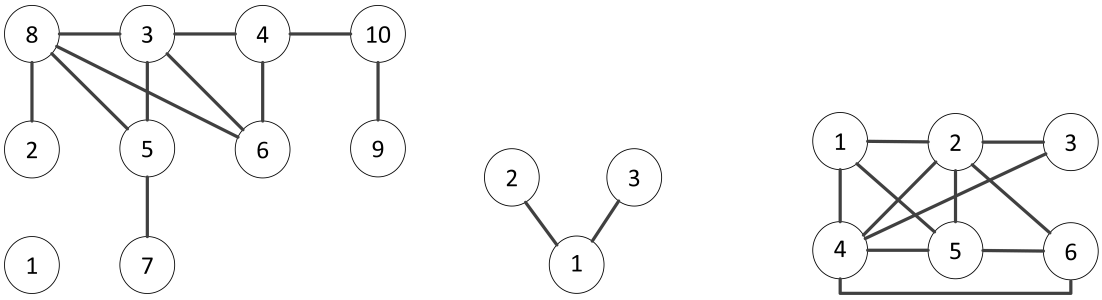
Parameter	Value
(a, b)	(1,4)
w_g	0.1
w_l	0.1
# iterations	10

3.8.2 Evaluation setup

The parameters of AP demand distribution were chosen as follows. The arrival rate of users at each AP is uniformly and independently generated between one and five users per minute. The expected stay time of users at each AP is randomly selected between one and five minutes. For the closed-class users, we set $N = 50$ and ν_p is randomly selected between 0.001 and 0.005. Table 3.2 summarizes the parameter settings in our simulations. The data rate per channel is assumed to be the same for all APs, i.e. $R_p \stackrel{\text{def}}{=} R, \forall p \in \mathcal{P}$.

The ratio R/d is used as a parameter in our simulations. The value of this ratio reveals how frequency channels can satisfy user demands. When R/d is small (i.e. < 1), multiple channels are required to satisfy a single user demand. On the contrary, large values of R/d indicate that a single channel might suffice to satisfy multiple user demands.

In our experiments, we use three network setups. The first network setup is used to demonstrate the benefits of the static stochastic allocation approach. The network consists of ten APs, each of which can have up to 15 users. The interference relationships between APs are shown in Fig. 3.5a, where a line between two APs exists if they are interfering. The second network setup is to show the advantage of the two-stage adaptive stochastic allocation approach and the performance of



(a) Network setup 1 (b) Network setup 2 (c) Network setup 3

Figure 3.5: Interference graphs.

the PSO-inspired algorithm. The network consists of three APs, each of which can have up to seven users. The interference graph of the second network is shown in Fig. 3.5b. We consider a general case where AP 1 interferes with both AP 2 and 3 while AP 2 and 3 are not interfering with each other. The third network setup is to show the performance of the PSO-inspired algorithm for larger network sizes. The network consists of six APs, each of which can have up to seven users. The interference graph of the third network is shown in Fig. 3.5c. We ran our experiments on a general-purpose desktop computer, which is Dell Precision T7600 with 16 processor cores (Intel Xeon CPU E5-2687W 0 @ 3.1 GHz) and 64 GB RAM. CPLEX was used to solve our optimization problems. For the deterministic and the static (single-stage) stochastic schemes, whenever CPLEX reports multiple optimal solutions (with the same value of W), we pick the solution that gives the maximum average probability of demand satisfaction across all APs.

We use the following metrics to evaluate our proposed stochastic allocation approaches: (i) the total number of utilized channels (W), (ii) the average probability of demand satisfaction across all APs, denoted as ξ, and (iii) the average AP demand dissatisfaction, denoted as χ, which considers only the cases where the AP demands are not satisfied. The relative deficit in each AP demand satisfaction is obtained then we calculate the average over all AP deficits. For the static (single-stage) and deterministic schemes, the relative deficit for each AP $p \in \mathcal{P}$ can be computed as

follows:

$$\chi_{\text{static-det}} = \frac{\sum\limits_{d_p \in \widetilde{D}_p} \left\{ \Pr\{\widetilde{D}_p = d_p\} \ \frac{\max(d_p - R\ b_p, 0)}{d_p} \right\}}{\sum\limits_{d_p \in \widetilde{D}_p : d_p > R\ b_p} \Pr\{\widetilde{D}_p = d_p\}} \times 100\%.$$

For the two-stage problem, the relative deficit for each AP $p \in \mathcal{P}$ is calculated as follows:

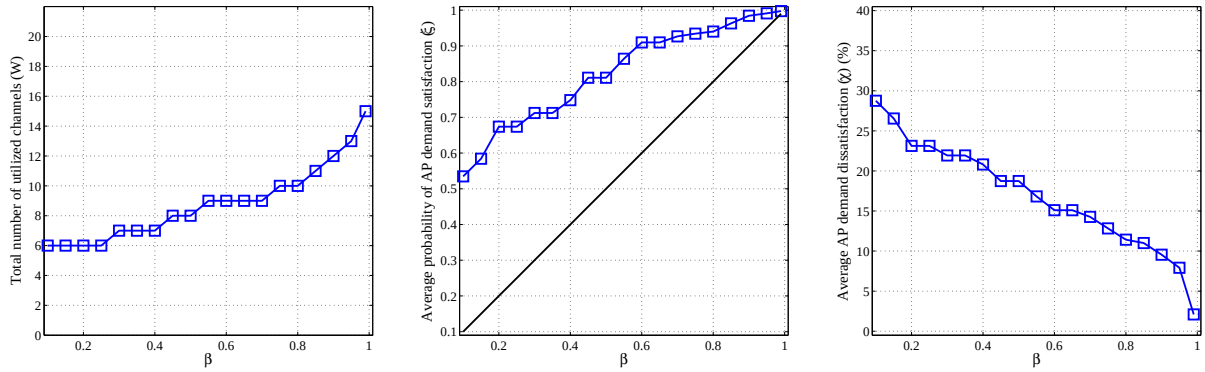
$$\chi_{\text{adaptive}} = \frac{\sum\limits_{\omega \in \Omega} \left\{ p^{(\omega)} \ \frac{\max\left(D_p^{(\omega)} - R\ b_p^{(\omega)}, 0\right)}{D_p^{(\omega)}} \right\}}{\sum\limits_{\omega \in \Omega : D_p^{(\omega)} > R\ b_p^{(\omega)}} p^{(\omega)}} \times 100\%.$$

3.8.3 Static approach

As shown in Fig. 3.6, the static approach provides flexibility in controlling the trade-off between W and both ξ and χ. As mentioned before, this controllability in the system performance is not possible under the deterministic approach. As shown in Fig. 3.6(b), when β is relatively small, the achieved ξ is significantly higher than the minimum required. This can be explained as follows. Some APs overlap with only one AP (e.g. AP 2 and 9) or does not overlap with any AP (e.g. AP 1). In such cases, when the total bandwidth utilized by the network is higher than these "isolated" APs' required bandwidth (for a given β), they can seize more channels as long as the total number of seized channels do not exceed the number utilized by the overall network. This results in achieving a probability of satisfaction much higher than β for these individual APs. Consequently, ξ exceeds β by a noticeable amount. This surplus decreases as β increases because the requirements become higher and the corresponding constraints are thus tighter.

3.8.4 Static vs. deterministic

In this subsection, we illustrate the gains of static approach. More specifically, we compare the performance of the deterministic scheme (which considers the mean value of the demand) with the static approach under two values of β. In our experiments, the probability that the actual demand



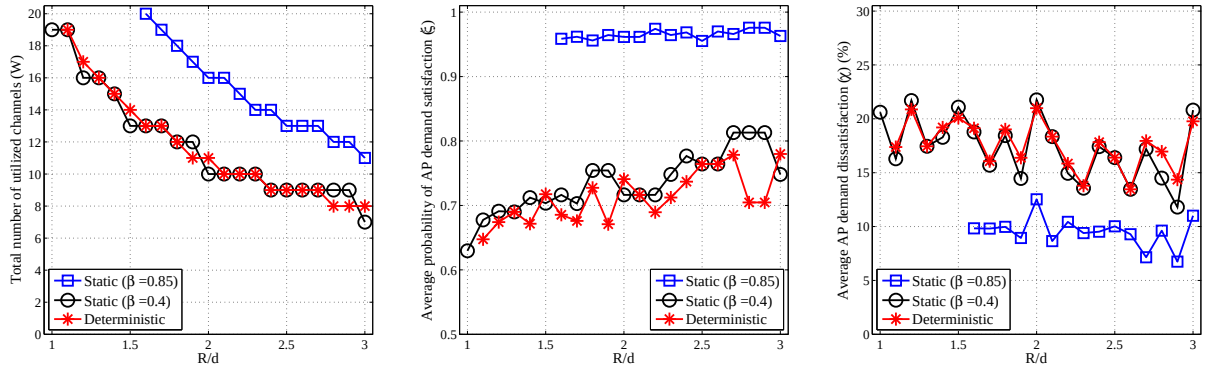
(a) (b) (c)

Figure 3.6: Behaviour of static approach (10 APs, 15 users, $R/d = 3.0$).

is smaller than or equal to the expected demand (i.e. $\Pr\{\widetilde{D}_p \leq \mathbb{E}[\widetilde{D}_p]\}$) is between 0.41 and 0.55. As shown in Fig. 3.7, when $\beta = 0.4$, the static approach performs similar to the deterministic one. When $\beta = 0.85$, both ξ and χ have much better values at the cost of more allocated channels. Also, when $R/d \leq 1.5$, there is no feasible solution as β is too large to be satisfied given the available number of channels. The fluctuations of ξ and χ in Fig. 3.7(b) and (c), respectively, can be explained through an example as follows. Let us consider the curves corresponding to $\beta = 0.4$. When the value of R/d increases from 1.9 to 2.0, W decreases from 12 to 10. Here, the increment in R/d could not compensate for the decrement in W. As a result, lower demand values could be covered by the current configuration. Consequently, ξ decreases and χ increases. On the other hand, when the value of R/d increases from 2.0 to 2.3, the value of W remains at 10, ξ increases, and χ decreases because the increment in R/d could satisfy a larger percentage of demands.

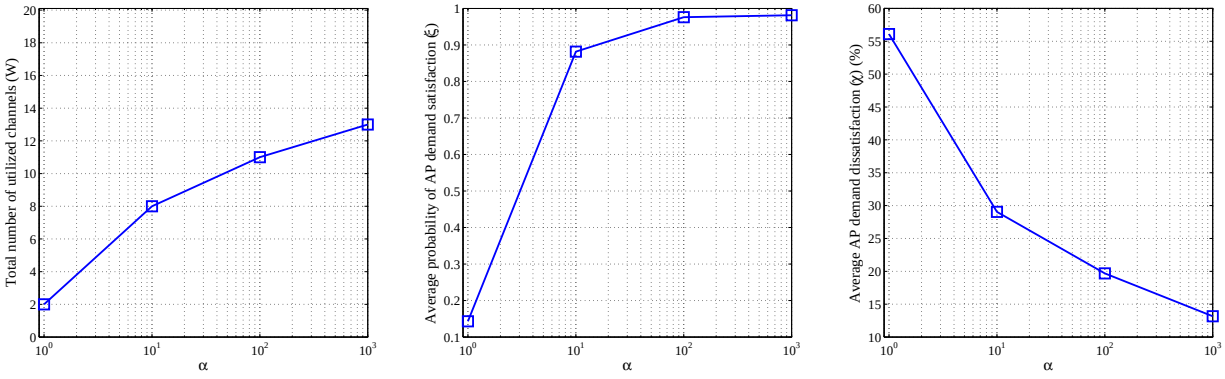
3.8.5 Adaptive approach

As discussed in Section 3.6, α controls the tradeoff between W and the APs demand satisfaction. Fig. 3.8 demonstrates the capability of the adaptive approach to control the trade-off between W and both ξ and χ. When α is sufficiently small (i.e. ≤ 1), the network seizes as few as one channel at the cost of small ξ and large χ. On the other hand, when α is sufficiently large (i.e. $\geq 10^3$), ξ approaches one, and χ drops to zero at the cost of seizing more channels.



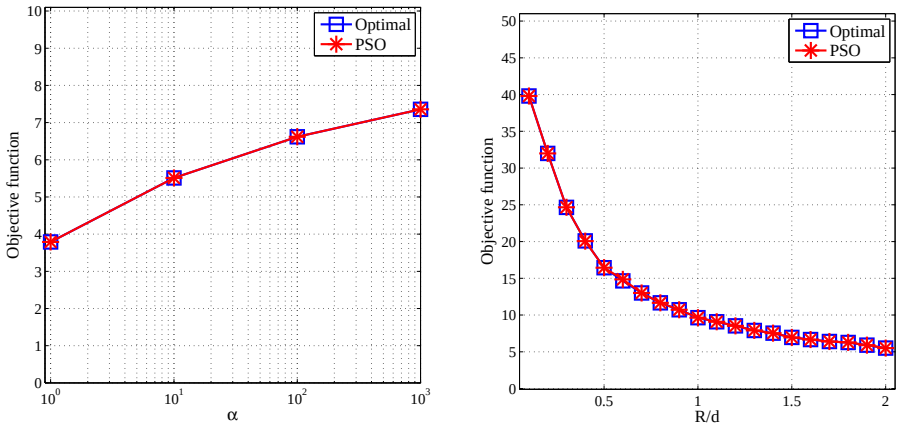
(a) (b) (c)

Figure 3.7: Comparison between deterministic and static approaches (10 APs, 15 users).



(a) (b) (c)

Figure 3.8: Behaviour of adaptive approach (3 APs, 7 users, $R/d = 1.0$).



