Exploring Performance Limits of Wireless Networks with Advanced Communication Technologies

Xiaoqi Qin

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

Y. Thomas Hou, Chair
Wenjing Lou
Scott F. Midkiff
Jeffrey H. Reed
Yi Shi
Yaling Yang

August 26, 2016
Blacksburg, Virginia

Keywords: Wireless network, MIMO, interference cancellation, full duplex, mmWave communication, modeling and optimization, algorithm design

Exploring Performance Limits of Wireless Networks with Advanced Communication Technologies

Xiaoqi Qin

ABSTRACT

Over the past decade, wireless data communication has experienced a phenomenal growth, which is driven by the popularity of wireless devices and the growing number of bandwidth-hungry applications. During the same period, various advanced communication technologies have emerged to improve network throughput. Some examples include multi-input multi-output (MIMO), full duplex, cognitive radio, mmWave, among others. An important research direction is to understand the impacts of these new technologies on network throughput performance. Such investigation is critical not only for theoretical understanding, but also can be used as a guideline to design algorithms and network protocols in the field.

The goal of this dissertation is to understand the impact of some advanced technologies on network throughput performance. More specifically, we investigate the following three technologies: MIMO, full duplex, and mmWave communication. For each technology, we explore the performance envelope of wireless networks by studying a throughput maximization problem. The main contributions of this dissertation can be summarized as follows:

- **Joint Flow Routing and DoF Allocation in Multi-hop MIMO Networks.** MIMO is a powerful physical (PHY) layer technology that can offer significant performance gain in capacity. Despite extensive work on MIMO at PHY layer for single-hop communications, there remains little knowledge on how to exploit MIMO capabilities in multi-hop networks. In this work, we employ a degree-of-freedom (DoF) based MIMO model for spatial multiplexing (SM) and interference cancellation (IC) and study a throughput maximization problem in a multi-hop network. The problem formulation involves joint consideration of flow routing and DoF allocation for SM/IC and falls in the form of a mixed-integer linear program (MILP). Our main contribution is an efficient polynomial time algorithm that offers a competitive solution.

- **Impact of Full Duplex Scheduling on End-to-end Throughput in Multi-hop Wireless**

Networks. Full duplex communication aims to double the throughput of a wireless node by allowing simultaneous transmission and reception. Recent efforts on full duplex has been focused on transceiver design to allow bi-directional communication. In this work, we investigate the potential benefit of full duplex for end-to-end session throughput in a multi-hop wireless network. Through rigorous mathematical modeling, we formulate a throughput maximization problem. We show that in many cases, the end-to-end session throughput in a full duplex network can exceed $2\times$ of that in a half duplex network. This finding offers new insights on the potential benefit of full duplex at the network level.

- **On AP Assignment and Transmission Scheduling for Multi-AP 60 GHz WLAN.** Millimeter wave (mmWave) communication in 60 GHz band is considered a promising technology to provide multi-gigabit data rate for Wi-Fi-based WLAN. To address the potential blockage problem for 60 GHz signals, multiple APs are proposed for such a WLAN. In this work, we study the important problem of AP assignment and transmission scheduling for a multi-AP WLAN in the 60 GHz regime. We propose several AP assignment schemes with different complexity to maximize user throughput. Among the proposed schemes, we advocate to use one-shot AP assignment-based scheduling due to its simplicity and competitive performance. We also propose an online algorithm to implement the one-shot AP assignment scheme in the field. We show that the proposed online algorithm is competitive compared with the optimal offline algorithm.