(a) $R/d = 2.0$ (b) $\alpha = 10$

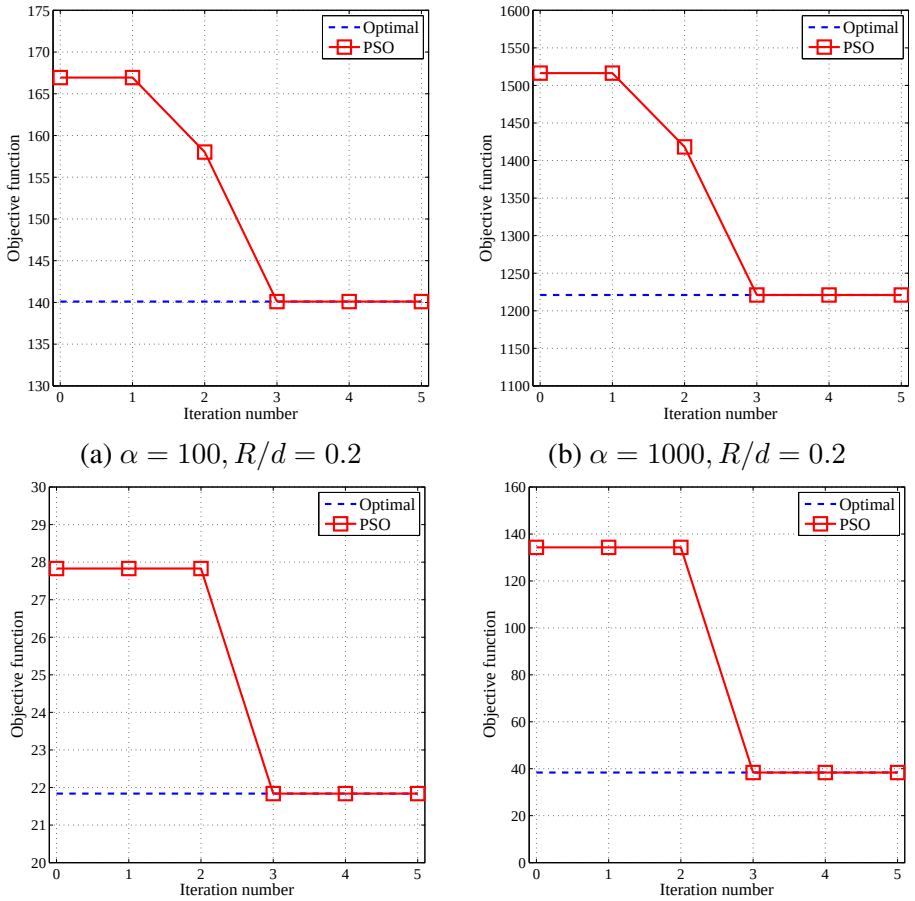
Figure 3.9: Performance of the proposed PSO-inspired framework (3 APs, 7 users).

3.8.6 PSO-inspired framework performance

In this section, we discuss the performance of the proposed PSO-inspired framework. Note that the shown results in this section are only a few examples from our experiments which spanned different network instances and gave similar results.

3.8.6.1 Performance for small networks

As discussed in Section 3.6, the problem size grows exponentially with the number of APs in the network. To tackle this challenge, a PSO-inspired framework was proposed in Section 3.7. Here, we compare this framework to the approach which solves the DEP using CPLEX. In the procedure of initial solution generation, we set the limits for the random number of channel separation as $a = 1$ and $b = 4$. The PSO algorithm parameters (w_g and w_l) are both set to 0.1. The maximum number of iterations is set to 10. Table 3.3 summarizes the PSO-inspired algorithm parameters. As shown in Fig. 3.9, the proposed framework was able to reach the same optimal solution obtained by CPLEX for different values of α and R/d within few iterations as will be discussed in the next subsection.



(a) $\alpha = 100, R/d = 0.2$ (b) $\alpha = 1000, R/d = 0.2$

(c) $\alpha = 100, R/d = 0.5$ (d) $\alpha = 1000, R/d = 0.5$

Figure 3.10: Convergence of the proposed PSO-inspired framework (3 APs, 7 users).

3.8.6.2 Convergence rate

Here, we show the convergence rate of the proposed framework. Each subfigure in Fig. 3.10 shows the behavior of the PSO-inspired algorithm for a specific combination of α and R/d. It is clear that the proposed framework takes three iterations in order to reach the optimal solution. When $R/d = 0.2$, the algorithm finds an intermediate solution before reaching the optimal one. When $R/d = 0.5$, it could not find intermediate solutions during the first and second iterations but could reach the optimal solution in the third iteration.

3.8.6.3 Performance for larger networks

As discussed in Section 3.6, CPLEX fails to obtain any solution for the DEP of the adaptive stochastic optimization problem. Our experiments reveal that three APs is the maximum network size for which CPLEX (being one of the most famous state-of-the-art solvers) can obtain any solution. We show here the ability of the proposed framework to obtain solutions for larger network size. We consider here the third network setup which consists of six APs, and up to seven users can be associated with each AP. Fig. 3.11 shows the values of the objective function, W, ξ, and χ, when $\alpha = 1$ and 1000. As expected, when R/d increases, the objective function value decreases. For $\alpha = 1000$, W decreases with R/d because all demands under all scenarios can be satisfied using smaller W. When $\alpha = 1$, ξ increases with R/d as larger portion of demand scenarios is satisfied.

3.9 Conclusions

We proposed a novel stochastic spectrum distribution framework for WLANs, which accounts for the APs' demand uncertainty. Static and adaptive stochastic bandwidth allocation approaches were developed. The objective of the static approach was to minimize the total number of channels allocated to the network. The static approach also considered guaranteeing a configurable minimum probability of satisfaction for each AP demand. On the other hand, the adaptive approach provided a trade-off between the number of channels and the level of AP demand satisfaction. Due to the complexity of the adaptive approach problem, we proposed a framework in which a PSO-inspired algorithm was developed. Through extensive simulations, we showed the superiority of the proposed stochastic allocation framework compared to the deterministic approach. Also, the results demonstrated the ability of the proposed PSO-inspired algorithm to solve the problem efficiently.

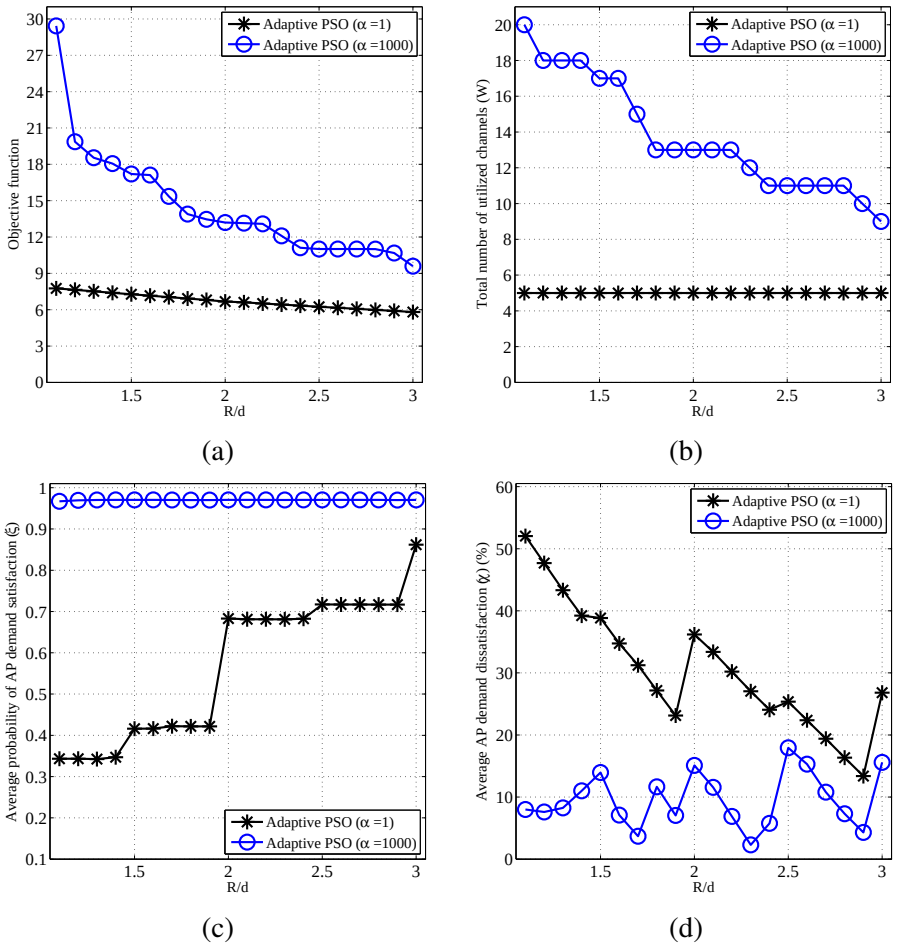


Figure 3.11: Behaviour of PSO-inspired framework for larger networks (6 APs, 7 users).

Chapter 4

On the Optimal Resource Allocation in Multi-RAT Wireless Networks with Receiver Characteristic Awareness [1]

4.1 Introduction

The number of connected devices and the demand for quality wireless services have seen an explosive growth in the recent past, and is expected to increase exponentially in the future [58]. This places enormous demands on the radio frequency (RF) spectrum, which is limited and hence expensive. In order to cope with this increasing demand, several researchers have proposed improving the spectrum utilization by replacing the static and exclusive spectrum allocation model with a dynamic allocation model [59–63]. This paves the way for network sharing and coexistence (see, for example, [64, 65]).

[1] The Introduction, Motivational Example, and Preliminaries sections of this chapter are based on the similarly titled sections in the conference and journal papers on which this chapter is based, and these sections of those papers were originally written by Aditya V. Padaki. They have been appropriately modified for appearance here, but I am grateful to Aditya for his contributions to this work.

To ensure harmonious coexistence, several frameworks based on sensing, beacons, and databases have been proposed [64]. Such frameworks will allow many radio access technologies (RATs) to operate in the same band while avoiding harmful interference to each other. However, realizing such frameworks imposes an onerous task of managing wireless networks with diverse technologies and optimizing several parameters to maximize spectral efficiency. The Spectrum Access System (SAS) is one database-driven framework in which a centralized entity manages spectrum in real time. SAS has been recently adopted by the Federal Communications Commission (FCC) to enable spectrum sharing in the 3.5 GHz band [66, 67].

Radio interference can be classified as co-channel and adjacent channel interference (ACI). The former is due to unwanted radio signals in overlapping frequencies with the desired signal. ACI is again of two types: One caused by transmitters on adjacent channels due to spectral leakage into the desired channel. The other type of ACI is due to the *nonlinear response of the receiver front-end* which causes the radio signals in adjacent channels to mix-up with the intended signal in the desired channel. This type of ACI primarily arises due to receiver imperfections and inherent non-linearity in their operation. The management of co-channel interference and ACI due to spectral leakage between coexisting systems has received much attention from several researchers [69–71]. However, there is not much literature on the management of ACI due to the nonlinear response of the receiver front-end.

In the conventional framework, receivers were protected from ACI during band planning and allocation by carefully-crafted guard bands customized to the technologies and receiver RF front-ends. This prevented adjacent channel signals from entering the receiver circuits. However, next-generation wireless networks will witness an unprecedented diversity in RATs and front-ends, accessing the same band of spectrum in a spatial-temporal neighborhood. Receivers in next generation dynamic spectrum access networks will have to encumber signals of unknown types, power levels, and spectrum masks on channels adjacent to its operation. While legacy receivers operating in opportunistic access bands were not designed to withstand the vagaries of dynamic spectrum access, even new radios can seldom be completely protected from hostile interference emanating from adjacent channels. Thus, the framework of allocating customized static guard bands col-

lapses when exceedingly diverse RATs need to be managed dynamically. Managing ACI due to the nonlinear response of the receiver front-end adds to the challenges for coexistence of wireless networks on a dynamic basis.

Considering the impact of receiver front end imperfections on the desired channels for resource allocation and network optimization is critical to the successful deployment and operations of agile opportunistic access networks. The adverse effects of allocating spectrum without accounting for receiver sensitivities is exemplified by the recent LightSquared (LS) controversy [72]. The company, LS, obtained a license to deploy LTE repeaters in a band adjacent to the civilian Global Positioning System (GPS) downlink and planned deployments worth close to $3B. However, post factum testing and analysis demonstrated that the LTE repeaters would potentially compromise GPS receiver operations because the latter poorly tolerated ACI. This resulted in the FCC suspending the license for LS. Note that FCC was only dealing with allocation of static bands in this case. If channel allocations remain agnostic to receiver characteristics, such issues will become rampant and magnified when disparate systems attempt to coexist dynamically. Multiple regulatory agencies and standardization bodies have also stated this fact repeatedly [73–77].

It is well acknowledged that efficient utilization of RF spectrum is of paramount importance. Network optimization methodologies that minimize the number of allocated channels and total transmit power whilst ensuring minimum data rate requirements are of immense interest for next-generation spectrum management systems (e.g., SAS). The significance of considering receiver performance in frameworks for efficient spectral utilization was first shown in [78]. In [80, 81], the authors proposed algorithms and frameworks to manage interference arising from solely from third order distortions. In [83], the authors proposed novel schemes for interference avoidance in wireless networks with nonlinear receivers approximated with third order polynomial model. The authors in [82] quantified the receiver nonlinearity from a spectrum-centric perspective. In [84], the linearity requirements for successful operations of opportunistic access cognitive radio for up to four blocking signals were addressed, assuming a third order nonlinear model with intermodulation distortion. However, these results largely concentrated on the distortions caused due to intermodulation spurs when the receiver operated in the weak nonlinear region.

In this chapter, we propose a scheme to minimize the number of allocated channels and total transmit power whilst ensuring minimum data rate requirements accounting for receiver image frequency and analog-to-digital converter (ADC) aliasing. We utilize the channelized spectrum representation models for image frequency rejection and ADC aliasing, described in [79,86] to formulate the optimization framework. The overall problem is in the form of Mixed-Integer-Linear-Programming (MILP) which is NP-hard, in general. Consequently, we develop an algorithm to solve it efficiently. The proposed algorithm decomposes the problem into two subproblems: channel allocation and power assignment. Then, it solves the decomposed problem iteratively using a two-phase structure. In Phase I, a particle swarm optimization (PSO)-based algorithm is tailored to obtain good solutions for the channel allocation subproblem. Phase II solves the power allocation subproblem optimally given the channel allocation solution obtained in Phase I. The algorithm alternates between the two phases until a stopping criterion is met.

Main contributions:

- We develop a receiver-characteristics-aware optimization framework for resource allocation in multi-RAT wireless networks.
- We provide insights on how to set the problem parameters under the proposed framework to control the trade-off between the number of allocated channels and the total power transmitted in the network.
- We show the effect of heterogeneous receiver characteristics on the overall utilization of network resources.
- We propose an efficient algorithm to solve the receiver-characteristic-aware joint channel allocation and power assignment problem. In this algorithm, the problem is solved iteratively, and in each iteration it is decomposed into two subproblems. A particle swarm optimization (PSO)-inspired algorithm is developed to solve the channel allocation subproblem. The power assignment subproblem is solved optimally given the solution of the channel allocation subproblem.