Exploring Performance Limits of Wireless Networks with Advanced Communication Technologies

Xiaoqi Qin

GENERAL AUDIENCE ABSTRACT

As everyone knows, we are now living in a connected world, where network access is available anytime and anywhere. According to Cisco's report [97], global Internet traffic is expected to reach 2.3 zettabytes per year by 2020, and wireless data traffic will account for 65% of the total Internet traffic. There are three primary contributors for the explosive growth of wireless data demand: the rising number of wireless devices, the increasing number of new applications, and the ever-growing amount of video traffic. Each year, all kinds of smart devices with increased intelligence are introduced in market. The number of wireless devices is predicted to reach 11.6 billion by 2020 [97]. The smart devices enable people to enjoy mobile applications for entertainment, such as social networking, video streaming, and gaming. Such bandwidth hungry applications have changed the wireless data consumption pattern. According to Ericssons report [98], video traffic dominates the mobile data consumption for all kinds of mobile devices. Moreover, the amount of video traffic is still growing more than 50 % annually.

To meet the ever-growing traffic demand, innovative technologies have been developed to expand the capacity of wireless networks. Some examples include multi-input multi-output (MIMO), full duplex, cognitive radio, mmWave, ultra-wideband, among others. In this dissertation, we aim to investigate the impact of such advanced technologies on network throughput performance. Such theoretical study is critical since it can be used as a guidline to design real-world network protocols.

Dedication

To my parents

Acknowledgments

First and foremost, I feel tremendously lucky to have Prof. Thomas Hou as my advisor. His ability to point out a promising research direction and to map out a productive agenda is incredible. He has been extremely helpful, encouraging, patient, and supportive throughout all my years at Virginia Tech. His pursuit of true scholarship inspired me to challenge myself. I can't thank him enough for spending significant amount of time on providing invaluable suggestions on both the technical part and writing of all my works in this dissertation. I will always be indebted to him for all that he has done for me.

I would also like to acknowledge the helpful feedback from the members of my dissertation committee. I thank Professors Wenjing Lou, Scott Midkiff, Jeffrey Reed, Yi Shi, and Yaling Yang for their time and interest in my work. Their comments and questions on my work have helped me to improve the quality of my dissertation.

I would like to thank my current and former colleagues in the research group: Yi Shi, Huacheng Zeng, Xu Yuan, and Biran Jalaian. Getting a doctorate is a long journey and it is these awesome people who encouraged me to keep going. I enjoyed our time together discussing research problems, sharing life experience and good food. When home is thousands of miles away, these people became my brothers and provided me with good company.

There aren't enough words for me to express how much I owe to my family. My parents have been providing me with all the support I needed and all the love I could ever imagine. From my mother's encouragement to my father's guidance, I never felt far away from them. They are my role models and most trustworthy friends. I cannot imagine what I would be today were it not for my family. Thanks you so much for your belief in me.

Funding Acknowledgments

I gratefully acknowledge the funding sources that made my Ph.D. work possible. This dissertation was supported in part by NSF grants CNS-1443889, CNS-1343222, CNS-1064953, ECCS-1102013 and ONR grant N00014-15-1-2926. I also acknowledge Advanced Research Computing (ARC) at Virginia Tech for providing me with the computing resources on powerful supercomputer BlueRidge and HokieOne.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and Goals

Wireless networks have become an essential element for most people's work, life and entertainment. Over the past fifteen years, wireless communication has seen a dramatic increase in data demands. In 2009, approximately 191 billion megabytes (MB) data traveled across U.S. wireless networks and five years later, the data traffic is increased to more than 4 trillion MB [96]. The rising tide of wireless data traffic is driven by the explosive growth in the number of service subscribers, who use high-speed Internet to enjoy bandwidth-hungry applications such as high-definition content streaming, video gaming, real-time data backup, etc. Looking ahead, the data consumption isn't expected to flatten out anytime soon. Attributed to innovation of mobile devices and extraordinary growth of multi-media applications, wireless data traffic is expected to increase more than eightfold by 2018.

The explosive growth of wireless data demand puts too much pressure on the existing network infrastructures. To alleviate this pressure, a number of innovative wireless technologies have been developed, including multiple-input and multiple-output (MIMO), full duplex, cognitive radio, mmWave communication, among others. MIMO is a powerful physical layer technology that employs multiple antennas to increase the capacity of a radio link without the need of additional spectrum. Full duplex breaks the long-held limitation in a wireless transceiver operating in half

duplex mode (i.e., either transmit or receive but not both at the same time). Instead, full duplex enables a transceiver to transmit and receive simultaneously on the same channel and is expected to double the channel capacity. Cognitive radio promises unprecedented flexibility in radio communications under which the transmitted and received waveforms can be programmed and configured dynamically by software. It is regarded as the enabling technology for dynamic spectrum access. Millimeter-wave communication is a promising technology to realize multi-gigabit Wi-Fi due to its massive bandwidth. It is enabled by recent advances in RF design and signal processing in the mmWave spectrum. With the emergence of these new technologies, an important question to ask is how these PHY layer technologies would impact throughput.

A fundamental understanding of these PHY layer technologies in the networking environment is much needed by the research community. But it is not trival. For example, MIMO is expected to increase the capacity between a pair of nodes linearly with the number of antennas. However, in a network where each node is equipped with M antennas, the network capacity may not necessarily be increased by M-fold. The achievable network throughput depends on algorithms at multiple layers, such as scheduling algorithm at link layer and routing algorithm at network layer. The benefit of MIMO at PHY layer could be diminished if it is not well coordinated with upper layer algorithms. Therefore, the impact of MIMO on user throughput in a network environment must take into considerations of link layer scheduling algorithm, network layer routing decisions, among others.

The goal of this dissertation is to study the impact of some advanced technologies on network throughput performance via exploring a set of network optimization problems. More specifically, we investigate the following three technologies: MIMO, full-duplex, and mmWave communication.

1.2 Summary of Contributions

This dissertation studies the impacts of three advanced technologies (MIMO, full-duplex, mmWave communication) on network throughput performance. The main contributions of each chapter are

summarized as follows:

- In Chapter 2, we study a throughput maximization problem in a multi-hop MIMO network. Since the mathematical characterization of MIMO's PHY layer signal processing involves complex matrix operations, degree-of-freedom (DoF) based models, which only require simple numeric operations on the number of DoFs, have been widely used to study MIMO network performance. In this work, we employ a novel DoF model that captures two key characteristics of MIMO: spatial multiplexing (SM) and interference cancellation (IC). We formulate the throughput maximization problem by jointly considering flow routing and DoF allocation for SM/IC. The problem formulation falls in the form of a mixed-integer linear program (MILP), which is computationally intractable (i.e., beyond the capability of any existing optimization solvers to find the optimal solution in a reasonable amount of time). Our main contribution is an efficient polynomial time algorithm that offers a competitive solution to the MILP through solving a series of linear programs. The algorithm employs a sequential fixing framework to obtain an initial feasible solution, and then improves the solution by exploiting the impact of node ordering on DoF consumption for IC at each node and route diversity in the network. Through simulation results, we show that the solutions obtained by our proposed algorithm are feasible and competitive compared with the optimal solutions obtained by CPLEX.

- In Chapter 3, we investigate the impact of full duplex scheduling on the throughput performance in a multi-hop network. Although the benefits of full duplex have been studied for single-hop wireless communications, its potential on throughput performance in a multi-hop network remains unclear. In this work, we study the impact of full duplex on end-to-end throughput in a multi-hop network. We formulate a throughput maximization problem for both full duplex and half duplex, and compare their performance. The resulting formulations fall in the form of a mixed-integer nonlinear program (MINLP), which is computationally intractable. We employ *Reformulation-Linearization Technique* and *piece-wise linear approximization* techniques to linearize the nonlinear constraints in the formulation. Through simulation results, we show that in many cases, the end-to-end session throughput in a full

duplex network can exceed $2\times$ of that in a half duplex network. Our findings demonstrate the much larger design space under full duplex scheduling in the throughput maximization problem.