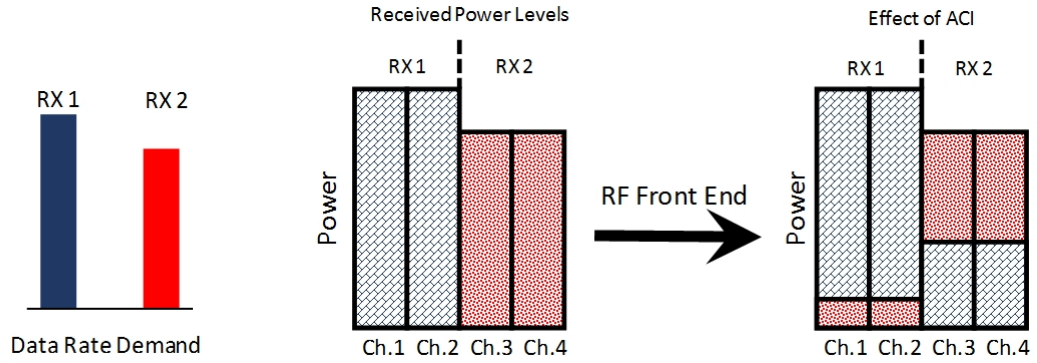


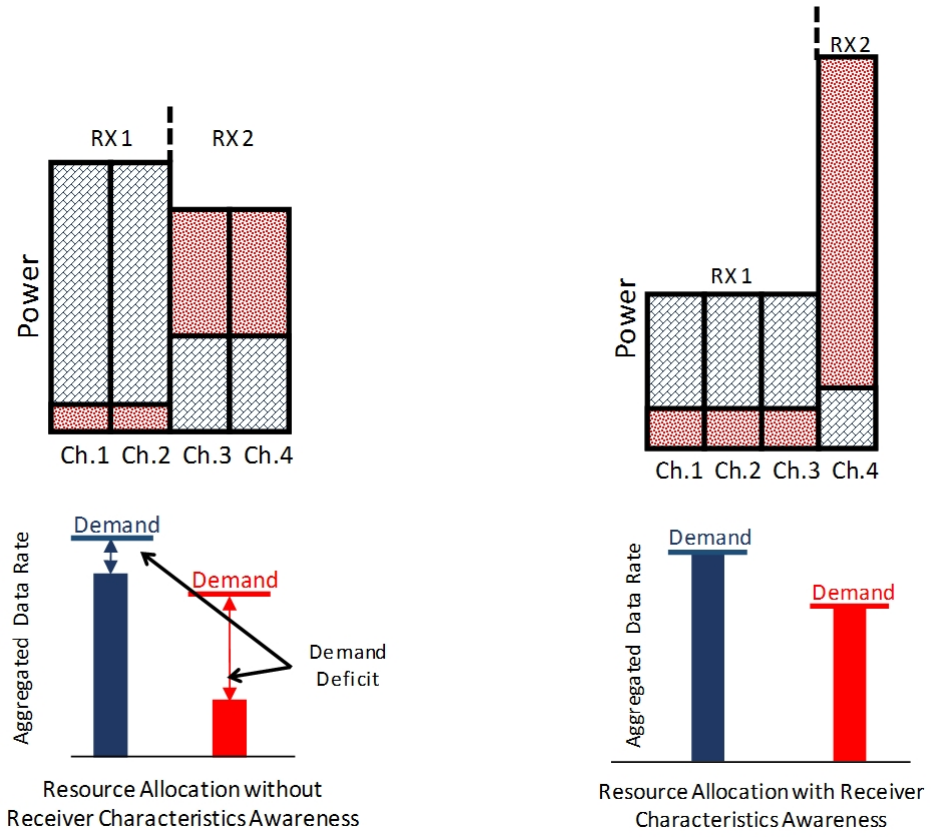
Figure 4.1: Adjacent channel interference due to receiver impairments

- We conduct extensive simulations to show the strength of the proposed optimization framework and the PSO-inspired algorithm for different network sizes.

The rest of the chapter is organized as follows. In Section 4.2, we motivate this work with an example. In Section 4.3, preliminaries on the sources of wireless receiver imperfections are provided. In Section 4.4, we introduce the system model and state our problem. In Section 4.5, we develop the proposed optimization framework. Section 4.7 shows numerical results and gives insights on how to set the configuration parameters of the proposed framework. In Section 4.8, we conclude our work and indicate directions for future research.

4.2 Motivational Example

In this section, we provide a qualitative explanation of the benefits of receiver characteristic aware resource allocation. For simplicity of explanation, consider two co-located receivers sharing four channels. Each receiver's link has its data rate requirement as shown in Fig. 4.1a. Assume that the impairments of Receiver 2 are much higher than Receiver 1. Thus, Receiver 2 will face higher adjacent channel interference than receiver 1. After the signals traverse the receiver front end, the interference faced by each receiver is illustrated in Fig. 4.1b.



(a) (b)

Figure 4.2: Illustrative diagram on the benefits of receiver-characteristics-aware resource allocation

Without the knowledge of receiver characteristics, we essentially are blind to the individual vulnerabilities of receivers. Thus, the spectral resources and power allocation for each receiver is proportional to its data rate demand only. However, since Receiver 2 is more vulnerable, it suffers a higher adjacent channel interference. Hence, such an allocation will potentially not be able to meet the individual Quality of Service (QoS) demands. This is shown in Fig. 4.2a. Receiver 2 faces enormous interference, and hence substantially falls short of the QoS demand.

Traditional responses to this will either be to (a) increase the number of allotted channels to Receiver 2, at the cost of QoS of Receiver 1, or (b) increase the power for Receiver 2, not only costing a higher power budget, but also in turn inflicting higher interference to Receiver 1. However, with the specific knowledge of the receiver impairments and vulnerabilities, an informed and efficient allocation may be carried out. As shown in Fig. 4.2b, as Receiver 1 can better handle

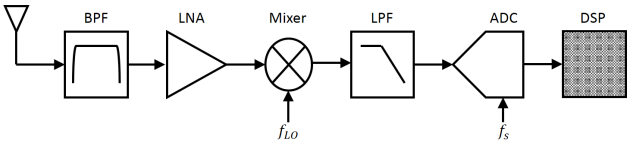


Figure 4.3: An example direct conversion receiver architecture.

adjacent channel interference, the power of Receiver 2 is increased on a single channel, while protecting it by allocating relatively lower power on adjacent channels for the other receiver. While this may result in distortions for Receiver 1, this along with lowered power are compensated by adding a channel to meet the required QoS demand. Thus, both receivers are potentially able to meet the QoS demand, while maintaining efficient spectral utilization and power budgets.

4.3 Preliminaries

In this section, we provide a synopsis of the receiver architecture considered and adjacent channel interference caused due to receiver imperfections. A Direct Conversion Receiver (DCR) architecture as shown in Fig. 4.3 is assumed. This is the most common architecture that will be used in next generation wireless systems [79, 87–89]. The front end Band Pass Filter (BPF) spans the entire range of RF frequencies in which the receiver is designed to operate. Thus, the desired signal bandwidth is typically a small fraction of the front end BPF. Moreover, since this is an RF filter, it generally exhibits poor selectivity. Consequently, unwanted signals from adjacent channels enter the receiver front end.

Certain aspects of receiver operation make then inherently non-linear (such as down conversion), and certain other front-end imperfections exhibit nonlinear behavior. This results in signals from adjacent channels interfering with the desired signal. Fig. 4.4 shows the transfer characteristics of a receiver. It can be largely divided into three regions: Linear, weak nonlinear, and strong nonlinear. The impact of distortions when the receiver operates in the nonlinear region was previously considered in [81]. However, optimal resource allocations accounting for ACI due to

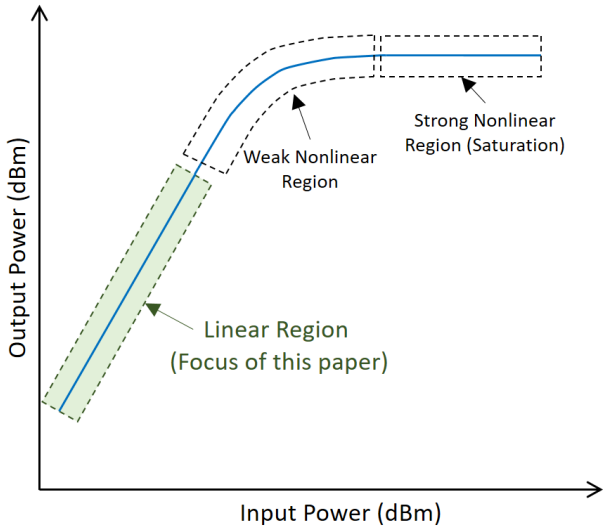


Figure 4.4: Transfer characteristics of a typical receiver.

receiver impairments in the linear region has not been addressed in the existing literature. In this chapter, we introduce a receiver-characteristics-aware resource allocation framework which takes into consideration receiver impairments occurring in the linear region of operation.

We consider two major receiver impairments that lead to ACI namely, incomplete image frequency rejection and imperfect sampling leading to ADC aliasing. Due to these effects, components of signals from adjacent channels appear in the desired channel at the output of the RF front end.

4.3.1 Impact of receiver impairments

In this section, we describe the quantitative formulation of the impact of the two receiver impairments based on the models developed in [79, 85, 86]. We describe the framework specific to our use case and direct readers to those references for further details.

Consider a receiver front end band pass filter spanning a set of contiguous channels $\mathcal{C} = \{1, 2, \ldots, C\}$ of equal bandwidth. Let $P_R[c]$ denotes the power received in channel $c \in \mathcal{C}$.

Table 4.1: Matrix representation example of receiver impairments for $C = 4$ and $B = 2$ [86].

Mixer	$\boldsymbol{P}_{\mathsf{mix}} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & \lambda \\ 0 & 1 & \lambda & 0 \\ 0 & \lambda & 1 & 0 \\ \lambda & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{Z}_{\mathsf{mix}}} \begin{bmatrix} P_R[1] \\ P_R[2] \\ P_R[3] \\ P_R[4] \end{bmatrix}$
ADC	$\boldsymbol{P}_{\mathsf{ADC}} = \underbrace{\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}}_{\mathbf{Z}_{\mathsf{ADC}}} \underbrace{\begin{bmatrix} \lvert G\rvert^2 & 0 & 0 & 0 \\ 0 & \lvert G\rvert^2 & 0 & 0 \\ 0 & 0 & \lvert G\rvert^2 & 0 \\ 0 & 0 & 0 & \lvert G\rvert^2 \end{bmatrix}}_{\mathbf{Z}_{\mathsf{AAF}}} \begin{bmatrix} P_{\mathsf{mix}}[1] \\ P_{\mathsf{mix}}[2] \\ P_{\mathsf{mix}}[3] \\ P_{\mathsf{mix}}[4] \end{bmatrix}$
Receiver Chain	$\boldsymbol{P}_{\mathsf{RF\_chain}} = \mathbf{Z}_{\mathsf{ADC}}\mathbf{Z}_{\mathsf{AAF}}\mathbf{Z}_{\mathsf{mix}}\boldsymbol{P}$, where $\boldsymbol{P} = [P_R[1]P_R[2]P_R[3]P_R[4]]^T$

Throughout the discussion in this chapter, we assume the receiver is operating in the linear region.

4.3.1.1 Mixer – image frequency rejection

We assume the local oscillator frequency to be the center of the front-end band pass filter. Imperfect mixing results in the image frequency signal component appearing at the desired channel c. Using the developments in [86], we re-formulate the mixer output at the desired channel $c \in \mathcal{C}$ specific to our use case as,

$$P_{\mathsf{mix}}[c] = P_R[c] + \lambda\, P_R[C - c + 1], \tag{4.1}$$

where $P_{\mathsf{mix}}[c]$ is the mixer output at channel c, and λ denotes the image frequency rejection ratio of the receiver.

4.3.1.2 ADC aliasing

The input signal to the ADC spans C channels. However, if there are only a set $\mathcal{B} = \{1, 2, \ldots, B\}$ of channels in the first Nyquist zone of the ADC, where $\mathcal{B} \subset \mathcal{C}, B < C$, there will be aliasing. The

set channels remaining at the output of the ADC can be obtained as,

$$\Psi_{B,C} = \left\{ \frac{C-B}{2} + 1, \frac{C-B}{2} + 2, \ldots, \frac{C+B}{2} \right\}. \tag{4.2}$$

In general, the mapping between c and b is given by

$$c = \left(\frac{C-B}{2} \right) + b;\ b \in [1, B]. \tag{4.3}$$

Assuming that an anti-aliasing-channel-select filter before the ADC filters out the unwanted frequencies, the output of this filter for channel c can be written as, $P_{\mathsf{AAF}}[c] = |G|^2 P_{\mathsf{mix}}[c]$, where $P_{\mathsf{AAF}}[c]$ is the anti-aliasing filter output at channel c, and G is the linear gain of the filter. We now re-formulate the analysis in [86] to obtain the ADC output for channel $c \in \mathcal{B}$ of the first Nyquist zone as:

$$P_{\mathsf{ADC}}[c] = \sum_{k=-\left\lceil \frac{C}{B} \right\rceil + 1}^{\left\lceil \frac{C}{B} \right\rceil - 1} P_{\mathsf{AAF}} \left[\left(\frac{C-B}{2} \right) + c - kB \right]. \tag{4.4}$$

4.3.1.3 Entire receiver chain

The output power of the channel c in the first Nyquist zone post sampling for the receiver chain can be expressed as:

$$P_{\mathsf{RF_{chain}}}[c] = \sum_{k=-\left\lceil \frac{C}{B} \right\rceil + 1}^{\left\lceil \frac{C}{B} \right\rceil - 1} \Big\{ |G|^2 \cdot (P_R[i - kB] + \lambda\, P_R[C - i + kB + 1]) \Big\}, \tag{4.5}$$

where $i = \left(\frac{C-B}{2}\right) + c$. The example matrix representation is shown in Table 4.1. In general, the output of each channel of the first Nyquist zone is:

$$\boldsymbol{P}_{\mathsf{RF_{chain}}} = \mathbf{Z}\boldsymbol{P}, \tag{4.6}$$

where $\mathbf{Z} = \mathbf{Z}_{\mathsf{ADC}}\mathbf{Z}_{\mathsf{AAF}}\mathbf{Z}_{\mathsf{mix}}$. Let $\mu_{c\hat{c}}$ represent the element of row c and column $\hat{c}$ of the matrix $\mathbf{Z}$. Post digitization, if the desired channel in the first Nyquist zone, then $\mu_{c\hat{c}}$ captures the impact of the signals in an adjacent channel $\hat{c} \in \mathcal{C} \setminus \{c\}$ on the desired channel c.

4.4 System Model and Problem Statement

We consider a set $\mathcal{M} = \{1, 2, \ldots, M\}$ of links belong to different networks in a multi-RAT environment. Each link $m \in \mathcal{M}$ can be assigned a transmission power on channel c (denoted by γ_c^m), where $0 \leq \gamma_c^m \leq \gamma_{\text{max}}$ and γ_{max} is the maximum total power that can be assigned to any link. We assume that the front end band pass filter of each link's receiver spans the whole $\mathcal{C}$. The channels allocated to link $m \in \mathcal{M}$ need to be contiguous. $\mathcal{C}_c^m$ represents a subset of channels so that a signal transmitted on channel $\hat{c} \in \mathcal{C}_c^m$ causes ACI to channel c at the receiver of link m with amount of $\mu_{c\hat{c}}^m$ of the signal power. Denote $\mathcal{I} = \{1, 2, \ldots, I\}$ as the set of data rates with which data transfer can be performed on any channel. The chosen data rate from this set depends on the received signal-to-interference-plus-noise (SINR) value on the corresponding channel taking into consideration intended signal's power and ACI.