- In Chapter 4, we study the AP assignment and transmission scheduling problem in a multi-AP 60 GHz WLAN. To address the potential blockage problem for 60 GHz signals, multiple APs are proposed for such WLAN. We propose several AP assignment schemes with different complexity, namely per-time slot AP assignment, one-shot AP assignment, and strongest-signal AP assignment. Under each scheme, we consider downlink communication and study how to maximize user throughput with joint consideration of AP assignment and transmission scheduling. Interference avoidance is considered as a constraint in all formulations by allowing only non-interfering links to be active simultaneously. Through simulation results, we show that both per-time slot AP assignment and one-shot AP assignment based schemes significantly outperform the strongest-signal AP assignment scheme, while the difference between the two is marginal. We advocate to use one-shot AP assignment based scheduling due to its simplicity. For practical implementation, we propose an online one-shot algorithm to address the AP re-assignment issue under real-time traffic with user arrivals/departures. The online algorithm optimizes the AP assignment for new users and time slot allocation for all the users without altering the AP assignment for existing users in the network. Through performance evaluation, we show that the proposed online algorithm is competitive compared with the optimal offline algorithm.

Chapter 2

Joint Flow Routing and DoF Allocation in Multi-hop MIMO Networks

2.1 Introduction

It is well known that MIMO technology delivers significant performance gain in terms of channel capacity and reliability [12, 32, 60, 82]. To date, MIMO has become an essential element of wireless communications including wireless LAN (802.11n) [101], WiMAX (802.16e) [102] and LTE (3GPP) [103]. Despite of rapid advances of physical layer techniques for single-hop communications [3, 14–16, 18, 22, 23], our knowledge on how to exploit the advantages of MIMO in multi-hop networks has been relatively limited. The main difficulty here is that mathematical characterization of MIMO's physical layer signal processing involves complex matrix operations (due to the use of multiple antennas). Such matrix manipulations pose a technical barrier in the design and analysis of network algorithms and protocols. Therefore, a tractable and accurate MIMO model is needed for multi-hop MIMO network research.

Realizing the difficulties, researchers in the networking community extended the concept of degree-of-freedom (DoF), which was originally defined by the information theory (IT) community to represent the multiplexing gain of a MIMO channel [40, 92, 93], to characterize the spatial freedom provided by multiple antennas at a MIMO node. Since then, various DoF-based models

have been developed to analyze the spatial multiplexing (SM) and interference cancellation (IC) capabilities of MIMO [9, 13, 36, 58, 77]. Instead of complex matrix operations, these DoF-based models only require simple numeric computations on the number of DoFs to identify a feasible DoF region. The main idea of these DoF-based models is as follows: (i) The number of available DoFs at a node is equal to the number of its antennas. (ii) A node consumes DoFs for SM. Specifically, a transmit node consumes DoFs to transmit its data streams while a receive node consumes DoFs to receive its desired data streams. (iii) A node consumes DoFs for IC. Specifically, a transmit node may cancel its interference to its neighboring unintended receive nodes by consuming its DoFs; likewise, a receive node may cancel the interference from its neighboring unintended transmit nodes by consuming its DoFs. (iv) A node can use some or all of its DoFs for SM and IC, as long as the total number of DoFs consumed for SM and IC does not exceed its available DoFs.

Based on different IC schemes, the DoF-based models can be classified into two major types: *conservative models* and *optimistic models*. The conservative models ensure feasibility by employing different kinds of predefined IC rules, such as IC at both transmit and receive nodes, restrictions on IC at receiver nodes, among others. As a result, a conservative model [9, 58, 77] loses some feasible solutions. The optimistic models, on the other hand, tend to incorrectly enlarge the feasible solution space due to the lack of a systematic IC scheme. As a result, an optimistic model [13, 36] may offer infeasible solutions. To address the problems associated with the conservative and optimistic models, a new DoF-based MIMO model was developed by Shi *et al.* [71, 90]. The novelty of this model is a disciplined IC scheme based on an ordered node list. Under this model, each node in the network is associated with an order and it is only responsible for canceling interference to/from those nodes that are before itself in the ordered list. That is, a transmit node only needs to cancel its interference to neighboring receive nodes ahead itself in the ordered node list, instead of all its neighboring receive nodes in the network. Similarly, a receive node only needs to cancel interference from neighboring transmit nodes that are before itself in the ordered node list, instead of all its neighboring transmit nodes in the network. By employing the novel "node ordering" concept, IC at each node can be performed in a systematic and disciplined manner. As a result, the potential waste of DoF resources is avoided by eliminating possible duplication in IC

(in conservative models). Further, DoF allocation solution is guaranteed to be feasible in contrast to the optimistic models [90].

With the new DoF model in [71], one can study MIMO multi-hop networks without the limitations associated with the conservative and optimistic DoF models. In this chapter, we explore how to apply this model to study a throughput maximization problem in a multi-hop MIMO network. Note that the scope of this chapter and [71] is fundamentally different. In [71], the authors proposed a new interference cancellation (IC) scheme that can be used for DoF allocation in a multi-hop MIMO network. Although they presented a case study involving a throughput maximization problem, the problem is much simpler than the one that we study in this chapter. For example, routing for each session was assumed to be given *a priori* in [71] rather than being considered as part of the optimization problem in this chapter. Further, there was no discussion on how to solve such an optimization problem. It merely relied on a commerical solver (CPLEX) to demonstrate results. This was OK as the main objective of [71] was to introduce a new MIMO IC scheme rather than pursuing an efficient solution procedure for a throughput maximization problem. In contrast, the objective of this chapter is to pursue an efficient solution procedure for a throughput maximization problem – a problem in a more complex (and general) form than that in [71]. Specifically, in this chapter, we consider a set of unicast sessions in a MIMO network. The objective is to maximize the minimum achievable session rate (in DoFs). For simplicity, we assume that fixed modulation and coding scheme (MCS) is used for each data stream and each data stream corresponds to one unit data rate. Through joint formulation of flow routing and DoF allocation, we obtain a mixed-integer linear program (MILP), which is NP-hard in general. Although a commercial solver (CPLEX) may solve our problem for small-sized networks, a more efficient (and polynomial time) algorithm is needed to handle networks of larger sizes.

The main contribution of this chapter is an efficient algorithm for the joint flow routing and DoF allocation problem. The proposed solution is an iterative greedy algorithm that solves a series (but limited number) of LPs, each of which has polynomial time complexity. The algorithm consists of two stages. In the first stage, the algorithm employs a *sequential fixing* (SF) technique [39, Chapter 10] to obtain an initial feasible solution. The SF technique is itself based on a series of

LPs. In the second stage, the algorithm improves the initial feasible solution. By identifying a "bottleneck" link, the algorithm attempts to increase DoF allocation (for SM) on the bottleneck link by altering the ordering of the node list. This idea exploits the unique property in the DoF model where a node's DoF consumption for IC depends on its position in the node-ordering list. When altering node-ordering can no longer improve the solution, our algorithm tries to make improvement by altering routes. It tries to find a parallel route to bypass the bottleneck link. Throughout the iterations, our algorithm ensures that DoF constraints for SM and IC at each node are satisfied. Simulation results show that our proposed algorithm can offer a competitive solution to the MILP. Further, the solution by our proposed algorithm is guaranteed to be feasible.

The reminder of this chapter is organized as follows. In Section 3.3, we develop a problem formulation for the throughput maximization problem by joint consideration of flow routing and DoF allocation. In Section 2.3 and Section 2.4, we present our algorithm to solve the throughput maximization problem in two stages. Section 2.5 gives a complexity analysis of the proposed algorithm. Section 2.6 presents simulation results. Section 4.8 concludes this chapter.

2.2 Modeling and Formulation

In this section, we study a throughput maximization problem for multi-hop MIMO networks. For a set of sessions in the network, our objective is to maximize the achievable session rate (in DoFs) among the sessions by optimizing variables in flow routing (network layer) and DoF allocation (link layer).