4.4.1 Channel effect on the received signal strength

The received signal strength is affected by the channel between the transmitter and the receiver. Channel effect is mainly due to path loss and small scale fading. We assume that each link's transmitter and receiver is independently positioned according to a Binomial Point Process (BPP) [90]. The path loss coefficient, denoted by PL^{mn}, is a function of the distance d^{mn} between transmitter n and receiver m, and path loss exponent α (i.e. $PL^{mn} = (d^{mn})^{-\alpha}$). Small scale propagation is modeled using a Rayleigh fading model. We denote the overall channel coefficient between transmitter n and receiver m on channel c as h_c^{mn}, and h_c^{mn} can be obtained by multiplying the path loss coefficient by the Rayleigh fading coefficient.

Table 4.2: Notation.

Sets and Indices:

$\mathcal{M}$	Set of links in the network
$\mathcal{C}$	Set of available contiguous channels in the network
$\mathcal{C}_c^m$	Subset of channels so that a signal transmitted on channel $\hat{c} \in \mathcal{C}_c^m$ causes ACI to channel $c \in \mathcal{C}$ at the receiver of link $m \in \mathcal{M}$
$\mathcal{I}$	Set of data rates that can be supported on any channel

Data:

B^m	Maximum number of channels that can be allocated to link $m \in \mathcal{M}$ which corresponds to the bandwidth of the ADC filter
γ_{max}	Maximum total power that can be assigned any link
η_i	SINR threshold level i
$\mu_{c\hat{c}}^m$	ACI effect on the receiver of link $m \in \mathcal{M}$ caused by channel $\hat{c}$' signal on channel $c \in \mathcal{C}$
h_c^{nm}	channel coefficient between the transmitter of link $n \in \mathcal{M}$ and the receiver of link $m \in \mathcal{M}$ on channel $c \in \mathcal{C}$
σ^2	Average power of the background noise
D^m	Data rate demand on link $m \in \mathcal{M}$

Decision Variables:

x_c^m	Binary variable to indicate whether or not channel $c \in \mathcal{C}$ is allocated to link $m \in \mathcal{M}$
γ_c^m	Assigned power to link $m \in \mathcal{M}$ on channel $c \in \mathcal{C}$
R_c^m	Achievable rate on channel $c \in \mathcal{C}$ if allocated to link $m \in \mathcal{M}$

4.4.2 Problem statement

Each link $m \in \mathcal{M}$ has a rate demand (denoted by D^m) to be met through aggregating data rates on its allocated channels. The main objective is to minimize the total number of utilized channels. Also, minimizing the total amount of power transmission from all links is desired. The overall objective is a linear combination of aforementioned two objectives while meeting the data rate demands of all links. Table 4.2 summarizes the important notation used in this chapter.

4.5 Optimization Framework

In this section, we present the proposed optimization framework which accounts for the unique characteristic of each link's receiver. We start with the problem constraints, then we introduce the objective function.

4.5.1 Optimization constraints

4.5.1.1 Avoiding co-channel interference between links

Denote x_c^m as a binary variable to indicate whether or not channel c is allocated to link m. Two links cannot utilize the same channel. This can be represented using the following set of constraints:

$$\sum_{m \in \mathcal{M}} x_c^m \leq 1, \quad \forall c \in \mathcal{C}. \tag{4.7}$$

4.5.1.2 Restricting maximum number of channels per link

The maximum number of channels that can be allocated to link m, denoted as B^m, is proportional to the bandwidth of the ADC filter.

$$\sum_{c \in \mathcal{C}} x_c^m \leq B^m, \quad \forall m \in \mathcal{M}. \tag{4.8}$$

4.5.1.3 Contiguity of link's allocated channels

All channels allocated to a specific link have to be contiguous. If a link utilizes a channel c and other channels, the consecutive channel $(c+1)$ or $(c-1)$ has to be allocated to the same link. For example, link m seizes two channels where one of them is channel 1, the other one has to be channel 2. Considering each channel in the set indexed from 1 to $C-1$, this restriction can be

mathematically expressed as follows:

$$\text{If } \left\{x_c^m = 1\right\} \text{ and } \left\{\sum_{\hat{c}>c} x_{\hat{c}}^m \geq 1\right\} \Rightarrow \left\{x_{c+1}^m = 1\right\}.$$

We first introduce an indicator variable
$y_c^m = \mathbb{1}_{\{\sum_{\hat{c}>c} x_{\hat{c}}^m\}}$ so that:

$$\left\{\sum_{\hat{c}>c} x_{\hat{c}}^m \geq 1\right\} \Leftrightarrow \left\{y_c^m = 1\right\}.$$

The forward relationship can be expressed as follows:

$$\sum_{\hat{c}>c} x_{\hat{c}}^m \leq (M_c + 1)\ y_c^m, \quad \forall c \in \mathcal{C} \setminus \{C\}, m \in \mathcal{M}, \tag{4.9}$$

where M_c is an upper bound on $\sum_{\hat{c}>c} x_{\hat{c}}^m - 1$. The backward relationship can be formulated as:

$$\sum_{\hat{c}>c} x_{\hat{c}}^m + m_c\ y_c^m \geq m_c + 1, \quad \forall c \in \mathcal{C} \setminus \{C\}, m \in \mathcal{M}, \tag{4.10}$$

where m_c is a lower bound on $\sum_{\hat{c}>c} x_{\hat{c}}^m - 1$. It is easy to see that $m_c = -1, \forall c \in \mathcal{C}$.

Then, the original expression can be represented as:

$$\text{If } \left\{x_c^m = 1\right\} \text{ and } \left\{y_c^m = 1\right\} \Rightarrow \left\{x_{c+1}^m = 1\right\},$$

which can be formulated as follows:

$$x_{c+1}^m \geq x_c^m + y_c^m - 1, \quad \forall c \in \mathcal{C} \setminus \{C\}, m \in \mathcal{M}. \tag{4.11}$$

4.5.1.4 Relationship between channel allocation and power transmission

A link is to transmit with a non-zero power level only on its allocated channels. This restriction can be modeled using:

$$\gamma_c^m \leq \gamma_{\text{max}}\, x_c^m, \quad \forall c \in \mathcal{C}, m \in \mathcal{M}. \tag{4.12}$$

4.5.1.5 Restricting link total power transmission

The total power transmitted from a link across its allocated channels should not exceed a predefined limit γ_{max}. This can be expressed as follows:

$$\sum_{c \in \mathcal{C}} \gamma_c^m \leq \gamma_{\text{max}}, \quad \forall m \in \mathcal{M}. \tag{4.13}$$

4.5.1.6 Link demand constraints

Each link's data rate demand must be met. Denote R_c^m as the achievable data rate on channel c if allocated to link m. This can be expressed as follows:

$$\text{If } x_c^m = 0 \text{ then } R_c^m = 0,$$
$$\text{else if } x_c^m = 1 \text{ then } R_c^m \geq 0.$$

The link demand constraints are:

$$\sum_{c \in \mathcal{C}} R_c^m \geq D^m, \quad \forall m \in \mathcal{M}. \tag{4.14}$$

4.5.1.7 SINR-to-rate mapping constraints

In practical wireless systems, each range of SINR values is mapped to a single Modulation and Coding Scheme (MCS) which defines the resulting date rate. Assume that there are I MCSs

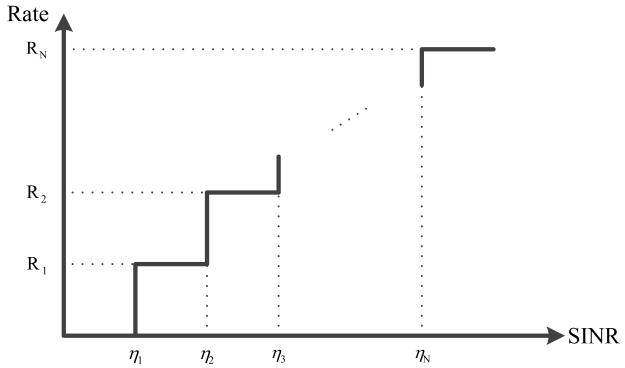


Figure 4.5: SINR-to-Rate mapping.

which correspond to the same number of supported data rates. Each SINR range is mapped to an achievable date rate value as shown in Fig. 4.5. In the context of our problem, the SINR at the receiver of link m on channel c can be expressed as:

$$\text{SINR}_c^m = \frac{h_c^{mm}\gamma_c^m}{\sigma^2 + \sum\limits_{\hat{c}\in\mathcal{C}_c^m} \mu_{c\hat{c}}^m \left(\sum\limits_{\substack{n\in\mathcal{M} \\ n\neq m}} h_{\hat{c}}^{nm}\gamma_{\hat{c}}^n \right)}, \quad \forall c \in \mathcal{C}, m \in \mathcal{M},$$

where σ^2 is the average power of the background noise, $h_{\hat{c}}^{nm}$ is the channel coefficient between the transmitter of link n and the receiver of link m on channel $\hat{c}$. When the SINR is below a minimum threshold η_1, the rate is set to zero. This can be expressed as follows. For each $c \in \mathcal{C}, m \in \mathcal{M}$:

$$\frac{h_c^{mm}\gamma_c^m}{\sigma^2 + \sum\limits_{\hat{c}\in\mathcal{C}_c^m} \mu_{c\hat{c}}^m \left(\sum\limits_{\substack{n\in\mathcal{M} \\ n\neq m}} h_{\hat{c}}^{nm}\gamma_{\hat{c}}^n \right)} < \eta_1 \Rightarrow R_c^m = 0.$$

The strict inequality can be converted to a non-strict one by adding a small number ε. Rearranging the terms and introducing a new binary variable t_c^m, the constraint can be represented as follows:

$$h_c^{mm}\gamma_c^m - \eta_1 \sum_{\hat{c}\in\mathcal{C}_c^m} \mu_{c\hat{c}}^m \left(\sum_{\substack{n\in\mathcal{M} \\ n\neq m}} h_{\hat{c}}^{nm}\gamma_{\hat{c}}^n \right) + \varepsilon \leq \eta_1\sigma^2 \Rightarrow t_c^m = 1 \Rightarrow R_c^m = 0,$$

which can be modeled using the following two inequalities:

$$h_c^{mm}\gamma_c^m - \eta_1 \sum_{\hat{c}\in\mathcal{C}_c^m} \mu_{c\hat{c}}^m \left(\sum_{\substack{n\in\mathcal{M}\\ n\neq m}} h_{\hat{c}}^{nm}\gamma_{\hat{c}}^n \right) + \varepsilon - (E_c^m - \epsilon_0)\ t_c^m \geq \eta_1\sigma^2 + \epsilon_0, \quad \forall c \in \mathcal{C}, m \in \mathcal{M}, \tag{4.15}$$

where $E_c^m = -\eta_1 \sum_{\hat{c}\in\mathcal{C}_c^m} \mu_{c\hat{c}}^m \left(\sum_{\substack{n\in\mathcal{M}\\ n\neq m}} h_{\hat{c}}^{nm}\gamma_{\text{max}} \right) - \eta_1\sigma^2$ and ϵ_0 is a small constant.

$$R_c^m \leq (1 - t_c^m)\ R_I, \quad \forall c \in \mathcal{C}, m \in \mathcal{M}. \tag{4.16}$$

When the SINR value exceeds the minimum threshold η_1, we can represent the relationship between the SINR and the achievable rate as follows. Setting $R_0 = 0$, for each $c \in \mathcal{C}, m \in \mathcal{M}, i \in \mathcal{I}$:

$$\frac{h_c^{mm}\gamma_c^m}{\sigma^2 + \sum_{\hat{c}\in\mathcal{C}_c^m} \mu_{c\hat{c}}^m \left(\sum_{\substack{n\in\mathcal{M}\\ n\neq m}} h_{\hat{c}}^{nm}\gamma_{\hat{c}}^n \right)} \geq \eta_i \Leftrightarrow R_c^m > R_{i-1}.$$

Rearranging the terms and converting the strict constraint to non-strict one (using the small constant ε):

$$h_c^{mm}\gamma_c^m - \eta_i \sum_{\hat{c}\in\mathcal{C}_c^m} \mu_{c\hat{c}}^m \left(\sum_{\substack{n\in\mathcal{M}\\ n\neq m}} h_{\hat{c}}^{nm}\gamma_{\hat{c}}^n \right) \geq \eta_i\sigma^2 \Leftrightarrow R_c^m \geq R_{i-1} + \varepsilon.$$

We introduce a new binary variable $s_{c,i}^m$ to break down the relationship between the two expressions as follows:

$$h_c^{mm}\gamma_c^m - \eta_i \sum_{\hat{c}\in\mathcal{C}_c^m} \mu_{c\hat{c}}^m \left(\sum_{\substack{n\in\mathcal{M}\\ n\neq m}} h_{\hat{c}}^{nm}\gamma_{\hat{c}}^n \right) \geq \eta_i\sigma^2 \Leftrightarrow s_{c,i}^m = 1,$$

and

$$s_{c,i}^{m} = 1 \Leftrightarrow R_{c}^{m} \geq R_{i-1} + \varepsilon.$$

Focusing on the first relationship, the forward part can be modeled as follows:

$$h_{c}^{mm}\gamma_{c}^{m} - \eta_{i} \sum_{\hat{c} \in \mathcal{C}_{c}^{m}} \mu_{c\hat{c}}^{m} \left(\sum_{\substack{n \in \mathcal{M} \\ n \neq m}} h_{\hat{c}}^{nm}\gamma_{\hat{c}}^{n} \right) - \eta_{i}\sigma^{2} \leq \left(J_{i}^{m} + \epsilon\right) s_{c,i}^{m} - \epsilon, \quad \forall c \in \mathcal{C}, m \in \mathcal{M}, i \in \mathcal{I}, \tag{4.17}$$

where $J_{i}^{m} = h_{c}^{mm}\gamma_{\max} - \eta_{i}\sigma^{2}$ and ϵ is a small constant. The backward part can be represented using the following constraint.

$$h_{c}^{mm}\gamma_{c}^{m} - \eta_{i} \sum_{\hat{c} \in \mathcal{C}_{c}^{m}} \mu_{c\hat{c}}^{m} \left(\sum_{\substack{n \in \mathcal{M} \\ n \neq m}} h_{\hat{c}}^{nm}\gamma_{\hat{c}}^{n} \right) - \eta_{i}\sigma^{2} \geq q_{c,i}^{m} \left(1 - s_{c,i}^{m}\right), \quad \forall c \in \mathcal{C}, m \in \mathcal{M}, i \in \mathcal{I}, \tag{4.18}$$

where $q_{c,i}^{m} = -\eta_{i} \sum\limits_{\hat{c} \in \mathcal{C}_{c}^{m}} \mu_{c\hat{c}}^{m} \left(\sum\limits_{\substack{n \in \mathcal{M} \\ n \neq m}} h_{\hat{c}}^{nm}\gamma_{\max} \right) - \eta_{i}\sigma^{2}$.

Now, focusing on the second relationship, the forward part can be modeled as follows:

$$R_{c}^{m} \geq \left(R_{i-1} + \varepsilon\right) s_{c,i}^{m}, \quad \forall c \in \mathcal{C}, m \in \mathcal{M}, i \in \mathcal{I}. \tag{4.19}$$

The backward part can be modeled as:

$$R_{c}^{m} \leq R_{i-1} + \varepsilon - \epsilon + \left(Q_{i} + \epsilon\right) s_{c,i}^{m}, \quad \forall c \in \mathcal{C}, m \in \mathcal{M}, i \in \mathcal{I}, \tag{4.20}$$

where $Q_{i} = R_{I} - R_{i-1} - \varepsilon$.

4.5.2 Optimization objective

Minimizing the spectrum usage allows other networks to operate and coexist with SAS networks. On the other hand, minimizing the network power consumption saves energy costs and prolongs

battery life in mobile networks. Also, it facilitates better coexistence with non-SAS networks because this will minimize the ACI caused to those networks. The goal of our problem then is to balance the minimization of the network power consumption and spectrum usage through a controlling parameter β. The objective function can be formulated as:

$$\min \left\{ \sum_{m \in \mathcal{M}} \sum_{c \in \mathcal{C}} x_c^m + \beta \sum_{m \in \mathcal{M}} \sum_{c \in \mathcal{C}} \gamma_c^m \right\}.$$

The overall joint power and band allocation problem with receiver characteristic awareness can be expressed as a mixed integer linear program (MILP):

Joint Power and Band Allocation Problem with Receiver Characteristic Awareness

$$\underset{\{x_c^m, \gamma_c^m, R_c^m, c \in \mathcal{C}, m \in \mathcal{M}\}}{\text{minimize}} \left\{ \sum_{m \in \mathcal{M}} \sum_{c \in \mathcal{C}} x_c^m + \beta \sum_{m \in \mathcal{M}} \sum_{c \in \mathcal{C}} \gamma_c^m \right\}$$

subject to:

Avoiding co-channel interference between links: (4.7);
Restricting maximum number of channels per link: (4.8);
Contiguity of link's allocated channels: $(4.9) - (4.11)$;
Relationship between channel allocation and power transmission: (4.12);
Restricting link total power transmission: (4.13);
Link demand constraints: (4.14);
SINR-to-rate mapping constraints: $(4.15) - (4.20)$.

4.6 A Proposed Algorithm to Solve the Joint Power and Band Allocation Problem

As discussed in the previous section, the joint power and band allocation problem is in the form of MILP which is NP-hard, in general. Branch-and Bound (B&B)-based methods can solve this MILP problem only for small network sizes. That is, in the basic form of the B&B method, a search tree is constructed by fixing one or more binary decision variables at the value of zero or one. For larger network sizes, the number of variables becomes dramatically big. Consequently, the B&B search tree grows fast and such methods fail to obtain any solution. In this section, we propose a novel framework to solve the overall problem efficiently.

4.6.1 Algorithm overview

The proposed framework solves the problem iteratively using a two-phase structure. In each iteration, Phase I implements a selection criterion to determine the band allocation for each link in the network. Then, Phase II computes the optimal power allocation given the band allocation in Phase I. Here, we use a suboptimal algorithm only to obtain solutions for the band allocation subproblem while solving the power allocation subproblem to optimality. The overall solution is fed back to Phase I of the next iteration to determine the new band allocation. The algorithm continues until a stopping criterion is reached. The goal of limiting the use of a suboptimal algorithm to find solutions to the channel allocation subproblem (as opposed to both channel and power allocation) is twofold. Firstly, possible sacrifice of the first subproblem's optimality is compensated by solving the power allocation subproblem to optimality. Secondly, the number of variables in the channel allocation subproblem is small compared to those of the overall problem. This leads to limiting the suboptimal algorithm's search space which leads to obtaining good solutions within small number of iterations. The overall framework is shown in Fig. 4.6.

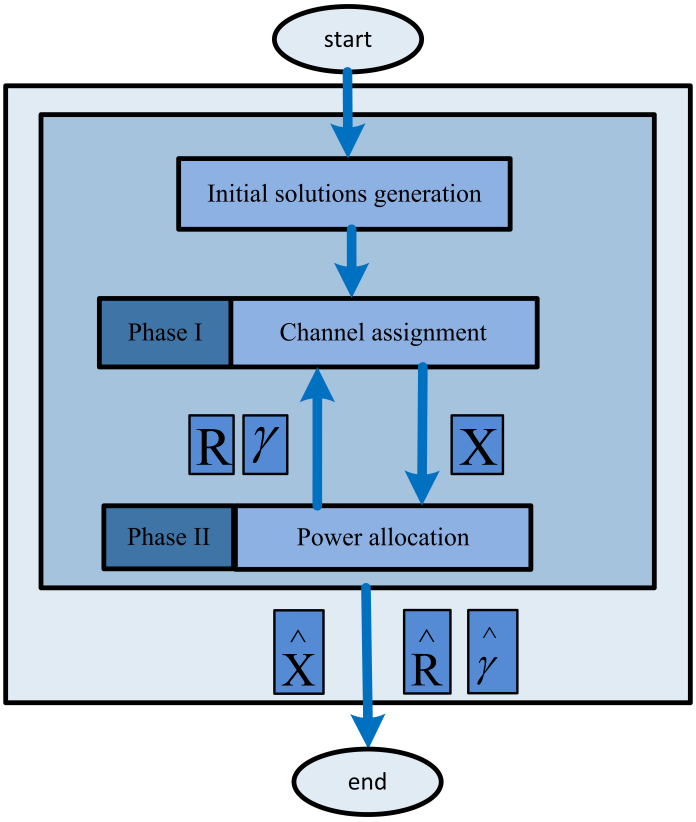


Figure 4.6: Proposed framework.

4.6.2 A PSO-inspired method for band allocation

Here, we introduce an efficient algorithm based on particle swarm optimization (PSO) to solve the channel allocation subproblem. PSO is a global search algorithm inspired by the social attitude of fishes schooling, animals herd and birds flock in which the groups search cooperatively for food [53]. Compared to other search algorithms (such as genetic algorithm and ant colony optimization), PSO is easier to implement and has fewer parameters to control. Over the last decade, PSO has been identified as an effective way to find good solutions for NP-hard problems [54–56]. In the original form of the algorithm, a set of particles explores the search space looking for good solutions. Each particle represents a current solution and its movement in each step is affected by the best solution found on its path and the best solution ever found among all particles. The particle definition is suitable for problems where one homogenous set of variables (e.g. channel index, ...) is changing in each iteration. However, in our problem, we modify two different sets of variables independently, namely lower-end and upper-end channels. Therefore, we introduce a "two-particle structure" where each particle represents one of the aforementioned two sets. In each two-particle structure, each upper-end channel is modified independently but within maximum distance from the corresponding lower-end channel. A PSO-based algorithm needs a set of initial solutions to start the iterating process. In the sequel, we describe how to determine these initial solutions.

4.6.2.1 Initial solution generation

Algorithm 1 shows the pseudocode for the initial solution generation. Each initial solution $i \in \mathcal{I}$ is found as follows. Denote the lower-end channel of link $m \in \mathcal{M}$ as $L_{i,m}$. A link m is chosen randomly and its $L_{i,m}$ is set to the value of f (initialized with the value zero) (lines 4-8). The number of contiguous channels assigned to the link m is chosen randomly between one and B^m (line 9). Denote upper-end channel allocated for link $m \in \mathcal{M}$ of solution i as $U_{i,m}$. A new lower-end channel is chosen at a random number, say between two chosen predefined numbers a and b, of channels apart from the $U_{i,m}$ of the previously selected link (line 12). Another link is randomly chosen among the remaining links (which have not yet been assigned a band yet) and assigned

the new calculated lower-end channel. The procedure continues until each link is assigned a band of contiguous channels. We then fix all channel-allocation-related variables (x_c^m and y_c^m) to the value zero or one given the obtained channel allocation in the previous steps. Then, the power allocation subproblem is solved using a B&B method (line 14). If the problem is infeasible, the demand of each link is decremented and the overall problem is resolved. The process continues until a feasible solution to the reduced-demand problem is found (lines 15-18). Please note that as we decrease the demand values, it is more likely to find a feasible solution. The resulting data rates on the allocated channels will not satisfy the original link demands. We count for this "demand deficit" using a fitness function as explained below.

Denote S_i and $\hat{S}$ as the current solution of the two-particle structure i, the best achieved position (solution) of the two-particle structure i, and overall best solution, respectively. To find the best solution among all initial solutions, we define a fitness function (denoted by F) weighted summation of the objective function and the total deficit in demand satisfaction as follows:

$$F = \sum_{m\in\mathcal{M}} \sum_{c\in\mathcal{C}} x_c^m + \beta \sum_{m\in\mathcal{M}} \sum_{c\in\mathcal{C}} \gamma_c^m + \zeta \sum_{m\in\mathcal{M}} \left(D^m - \sum_{c\in\mathcal{C}} R_c^m \right)^+, \tag{4.21}$$

where ζ is a weighting factor, and its value is set such that a unit of demand dissatisfaction weighs out the maximum value of the original objective function (i.e. $\zeta \geq C + \beta\gamma_{\max} M$)). The third term in Equation 4.21 is the total deficit in demand satisfaction. Denote F_i and $\hat{F}$ as the fitness function values corresponding to S_i and $\hat{S}$, respectively. The best solution ($\hat{S}$) will have the minimum F_i value among all obtained initial solutions (line 21).

4.6.2.2 Core PSO-inspired function

Now, we describe the core part of the proposed algorithm for Phase I. Algorithm 2 shows the pseudocode for the overall procedure. Denote $\bar{S}_i$ and $\bar{F}_i$ as the best achieved position of the two-particle structure i and the corresponding fitness function value. The algorithm starts each iteration by determining the coordinates of the two-particle structure (the solution vector) $i \in \mathcal{I}$ through the

Algorithm 3 Procedure of initial solutions generation

for $i \in \mathcal{I}$ **do**
  Set $\mathcal{H} = \mathcal{M}$; $f = 0$;
  while $\mathcal{H} \neq \emptyset$ **do**
    Pick a link m randomly;
    while $m \notin \mathcal{H}$ **do**
      Pick another link m randomly;
    end while
    $L_m := f$;
    $U_m := L_m + rand(0, B^m - 1)$;
    $\mathcal{H} := \mathcal{H} \setminus m$;
    $f := U_m + rand(a, b)$;
  end while
  Solve the power allocation subproblem using a B&B method;
  while S_i is infeasible **do**
    Reduce the link demands;
    Solve the power allocation subproblem using a B&B method;
  end while
  Calculate F_i;
end for
$\hat{S} = \underset{S_i, i \in \mathcal{I}}{\text{argmin}}(F_i)$;

following four equations (lines 5-10):

$$v1_{i,m} := v1_{i,m} + w_g \cdot (\hat{S}(m) - L_{i,m}) + w_l \cdot (\bar{S}_i(m) - L_{i,m}), \tag{4.22}$$

$$v2_{i,m} := v2_{i,m} + w_g \cdot (\hat{S}(m) - U_{i,m}) + w_l \cdot (\bar{S}_i(m) - U_{i,m}), \tag{4.23}$$

$$L_{i,m} := \text{max}(0, \text{min}(L_{i,m} + \text{round}(v1_{i,m}), C)), \tag{4.24}$$

$$U_{i,m} := \text{max}(L_{i,m}, \text{min}(U_{i,m} + \text{round}(v2_{i,m}), \text{min}(C, L_{i,m} + B^m - 1))), \tag{4.25}$$

where $v1_{i,m}$ and $v2_{i,m}$ are the two velocity components of two-particle structure i in the m^{th} dimension, w_g and w_l are the global and local weighting factors, respectively, $L_{i,m}, U_{i,m}, \bar{S}_i(m)$ and $\hat{S}(m)$ are the m^{th} entries in lower-end and upper-end channels in the current solution vector i, best found solution for the two-particle structure i, and the best solution among all two-particle structures, respectively. The global weighting factor determines how much the difference between a two-particle structure's current solution and the best ever-found solution ($\hat{S}$) affects its next movement of the two-particle structure. Similarly, the local weighting factor determines how much the difference between a two-particle structure i's current solution and its best obtained solution ($\bar{S}_i$) affects its next movement. Note that in Equation 4.24, the updated value of $v1_{i,m}$ is rounded because the two-particle structure move in any direction is restricted to be an integer number of steps since it represents a channel index. Also, Equation 4.24 avoids moving the two-particle structure outside the search space in any of the M directions through limiting the resulting value of $L_{i,m}$ in the range [1, C]. In Equation 4.25, three restrictions on the updated value of $U_{i,m}$ are enforced: (i) $U_{i,m} \geq L_{i,m}$, (ii) $U_{i,m} \leq C$, and (iii) $U_{i,m} \leq L_{i,m} + B^m - 1$. The first restriction keeps the upper-end channel at least as high as the lower-end channel. The second restriction prevents the upper-end channel from being allocated outside the available range of frequencies. The third restriction keeps the number of allocated channels for any link $m \in \mathcal{M}$ at most B^m as stated in Constraint 4.8 in the original problem formulation.

If the channel allocation results in overlapping between links' channels, it is resolved as follows. A link is selected randomly and its allocated channels are not changed. Another link is

selected, and its overlapping channels with any of the previously selected links are deallocated. The process continues until all links are selected. Similar to the initial solution procedure, all channel-allocation-related variables (x_c^m and y_c^m) are then fixed to the value zero or one given the obtained channel allocation in the previous steps. Then, the power allocation subproblem is solved using a B&B method. If the problem is infeasible, the demand of each link is decremented and the overall problem is resolved. The process continues until a feasible solution to the reduced-demand problem is found (lines 12-16). The fitness function of the current solution is then evaluated (line 17). Then, the best achieved position (solution) and the fitness function value for each two-particle structure $(\bar{S}_i, \bar{F}_i)$ are updated, if needed (lines 18-21). At the end of the iteration, the overall best solution ($\hat{S}$) is updated (lines 23). The algorithm terminates when the maximum number of iterations (denoted by Max-Iterations) is reached.

4.6.3 Power allocation function

The goal of this function is to determine the power allocation on each allocated channel for each link in the network. Suboptimal algorithms can be designed to solve this power allocation subproblem. However, we show here how to obtain an "optimal" solution for this subproblem. Revisiting the original problem formulation, we can now set all the binary variables x_c^m and y_c^m, $\forall m \in \mathcal{M}, c \in \mathcal{C}$ either to the value one or zero given the obtained channel allocation in Phase I. We call the resulting formulation the "reduced problem". Although the reduced problem formulation is still MILP due to the fact that other discrete variables are still not fixed, the number of these decision variables is relatively small compared to the original joint channel and power allocation problem. Then, it becomes relatively easier to solve it using B&B methods. This enables us to obtain the "optimal" power allocation for the reduced problem. As we will discuss in Section 4.7, our experiments show that state-of-the-art optimizers (which implement B&B methods) can easily obtain the optimal solution for the reduced problem.