2.2.1 Mathematical Modeling

Table 4.1 lists notation in this chapter. We consider a multi-hop MIMO network with a set of $\mathcal{N}$ nodes, with $N = |\mathcal{N}|$ being the number of nodes. Each MIMO node i has A_i antennas. Denote $\mathcal{F}$ as a set of sessions in the network. For each session $f \in \mathcal{F}$, denote $s(f)$ and $d(f)$ as its source and destination nodes, respectively. Denote $r(f)$ as the achieved throughput (in DoFs) of session $f \in \mathcal{F}$. We assume that time slot based scheduling is used at the link layer. Denote T as the

Table 2.1: Notation in Chapter 2

Symbol	Definition
$\mathcal{N}$	The set of nodes in the network
$\mathcal{F}$	The set of sessions in the network
T	The number of time slots in a frame
$\mathcal{T}_i$	The set of nodes within the transmission range of node i
$\mathcal{I}_i$	The set of nodes within the interference range of node i
A_i	Number of antennas at node i
$x_i[t]$	=1 if node $i \in \mathcal{N}$ is a transmitter in time slot t, and is 0 otherwise
$y_i[t]$	=1 if node $i \in \mathcal{N}$ is a receiver in time slot t,
	and is 0 otherwise
$z_{ij}[t]$	The number of data streams transmitted from node i
	to node j in time slot t, $i, j \in \mathcal{N}$
$\pi[t]$	An ordered node list for IC in time slot t
$\pi_i[t]$	The position of node $i \in \mathcal{N}$ in $\pi[t]$
$\theta_{ji}[t]$	Binary indicator showing the relationship between nodes i
	and j in $\pi[t]$, $i, j \in \mathcal{N}$
$r(f)$	The data rate of session $f \in \mathcal{F}$
$r_{ij}(f)$	The data rate attributed to session $f \in \mathcal{F}$ on link (i, j) in time slot t
$r_{\min}$	The minimum data rate among all sessions

number of time slots in a frame.

Half-duplex Constraint. We assume that wireless transceivers are half-duplex [1]. In a time slot based system, a half-duplex transceiver operates in one of three modes in a time slot: transmit, receive, or idle. To model this behavior in a time slot, we define two binary variables $x_i[t]$ and $y_i[t]$ to indicate whether node i is a transmit node or a receive node in time slot t, respectively. That is, $x_i[t] = 1$ if node i is a transmit node in time slot t and 0 otherwise (either idle or receive); $y_i[t] = 1$ if node i is a receive node in time slot t and 0 otherwise (either idle or transmit); Then the half-duplex constraint can be modeled as

$$x_i[t] + y_i[t] \leq 1 \ , \quad (i \in \mathcal{N}, 1 \leq t \leq T) \ . \tag{2.1}$$

We assume the total number of DoFs at a node i is equal to its number of antenna elements A_i. Denote $z_{ij}(t)$ as the number of data streams from node i to node j in time slot t. If node i is not an active transmitter in time slot t, then no data stream is transmitted from this node, i.e., $\sum_{j \in \mathcal{T}_i} z_{ij}[t] = 0$ if $x_i[t] = 0$, where $\mathcal{T}_i$ is the set of nodes within the transmission range of node i. Otherwise, the total number of DoFs used for transmission cannot exceed the total number of antennas A_i at this node, i.e., $1 \leq \sum_{j \in \mathcal{T}_i} z_{ij}[t] \leq A_i$ if $x_i[t] = 1$. These two cases can be formulated as

$$x_i[t] \leq \sum_{j \in \mathcal{T}_i} z_{ij}[t] \leq A_i \cdot x_i[t] \ , \quad (i \in \mathcal{N}, 1 \leq t \leq T) \ . \tag{2.2}$$

Similarly, considering whether or not node i is a receive node in time slot t, we have

$$y_i[t] \leq \sum_{j \in \mathcal{T}_i} z_{ji}[t] \leq A_i \cdot y_i[t] \ , \quad (i \in \mathcal{N}, 1 \leq t \leq T) \ , \tag{2.3}$$

where we assume node j and node i have the same transmission range, i.e., if $i \in \mathcal{T}_j$ then $j \in \mathcal{T}_i$, and vice versa.