Algorithm 4 PSO-inspired algorithm for band allocation

Find a set of $\mathcal{I}$ initial solutions using Algorithm 1;
Set $k = 0, \bar{F}_i = F_i, \forall i \in \mathcal{I}$;
while $k \leq$ Max-Iterations **do**
  for $i \in \mathcal{I}$ **do**
    for $m \in \mathcal{M}$ **do**
      $v1_{i,m} := v1_{i,m} + w_g \cdot (\hat{S}(m) - L_{i,m}) + w_l \cdot (\bar{S}_i(m) - L_{i,m})$;
      $v2_{i,m} := v2_{i,m} + w_g \cdot (\hat{S}(m) - U_{i,m}) + w_l \cdot (\bar{S}_i(m) - U_{i,m})$;
      $L_{i,m} := \max(0, \min(L_{i,m} + \text{round}(v1_{i,m}), C))$;
      $U_{i,m} := \max(L_{i,m}, \min(U_{i,m} + \text{round}(v2_{i,m}), \min(C, L_{i,m} + B^m - 1)))$;
    end for
    Resolve any resulting overlapping in the channel allocation.
    Solve the power allocation subproblem using a B&B method;
    while S_i is infeasible **do**
      Reduce the link demands;
      Solve the power allocation subproblem using a B&B method;
    end while
    Calculate F_i;
    if $\bar{F}_i > F_i$ **then**
      $\bar{S}_i := S_i$;
      $\bar{F}_i := F_i$;
    end if
  end for
  $\hat{S} = \underset{\bar{S}_i, i \in \mathcal{I}}{\text{argmin}}(\bar{F}_i)$;
  $k := k + 1$;
end while

4.7 Performance Evaluation

In this section, we show the performance of the proposed optimization framework under both homogeneous and heterogenous network setups. Besides, we show how to control the trade-off between channel allocation and power assignment. Then, we discuss the performance of the proposed PSO-based algorithm.

4.7.1 Evaluation setup

In our experiments, we use two network setups. The first network setup is a small one and is used to demonstrate the benefits of the proposed receiver-characteristic-aware framework. Moreover, it shows the performance of the PSO-inspired algorithm compared to the optimal solution. We consider the network topology shown in Fig. 4.7. This network consists of five links belonging to different wireless systems where all transmitters and receivers exist within a square area with side length of 100 m. The other network setup is used to show the performance of the PSO-inspired algorithm for larger network sizes. This network consists of 10 links as shown in Fig. 4.8 where all transmitters and receivers exist within a square area with side length of 100 m. In both network setups, the location of each link's transmitter is generated using a binomial point process (BPP). Each receiver's location is randomly generated within a distance between 20-50 m from its transmitter's location. The number of available channels is set to 20 and 50 for the first and second network setups, respectively. The number of SINR levels are four. $\gamma_{\max}$ is set to 1 watt. The path loss exponent (α) is 3.5. Background noise power is 10^{-7} watts. The demand of each link is a parameter in our simulations. In each experiment, we assume the same demand for all links, i.e. $D_m \stackrel{\text{def}}{=} D, \forall m \in \mathcal{M}$. Each receiver's ADC is assumed to process signals within at most four channels. Also, each of the two receiver characteristics parameters (namely, λ and G) are the same across all channels for each receiver, and are randomly generated from two sets as follows. If the mixer is marked as 'good', λ is selected randomly in the range [0.001, 0.01]. In case of 'bad' mixers, the value is generated in the range [0.01, 0.05]. On the other hand, G is generated within

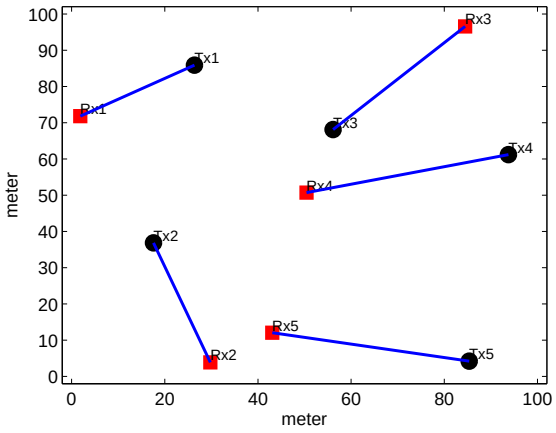


Figure 4.7: Network topology 1.

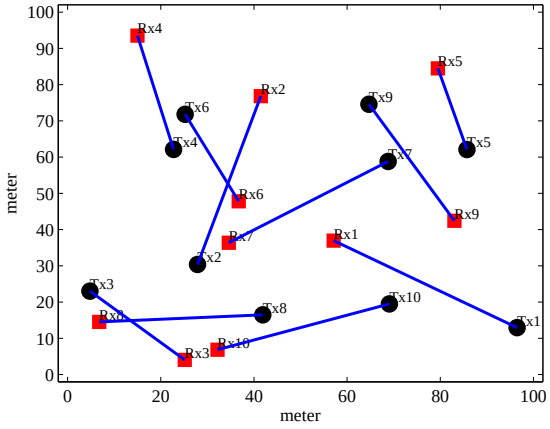


Figure 4.8: Network topology 2.

the range [0.01, 0.1] when the receiver's ADC is 'good'. For 'bad' ADCs, the value is chosen from the range [0.1, 0.3]. These ranges have been carefully chosen in consistence with the experimental results obtained at Wireless @ VT labs. 'Bad Receivers' refers to the class of receivers where both mixers and ADCs are 'bad'. Table 4.3 summarizes the parameter settings in our simulations.

We used CPLEX to solve our optimization problems. All experiments were run using a cluster at Virginia Tech, called BlueRidge [23]. More specifically, each experiment was executed on a single node of BlueRidge that has 16 processors (utilized by CPLEX when possible) and 64 GB memory. The running time of each experiment ranges from few seconds to few minutes. We use the following metrics to evaluate our proposed allocation scheme: (i) total number of allocated

Table 4.3: System parameter values.

Parameter	Value
γ_{max}	1 watt
α	3.5
σ^2	10^{-7}
$\lambda - good$	[0.001-0.01]
$\lambda - bad$	[0.01-0.05]
$G - good$	[0.01-0.1]
$G - bad$	[0.1-0.3]

Table 4.4: Algorithm parameter values.

Parameter	Value
(a, b)	(1,3)
w_g	0.1
w_l	0.1
ζ_1	$5\beta + 20$
ζ_2	$10\beta + 50$

channels (C_{total}), and (ii) total amount of power generated collectively by all links (P_{total}).

4.7.2 The case of homogeneous networks

Here, networks in which all receivers have similar characteristics are referred to as homogeneous networks. Our initial simulations revealed that a naive optimization framework that does not account for the receiver characteristics fails to satisfy link demands in most cases. Under our proposed framework, all demands are satisfied.

Table 4.5 shows the channel allocation and power assignment for homogeneous network setup. It is clear that with the same channel allocation and slight increase in the power assignment (compared to the naive approach) under the proposed optimization framework results in satisfying all link demands. For example, the naive approach fails to satisfy a rate demand per link of 20 Mb/s while $C_{\text{total}} = 5$ and $P_{\text{total}} = 1.461$. Under the proposed framework, the channel allocation stays the same while P_{total} increased by only 6.63% to consider the effect of 'Bad' ADCs. While C_{total} does

Table 4.5: Channel allocation and power assignment (in Watts) w/ and w/o RF-characteristics awareness for homogeneous networks ($\beta = 1$).

D	w/o RF awareness		w/ RF awareness					
			w/ 'Bad' Mixers		w/ 'Bad' ADCs		w/ 'Bad' Receivers	
	C_{total}	P_{total}	C_{total}	P_{total}	C_{total}	P_{total}	C_{total}	P_{total}
5	5	0.1851	5	0.1891	5	0.1919	5	0.1934
10	5	0.367	5	0.383	5	0.384	5	0.3877
15	5	0.9218	5	0.992	5	0.9774	5	1.0015
20	5	1.461	5	1.5954	5	1.5578	5	1.6189
25	5	1.461	5	1.5954	5	1.5578	5	1.6189
30	10	1.6802	10	1.8374	10	2.0071	10	2.0631
35	10	2.2865	10	2.5967	10	2.8697	10	3.026

not change, the percentage of total power increase becomes 9.2% and 10.81% when considering the effect of 'Bad' mixers and receivers, respectively.

4.7.3 Balancing channel allocation and power assignment

Here, we show the ability of our framework to control the trade-off between C_{total} and P_{total}. We consider the setup of 'bad' receivers for this set of experiments. As shown in Table 4.6, when β increases, the proposed framework favors minimizing the total assigned power at the cost of utilizing larger number of channels.

4.7.4 The case of heterogeneous networks

Here, we consider a general case where only a subset of the receivers has good characteristics while the others are not. Table 4.7 shows the channel allocation and power assignment for heterogeneous networks setup. As expected, the higher the percentage of 'bad' receivers, the larger number of channels and total power are required to satisfy all demands.

Table 4.6: The trade-off between channel allocation and power assignment (in Watts) ('bad' receivers case).

D	$\beta = 1$		$\beta = 10$		$\beta = 100$		$\beta = 1000$	
	C_{total}	P_{total}	C_{total}	P_{total}	C_{total}	P_{total}	C_{total}	P_{total}
5	5	0.1934	5	0.1934	5	0.1934	5	0.1934
10	5	0.3877	5	0.3877	5	0.3877	5	0.3877
15	5	1.0015	5	1.0015	11	0.7637	16	0.7490
20	5	1.6189	8	1.2453	16	1.0718	16	1.0718
25	5	1.6189	8	1.2453	16	1.0718	16	1.0718
30	10	2.063	11	1.9403	16	1.8032	16	1.8032
35	10	3.026	12	2.7241	16	2.5548	16	2.5548

4.7.5 Performance of the proposed PSO-based algorithm

In this section, we discuss the performance of the proposed PSO-inspired framework. The PSO algorithm parameters (w_g and w_l) are both set to 0.1. The weight of the penalty term in the fineness function (ζ) is set to $C + \beta\gamma_{\text{max}}M$. For the small network setup, $\zeta := \zeta_1 = 5\beta + 20$. For the large network setup, $\zeta := \zeta_2 = 10\beta + 50$. Table 4.4 summarizes the PSO-inspired algorithm parameters. For small network, we compare the proposed PSO-based algorithm under different configurations to the optimal solution using two sets of experiments. In the first set, we set the number of particles to 10 and the number of iterations to 10000 (we refer to it as PSO-10-10000). In the second set, the number of particles is set to 50 and the number of iterations to 1000 (we refer to it as PSO-50-1000). Note that the shown results in this section are only a few examples from our experiments which spanned different network instances and gave similar results.

4.7.5.1 Performance for small networks

Here, we compare the solutions obtained by PSO framework to the optimal solution obtained by CPLEX for the small network setup. We use the same network topology in Fig. 4.7 and all link receivers are 'Bad'. As shown in Fig. 4.9, the proposed framework was able to reach solutions which are very close to the optimal ones for different demand values under the two PSO algorithm

Table 4.7: Channel allocation and power assignment (in Watts) for heterogeneous networks ($\beta = 1$).

D	40% 'Bad' Receivers		60% 'Bad' Receivers		80% 'Bad' Receivers		100% 'Bad' Receivers	
	C_{total}	P_{total}	C_{total}	P_{total}	C_{total}	P_{total}	C_{total}	P_{total}
5	5	0.1851	5	0.1859	5	0.1916	5	0.1934
10	5	0.3672	5	0.3701	5	0.385	5	0.3877
15	5	0.9233	5	0.9415	5	0.9806	5	1.0015
20	5	1.4647	5	1.5106	5	1.5657	5	1.6189
25	5	1.4647	5	1.5106	5	1.5657	5	1.6189
30	10	1.7503	10	1.7828	10	1.8861	10	2.063
35	10	2.401	10	2.4475	10	2.628	10	3.026
40	10	3.4031	10	3.5188	11	4.0482	Infeasible	
45	10	3.4031	10	3.5188	11	4.0482	Infeasible	
50	12	3.6962	13	3.9345	Infeasible		Infeasible	
55	17	4.1318	Infeasible		Infeasible		Infeasible	
60	Infeasible		Infeasible		Infeasible		Infeasible	

configurations.

4.7.5.2 Convergence rate

Here, we show the convergence rate of the proposed framework. Each subfigure in Fig. 4.10 shows the behavior of the PSO-inspired algorithm under the two configurations for a specific combination of β and demand value D. For $\beta = 1$, when $D = 15$, the algorithm was able to find the optimal solution in less than 200 iterations under the second PSO configuration (PSO-50-1000) while it needed around 800 iterations to reach the optimal solution under the first configuration (PSO-10-10000). This can be explained as follows. In the first configuration, the number of particles is 10. This resulted in limited number of searching arms for the algorithm to exploit different areas in the solution space within small number of iterations. In the second configuration, the number of particles is 50 which resulted in better exploitation of the solution space and the algorithm could reach the optimal solution within much lower number of iterations. This comes with the price of extra computations in each iteration. As expected, as the demand value increases (e.g. $D = 30$),

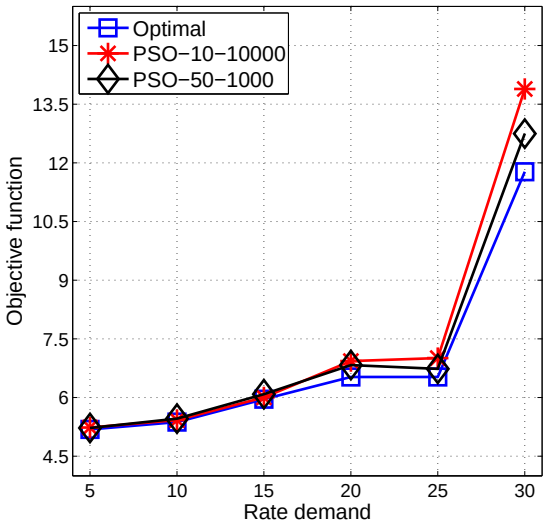
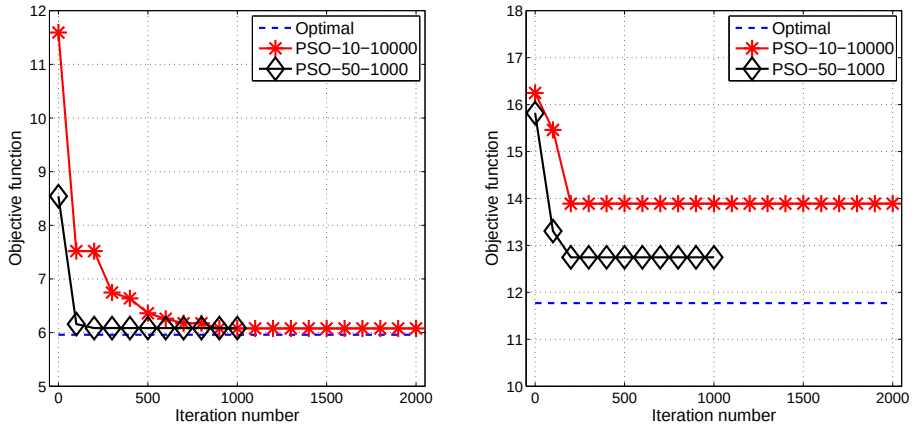


Figure 4.9: Performance of the proposed PSO-inspired framework (5 links, $\beta = 1$).

the algorithm can find near-optimal solutions. In this case, the exact optimal solution was not found within each experiment's iteration because the number of optimal solutions is smaller as the demand values increase and hence, harder to find. For $\beta = 100$, the algorithm could find good solutions within the predetermined maximum number of iterations.

4.7.5.3 Performance for larger networks

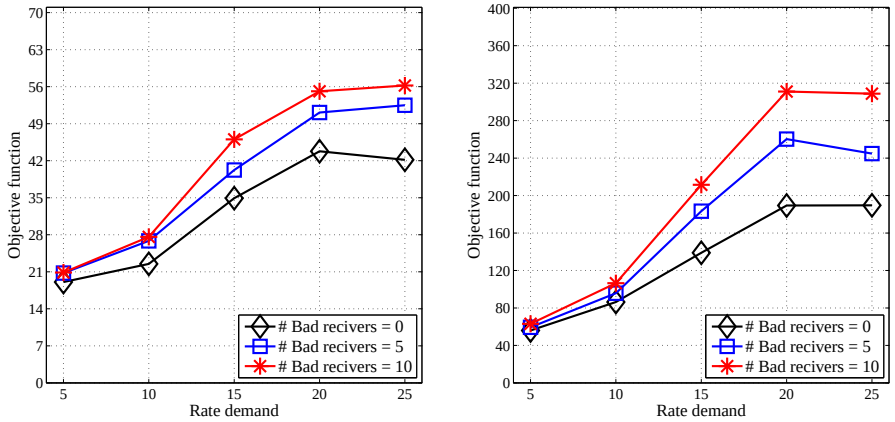
Here, we show the performance of our proposed algorithm for a larger network shown in Fig. 4.8 where the number of links is 10. Note that it is not possible here to compare the obtained results of the PSO algorithm with the optimal solution. This is because a state-of-the-are tool (e.g. CPLEX) fails to provide solutions for such large network size. Moreover, as it was shown in the previous section that using 50 particles gives satisfactory results, we use the same configuration for the experiments here. Fig. 4.11 shows the objective function value for $\beta = 10$ and 1000 under three different receiver configurations. In the first, second and third configurations, the number of bad receivers are 0, 5, and 10, respectively. It is easy to see that as the number of bad receivers increases, the value of objective function increases, as expected. Also, as the demand value increases, the difference in the objective function value between the three configuration increases,



(a) $\beta = 1, D = 15$ (b) $\beta = 1, D = 30$

(c) $\beta = 100, D = 15$ (d) $\beta = 100, D = 30$

Figure 4.10: Convergence of the proposed PSO-inspired framework for the small network setup.



(a) $\beta = 10$ (b) $\beta = 100$

Figure 4.11: Behaviour of PSO-inspired framework for larger networks (10 links).

in general. This can be explained as follows. When the demand value for each link is small (i.e. 5 and 10), the amount of power allocated to each link is relatively small which does not qualify to cause ACI to other links. As a result, almost no power adjustment is needed at any link and all demands can be satisfied. This makes the network under the three configurations has the same behavior. As the demand value increases, more power and channels are allocated to each link. This increases the amount of ACI caused to other links which results in demand dissatisfaction. To obtain a feasible solution, the power (and possibly, the number of channels) for each link should be increased to compensate the ACI effect. This will in turn increase the mutual ACI between links. The process of increasing the power level and/or number of channels at each link will continue until all demands are satisfied. Of course, as the number of bad receivers increases, the amount of ACI also increases which causes this process to continue longer leading to significant increase in the objective function value. In this case, the objective function value becomes noticeably different under the three configurations.

4.8 Conclusions

In this chapter, we proposed a novel receiver-characteristic aware framework in a multi-RAT environment. The framework allows controlling a trade-off between channel allocation and total power emission by all links in the network. The objective was to minimize the total number of utilized channels and the aggregated power emission while meeting all link rate demands. Through extensive simulations, we showed the criticality of the receiver characteristic awareness when designing resource allocation schemes. To tackle the complexity of the overall problem, we developed an efficient algorithm. By decomposing the problem into two subproblems, the algorithm solves it iteratively using a two-phase structure. In Phase I, a particle swarm optimization (PSO)-based algorithm is tailored to obtain good solutions for the channel allocation subproblem. Phase II solves the power allocation subproblem optimally given the channel allocation solution obtained in Phase I. The results demonstrate the superiority of the proposed algorithm in solving the overall problem efficiently and obtaining near-optimal solutions.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

In this dissertation, we covered three research problems which can be summarized as follows:

In Chapter 2, we discussed different approaches to tackle the problem of excessive memory consumption when solving MILP problems. Generic formulations are often not sufficiently attractive from the problem-solving perspective. We demonstrated that generating special cuts through exploiting the structure of the problem offers a better strategy. In most cases, combining different kinds of special cuts outperformed the performance of the formulations that use these cuts individually (or not at all). Moreover, introducing auxiliary binary variables to provide partitioning opportunities based on these cuts, when applicable, significantly enhanced the performance for some instances. Overall, this work demonstrates how the use of proper combinations of model enhancement techniques can help optimize (or further reduce the optimality gap) for challenging instances that were hopelessly unsolvable using traditional formulations.

In Chapter 3, we proposed a novel stochastic spectrum distribution framework for WLANs, which accounts for the APs' demand uncertainty. Static and adaptive stochastic bandwidth allo-

cation approaches were developed. The objective of the static approach was to minimize the total number of channels allocated to the network. The static approach also considered guaranteeing a configurable minimum probability of satisfaction for each AP demand. On the other hand, the adaptive approach provided a trade-off between the number of channels and the level of AP demand satisfaction. Due to the complexity of the adaptive approach, we proposed a PSO-inspired algorithm to solve the ACA problem. Through extensive simulations, we showed the superiority of the proposed stochastic allocation framework compared to the deterministic approach.

In Chapter 4, we proposed a novel receiver-characteristic aware framework in a multi-RAT environment. The framework allows controlling a trade-off between channel allocation and total power emission by all links in the network. The objective was to minimize the total number of utilized channels and the aggregated power emission while meeting all rate demands of the links. Through extensive simulations, we showed the criticality of the receiver characteristic awareness when designing resource allocation schemes. Given the complexity of the overall problem, we develop an efficient algorithm. After decomposing the problem into two subproblems, the algorithm solves it iteratively using a two-phase structure. In Phase I, a particle swarm optimization (PSO)-based algorithm is tailored to obtain good solutions for the channel allocation subproblem. Phase II solves the power allocation subproblem optimally given the channel allocation solution obtained in Phase I. The results demonstrated the superiority of the proposed algorithm in solving the overall problem efficiently and obtaining near-optimal solutions.

5.2 Publications

The following list summarizes the publications resulting from this dissertation:

- Journal articles:
  - **A. Nabil**, M. J. Abdel-Rahman, and A. B. MacKenzie, "A Stochastic Optimization Framework for Channel Bonding in Wireless LANs Under Demand Uncertainty," *in*

preparation.

  - **A. Nabil**, A. V. Padaki,, M. J. Abdel-Rahman, M. ElNainay, A. B. MacKenzie, and J. H. Reed, "On the Optimal Resource Allocation in Multi-RAT Wireless Networks with Receiver Characteristic Awareness," *in preparation*.

- Magazine papers:

  - **A. Nabil**, R. Zhu, Y. T. Hou, W. Lou, and S. F. Midkiff, "Recent Advances in Interference Management for Wireless Networks," IEEE Network, vol. 29, no. 5, pp. 83-89, Sept.-Oct. 2015.

- Conference papers:

  - **A. Nabil**, H. D. Sherali, and M. ElNainay, "Enhancing the Solvability of Network Optimization Problems through Model Augmentations," in Proc. IEEE/IFIP WONS Conference, Feb. 2017.
  - **A. Nabil**, M. J. Abdel-Rahman, and A. B. MacKenzie, "Adaptive Channel Bonding in Wireless LANs Under Demand Uncertainty," in Proc. IEEE PIMRC Conference, Oct. 2017.
  - **A. Nabil**, A. V. Padaki,, M. J. Abdel-Rahman, A. B. MacKenzie, and J. H. Reed, "Receiver Characteristic Aware Optimal Resource Allocation in Multi-RAT Wireless Networks," in Proc. IEEE PIMRC Conference, Oct. 2017.

5.3 Future Research Directions

Potential future research directions for each part of this dissertation are discussed here.

- **Solvability of Network Optimization Problems.** The general multi-hop network cuts can be applied to different problems to study their relative effect. On the other hand, additional approaches following a like philosophy can be explored to obtain better performance. For

example, we could specify specialized branching priorities within CPLEX for binary and integer variables, or introduce partitioning based on different types of disjunctive constraints according to our understanding of the network structure.

- **Channel Bonding in Wireless LANs Under Demand Uncertainty.** In this work, the second stage of the ACA problem was solved using CPLEX without exploiting its decomposability in the implementation. To better exploit it, parallelism in solving the second stage problem should be implemented. State-of-the-art graphics processing units (GPUs) is suitable for this purpose. Besides, we assumed only one source of uncertainly in our problem: AP demands. Channel condition and per-user demand are other potential sources of uncertainly. On the other hand, it was reported that CB results in reducing the transmission range of the AP. Exploiting the capabilities of MIMO can alleviate this effect. It is promising to study our framework under the assumption that some devices in the network are MIMO-capable.

- **Resource Allocation with Receiver Characteristic Awareness.** In this work, a PSO-based algorithm was used to determine the channel allocation for each link in the network while using an optimizer to get the power assignment. An extension to this work is to use PSO to solve the overall problem and compare the performance of the two approaches. On the other hand, receiver parameters were assumed to be deterministic. In reality, they are stochastic where the values cannot be known for certain. Considering the uncertainty in these parameters is a promising direction to explore. This adds a level of difficulty in designing resource allocation algorithms. If the parameters of the receiver characteristics follow specific distributions, a stochastic optimization framework can be employed in order to solve related resource allocation problems.

Finally, the ideas and techniques in Chapter 2 can be extended to enhance the solvability of Chapter 3 and 4 problems.

Bibliography

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?," in *IEEE J. Sel. Areas Commun.,* vol. 32, no. 6, pp. 1065–1082, June 2014.

[2] A. Osseiran, J. F. Monserrat, and P. Marsch, *5G Mobile and Wireless Communications Technology,* Cambridge University Press, New York, NY, 2016.

[3] Y. Li, F. Baccelli, J. G. Andrews, T. D. Novlan and J. C. Zhang, "Modeling and analyzing the coexistence of Wi-Fi and LTE in unlicensed spectrum," in *IEEE Trans. Wireless Commun.,* vol. 15, no. 9, pp. 6310–6326, Sept. 2016.

[4] H. Zhang, X. Chu, W. Guo and S. Wang, "Coexistence of Wi-Fi and heterogeneous small cell networks sharing unlicensed spectrum," in *IEEE Commun. Mag.,* vol. 53, no. 3, pp. 158–164, March 2015.

[5] R. Zhang, M. Wang, L. X. Cai, Z. Zheng, X. Shen and L. L. Xie, "LTE-unlicensed: the future of spectrum aggregation for cellular networks," in *Wireless Commun.,* vol. 22, no. 3, pp. 150–159, June 2015.

[6] A. Mukherjee, J.-F. Cheng, S. Falahati, H. Koorapaty, D. H. Kang, R. Karaki, L. Falconetti and D. Larsson, "Licensed-Assisted Access LTE: coexistence with IEEE 802.11 and the evolution toward 5G," in *IEEE Commun. Mag.,* vol. 54, no. 6, pp. 50–57, June 2016.

[7] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. Sukhavasi, C. Patel and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5G," in *IEEE Commun. Mag.,* vol. 52, no. 2, pp. 82–89, Feb. 2014.

[8] Presidents Council of Advisors on Science and Technology (PCAST) report, Realizing the Full Potential of Government-Held Spectrum to Spur Economic Growth, July 2012.

[9] "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," IEEE Std. 802.11-2016 (Revision of IEEE Std 802.11-2012), Dec. 2016.

[10] X. G. Meng, S. H. Wong, Y. Yuan, and S. Lu, "Characterizing flows in large wireless data networks," in *Proc. ACM MobiCom,* pp. 174–186, Philadelphia, PA, Sept. 2004.

[11] Y. T. Hou, Y. Shi, and H. D. Sherali, *Applied Optimization Methods for Wireless Networks,* Cambridge University Press, Cambridge, UK, 2014.

[12] H. D. Sherali and W. P. Adams, "A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems," *SIAM J. Discrete Math.,* vol. 3, no. 3, pp. 411–430, Aug. 1990.

[13] http://www.ilog.com/products/cplex/

[14] G. L. Nemhauser and L. A. Wolsey, *Integer and Combinatorial Optimization,* John Wiley & Sons, New York, NY, 1999.

[15] Y. T. Hou, Y. Shi, and H. D. Sherali, "Spectrum sharing for multi-hop networking with cognitive radios," *IEEE J. Sel. Areas Commun.,* vol. 26, no. 1, pp. 146–155, Jan. 2008.

[16] M. Johansson and L. Xiao, "Cross-layer optimization of wireless networks using nonlinear column generation," *IEEE Trans. Wireless Commun.,* vol. 5, no. 2, pp. 435–445, Feb. 2006.

[17] M. Pan, C. Zhang, P. Li and Y. Fang, "Joint routing and link scheduling for cognitive radio networks under uncertain spectrum supply," in *Proc. IEEE INFOCOM,* Shanghai, China, pp. 2237–2245, April 2011.

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

[19] P. Gupta and P. R. Kumar, "The capacity of wireless networks." *IEEE Trans. Inf. Theory,* vol. 46, no. 2, pp. 388–404, March 2000.

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

[21] O. Simone, I. Stanojev, S. Savazzi, Y. Bar-Ness, U. Spagnolini, and R. Pickholtz, "Spectrum leasing to cooperating secondary ad hoc networks," *IEEE J. Sel. Areas Commun.,* vol. 26, no. 1, pp. 203–213, Jan. 2008.

[22] X. Yuan, Y. Shi, Y. T. Hou, W. Lou, and S. Kompella, "UPS: A united cooperative paradigm for primary and secondary networks," in *Proc. IEEE MASS,* Hangzhou, Zhejiang, P.R.C., pp. 78–85, Oct. 2013.

[23] http://www.arc.vt.edu/resources/hpc/blueridge.php

[24] A. Nabil, M. J. Abdel-Rahman, and A. B. MacKenzie, "Adaptive channel bonding in wireless LANs under demand uncertainty," to appear in *Proc. IEEE PIMRC,* Montreal, QC, Canada, Oct. 2017.

[25] https://www.ericsson.com/mobility-report

[26] http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[27] "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications," IEEE Std. 802.11ac, Dec. 2013.

[28] T. Moscibroda, R. Chandra, Y. Wu, S. Sengupta, P. Bahl, and Y. Yuan, "Load-aware spectrum distribution in wireless LANs," in *Proc. IEEE ICNP,* pp. 137–146, Orlando, FL, Oct. 2008.

[29] 3GPP TR 25.913 V8.0.0 (200812), "Requirements for evolved UTRA (E-UTRA) and evolved UTRAN (E-UTRAN), Rel. 8, 2008.

[30] P. Kali and S. W. Wallace, *Stochastic programming,* Springer, New York, NY, 1994.

[31] M. J. Abdel-Rahman, M. AbdelRaheem, A. B. MacKenzie, K. Cardoso, and M. Krunz, "On the orchestration of robust virtual LTE-U networks from hybrid half/full-duplex Wi-Fi APs," in *Proc. IEEE WCNC,* pp. 1–6, Doha, Qatar, April 2016.

[32] M. J. Abdel-Rahman, K. Cardoso, A. B. MacKenzie, and L. A. DaSilva, "Dimensioning virtualized wireless access networks from a common pool of resources," in *Proc. IEEE CCNC,* pp. 1049–1054, Las Vegas, NV, Jan. 2016.

[33] N. Y. Soltani, S. J. Kim, and G. B. Giannakis, "Chance-Constrained optimization of OFDMA cognitive radio uplinks," *IEEE Trans. Wireless Commun.,* vol. 12, no. 3, pp. 1098–1107, March 2013.

[34] L. Deek, E. Garcia-Villegas, E. Belding, S. J. Lee, and K. Almeroth, "The impact of channel bonding on 802.11 n network management," in *Proc. ACM CoNext,* pp. 11–22, Tokyo, Japan, Dec. 2011.

[35] S. Joshi, P. Pawelczak, D. Cabric, and J. Villasenor, "When channel bonding is beneficial for opportunistic spectrum access networks," *IEEE Trans. Wireless Commun.,* vol. 11, no. 11, pp. 3942–3956, Nov. 2012.

[36] F. Zarinni and S. R. Das, "Adaptive Spectrum Distribution in WLANs," in *Proc. IEEE GLOBECOM,* pp. 1–6, Miami, FL, Dec. 2010

[37] M. Y. Arslan, K. Pelechrinis, I. Broustis, S. V. Krishnamurthy, S. Addepalli, and K. Papagiannaki, "Auto-configuration of 802.11n WLANs," in *Proc. ACM CoNext,* pp. 27–38, Philadelphia, PA, Nov. 2010.

[38] J. Herzen, R. Merz, and P. Thiran, "Distributed spectrum assignment for home WLANs," in *Proc. IEEE INFOCOM,* pp. 1573–1581, Turin, Italy, April 2013.

[39] K. Hanada, K. Yamamoto, M. Morikura, K. Ishihara, and K. U. D. O. Riichi, "Game-theoretic analysis of multibandwidth channel selection by coordinated APs in WLANs," *IEICE Trans. Commun.,* vol. 96, no. 6, pp. 1277–1287, June 2013.

[40] L. Xu, K. Yamamoto and S. Yoshida, "Performance comparison between channel-bonding and multi-channel CSMA," in *Proc. IEEE WCNC,* pp. 406–410, Kowloon, China, March 2007.

[41] Z. Khan, H. Ahmadi, E. Hossain, M. Coupechoux, L. A. Dasilva and J. J. Lehtomäki, "Carrier aggregation/channel bonding in next generation cellular networks: methods and challenges," *IEEE Network,* vol. 28, no. 6, pp. 34–40, Nov.-Dec. 2014.

[42] L. Deek, E. Garcia-Villegas, E. Belding, S. J. Lee and K. Almeroth, "Intelligent channel bonding in 802.11n WLANs," *IEEE Trans. Mobile Comput.,* vol. 13, no. 6, pp. 1242–1255, June 2014.

[43] B. Bellalta, A. Checco, A. Zocca, and J. Barcelo, "On the interactions between multiple overlapping WLANs using channel bonding," *IEEE Trans. Veh. Technol.,* vol. 65, no. 2, pp. 796–812, Feb. 2016.

[44] S. H. R. Bukhari, M. H. Rehmani, and S. Siraj, "A Survey of Channel Bonding for Wireless Networks and Guidelines of Channel Bonding for Futuristic Cognitive Radio Sensor Networks," *Commun. Surveys Tuts.,* vol. 18, no. 2, pp. 924–948, Secondquarter 2016.

[45] R. Atawia, H. Abou-zeid, H. S. Hassanein, and A. Noureldin, "Robust resource allocation for predictive video streaming under channel uncertainty," in *Proc. IEEE GLOBECOM,* pp. 4683–4688, Austin, TX, Dec. 2014.

[46] M. J. Abdel-Rahman and M. Krunz, "Stochastic guard-band-aware channel assignment with bonding and aggregation for DSA networks," *IEEE Trans. Wireless Commun.,* vol. 14, no. 7, pp. 3888–3898, July 2015.

[47] M. J. Abdel-Rahman, M. AbdelRaheem, and A. B. MacKenzie, "Stochastic resource allocation in opportunistic LTE-A networks with heterogeneous self-interference cancellation capabilities," in *Proc. IEEE DySPAN,* pp. 200–208, Stockholm, Sweden, Sept.-Oct. 2015.

[48] M. J. Abdel-Rahman, E. A. Mazied, A. B. MacKenzie, S. Midkiff, M. R. Rizk, and M. El-Nainay, "On stochastic controller placement in software-defined wireless networks," in *Proc. IEEE WCNC,* pp. 1–6, San Francisco, CA, March 2017.

[49] R. Atawia, H. S. Hassanein, and A. Noureldin, "Energy-efficient predictive video streaming under demand uncertainties," in *Proc. IEEE ICC,* pp. 1–6, Paris, France, May 2017.

[50] Y. C. Chen, J. Kurose, and D. Towsley, "A mixed queueing network model of mobility in a campus wireless network," in *Proc. IEEE INFOCOM,* pp. 2656–2660, Orlando, FL, March 2012.

[51] G. Laporte and F. V. Louveaux, "The integer L-shaped method for stochastic integer programs with complete recourse," *Operations research letters,* vol. 13, no .3, pp. 133–142, April 1993.

[52] G. Angulo, S. Ahmed, and S. S. Dey, "Improving the integer L-shaped method," *INFORMS Journal on Computing,* vol. 28, no. 3, pp. 483–499, May 2016.

[53] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in *Proc. IEEE ICNN,* pp. 1942–1948, Perth, Western Australia, Nov.-Dec. 1995.

[54] R. V. Kulkarni and G. K. Venayagamoorthy, "Particle Swarm Optimization in Wireless-Sensor Networks: A Brief Survey," *IEEE Trans. Syst. Man, Cybern. B, Cybern.,* Part C (Applications and Reviews), vol. 41, no. 2, pp. 262–267, March 2011.

[55] X. Zhuang, H. Cheng, N. Xiong, and L. T. Yang, "Channel Assignment in Multi-Radio Wireless Networks Based on PSO Algorithm," in *Proc. FutureTech,* pp. 1–6, Busan, South Korea, May 2010.

[56] S. N. Ohatkar and D. S. Bormane, "Hybrid channel allocation in cellular network based on genetic algorithm and particle swarm optimisation methods," *IET Communications,* vol. 10, no. 13, pp. 1571–1578, Sept. 2016.

[57] A. Nabil, A. V. Padaki, M. J. Abdel-Rahman, A. B. MacKenzie, and J. H. Reed, "Receiver Characteristic Aware Optimal Resource Allocation in Multi-RAT Wireless Networks," to appear in *Proc. IEEE PIMRC,* Montreal, QC, Canada, Oct. 2017.

[58] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021, White Paper. Available: https://goo.gl/zrFLFn

[59] President's Council of Advisors on Science and Technology (PCAST) report, Realizing the Full Potential of Government-Held Spectrum to Spur Economic Growth, July 2012. Available: `http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast_spectrum_report_final_july_20_2012.pdf`

[60] T. M. Taher, R. B. Bacchus, K. J. Zdunek, and D. A. Roberson, "Long-term Spectral Occupancy Findings in Chicago," in *Proc. IEEE DySPAN,* pp. 100–107, Aachen, Germany, May 2011.

[61] K. Patil, R. Prasad, and K. Skouby, "A Survey of Worldwide Spectrum Occupancy Measurement Campaigns for Cognitive Radio," in *Proc. ICDeCom,* pp. 1–5, Ranchi, India, Feb. 2011.

[62] A. Martian, C. Vladeanu, I. Marcu, and I. Marghescu, "Evaluation of Spectrum Occupancy in an Urban Environment in a Cognitive Radio Context," *International Journal on Advances in Telecommunications*, vol. 3, no. 3, pp. 172–181, 2010.

[63] J. M. Peha, "Sharing Spectrum Through Spectrum Policy Reform and Cognitive Radio," in *Proc. IEEE,* vol. 97, no. 4, pp. 708–719, April 2009.

[64] S. M. Dudley *et al.*, "Practical Issues for Spectrum Management with Cognitive Radios," in *Proc. IEEE,* vol. 102, no. 3, pp. 242–264, 2014.

[65] T. M. Taher, R. B. Bacchus, K. J. Zdunek, and D. A. Roberson, "Spectrum Utilization Study in Support of Dynamic Spectrum Access for Public Safety," in *Proc. IEEE DySPAN,* pp. 1–11, Singapore, April 2010.

[66] FCC Report and Order and Second Notice of Proposed Rule Making, Amendment of the Commission's Rules with Regard to Commercial Operations in the 3550-3650 MHz Band, FCC 15-47, GN Docket No. 12-354, April 21, 2015.

[67] FCC News Release, "FCC Makes 150 MHz of Contiguous Spectrum Available for Mobile Broadband and Other Uses Through Innovative Sharing Policies," April 17, 2015. Available: https://apps.fcc.gov/edocs_public/attachmatch/DOC-333083A1.pdf

[68] FCC Report and Order, Amendment of the Commissions Rules with Regard to Commercial Operations in the 1695-1710 MHz, 1755-1780 MHz, and 2155-2180 MHz Bands, FCC 14-31, GN Docket No. 13-185, March 2014.

[69] C. Ghosh, S. Roy, and D. Cavalcanti, "Coexistence Challenges for Heterogeneous Cognitive Wireless Networks in TV White Spaces," *Wireless Commun.,* vol. 18, no. 4, pp. 22–31, Aug. 2011.

[70] N. Golmie, *Coexistence in Wireless Networks,* Cambridge University Press, New York, NY, 2006.

[71] J. H. Reed *et al.*, "On the Co-existence of TD-LTE and Radar over 3.5 GHz Band: An Experimental Study," *Wireless Commun.,* vol. 5, no. 4, pp. 368–371, Aug. 2016.

[72] D. Schiendler, "GPS-Interference Controversy Comes to a Boil," *IEEE Spectrum,* vol. 49, no. 2, pp. 13-14, Feb. 2012.

[73] FCC Technological Advisory Council, "Interference Limits Policy: The use of harm claim thresholds to improve the interference tolerance of wireless systems," White Paper Version 1.0, Feb. 2013.

[74] European Communications Office (ECO) Report 02, "The Impact of Receiver Parameters on Spectrum Management Regulations: A Pilot Study," June 2010.

[75] European Communications Committee (ECC) Report 127, "The Impact of Receiver Standards on Spectrum Management," Oct. 2008.

[76] L. Davies and P. Winter, "Study of Current and Future Receiver Performance," Final Report of OFCOM, Jan. 2010.

[77] Wireless Innovation Forum, "Receiver Performance Technology," Document WINNF-16-P-0020, Oct. 2016. Available:http://www.wirelessinnovation.org/assets/work_products/Reports/winnf-16-p-0020-v1.0.0%20receiver%20performance%20technology.pdf

[78] A. V. Padaki, V. Marojevic, and J. H. Reed, "Role of Receiver Performance Data in Efficient Spectrum Utilization," in *Proc. IEEE DySPAN,* pp. 366–369, McLean, VA, April 2014.

[79] E. Y. Imana, T. Yang, and J. H. Reed, "Relaxing Pre-selector Filter Selectivity Requirements Using Cognitive RF Front-end Control," *Analog Integr. Circuits Signal Process.,* vol. 78, no. 3, pp. 705–716, March 2014.

[80] A. V. Padaki, R. Tandon, and J. H. Reed, "Receiver Non-Linearity Aware Resource Allocation for Dynamic Spectrum Access Systems," in *Proc. IEEE MILCOM,* pp. 1393–1398, Baltimore, MD, Oct. 2014.

[81] A. V. Padaki, R. Tandon, and J. H. Reed, "Efficient Spectrum Sharing with RF Diversity: Adapting to Nonlinearity of Front Ends," in *Proc. IEEE GLOBECOM,* pp. 1–6, San Diego, CA, Dec. 2015.

[82] A. V. Padaki and J. H. Reed, "Impact of Intermodulation Distortion on Spectrum Preclusion for DSA: A New Figure of Merit," in *Proc. IEEE DySPAN,* pp. 358–361, McLean, VA, April 2014.

[83] A. V. Padaki, R. Tandon, and J. H. Reed, "On Scalability and Interference Avoidance in Nonlinear Adjacent Channel Interference Networks," in *Proc. IEEE ICC,* pp. 1–6, Paris, France, May 2017.

[84] D. H. Mahrof, E. A. M. Klumperink, J. C. Haartsen, and B. Nauta, "On the Effect of Spectral Location of Interferers on Linearity Requirements for Wideband Cognitive Radio Receivers," in *Proc. DySPAN,* pp. 1–9, Singapore, April 2010.

[85] E. Y. Imana, *Cognitive RF Front-end Control*, Ph.D. Dissertation, Department of Electrical and Computer Engineering, Virginia Tech, Dec. 2014.

[86] E. Y. Imana, T. Yang, and J. H. Reed, "Suppressing the Effects of Aliasing and IQ Imbalance on Multi-band Spectrum Sensing," *in IEEE Trans. Veh. Technol.,* vol. 66, no. 2, pp. 1074–1086, Feb. 2017.

[87] B. Razavi, *RF Microelectronics,* Prentice Hall Press Upper Saddle River, NJ, 2012.

[88] L. Smaini, *RF Analog Impairments Modeling for Communication Systems Simulation: Application to OFDM-based Transceivers,* J. Wiley & Sons, Hoboken, NJ, 2012.

[89] K. McClaning, *Wireless Receiver Design for Digital Communications,* SciTech Publishing, Raleigh, NC, 2012.

[90] M. Haenggi, *Stochastic Geometry for Wireless Networks,* Cambridge University Press, New York, NY, 2012.