Ordering and DoF Constraints. The "node ordering" concept was proposed in [71] to ensure feasibility at the physical (PHY) layer while avoiding duplication in canceling the same interfer-

[1] Although full-duplex MIMO node has been demonstrated in recent research literatures [11, 21], its near-term availability remains unclear. Should it become widely available in the future, only some minor extensions are needed in our formulation and proposed algorithm.

ence by both the transmitter and receiver in a MIMO network. An optimal ordering can be found by incorporating ordering variables in the problem formulation.

Denote $\pi[t]$ as a list of length N, with each element containing a node in the network. The position of the node in the list defines the "order" of that node. Denote $\pi_i[t]$ as the order (position) of node i in $\pi[t]$. For example, if $\pi_i[t] = 5$, then it means that node i is in the fifth element in the ordered list. Therefore, we have

$$1 \leq \pi_i[t] \leq N \ , \quad (i \in \mathcal{N}, 1 \leq t \leq T) \ . \tag{2.4}$$

We use a binary variable $\theta_{ji}[t]$ to indicate the relative position between two nodes i and j in $\pi[t]$. i.e., $\theta_{ji}[t] = 1$ if node j is before node i in $\pi[t]$ and 0 otherwise. Based on the definition of $\theta_{ji}[t]$, it can be easily verified that the following relationships hold between $\pi_i[t]$ and $\pi_j[t]$:

$$\pi_i[t] - N \cdot \theta_{ji}[t] + 1 \leq \pi_j[t] \leq \pi_i[t] - N \cdot \theta_{ji}[t] + N - 1 \ , (i, j \in \mathcal{N}, i \neq j, 1 \leq t \leq T) \ . \tag{2.5}$$

A node can use its DoFs for either SM or IC, as long as the number of consumed DoFs does not exceed its total available DoFs. Depending on whether the node is a transmit or receive node in time slot t, it has different IC behavior as follows:

- If the node is a transmit node, in addition to SM, the node should use its DoFs to cancel its interference to all the unintended neighboring receive nodes (within its interference range) that are before itself in the ordered node list $\pi[t]$. The number of DoFs consumed by this transmit node for IC is equal to the sum of intended data streams received by those unintended receive nodes.

- If the node is a receive node, in addition to SM, the node should use its DoFs to cancel interference from all unintended neighboring transmit nodes (whose interference range covers this receive node) that are before itself in the ordered node list $\pi[t]$. The number of DoFs consumed by this node for IC is equal to the sum of data streams transmitted by those unintended transmit nodes.

Note that in either case (transmitter or receiver), the node only needs to consider the nodes before itself in $\pi[t]$ for IC. Interference to/from nodes that are after it in the ordered node list $\pi[t]$ will be

taken care of by those nodes later when we go through the list. This is the key in allocating DoF resources for IC. Denote $\mathcal{I}_i$ as the set of nodes within the interference range of node i. We now model both cases mathematically as follows:

$$\text{If } x_i[t]=1, \text{ then } \quad \sum_{j\in\mathcal{T}_i} z_{ij}[t] + \sum_{j\in\mathcal{I}_i}(\theta_{ji}[t]\sum_{k\in\mathcal{T}_j}^{k\neq i} z_{kj}[t]) \leq A_i \ , (i\in\mathcal{N}, 1\leq t\leq T) \ , \tag{2.6}$$

where on the left side of the inequality, the first and second terms represent the number of DoFs consumed by node i for SM and IC, respectively.

$$\text{If } y_i[t]=1, \text{ then } \quad \sum_{j\in\mathcal{T}_i} z_{ji}[t] + \sum_{j\in\mathcal{I}_i}(\theta_{ji}[t]\sum_{k\in\mathcal{T}_j}^{k\neq i} z_{jk}[t]) \leq A_i \ , (i\in\mathcal{N}, 1\leq t\leq T) \ , \tag{2.7}$$

where on the left side of the inequality, the first and second terms represent the number of DoFs consumed by node i for SM and IC, respectively.

Flow Balance Constraints. For flexibility and better load balancing, we allow flow splitting in the network. That is, the flow of a session (with granularity of one DoF) may split and merge inside the network in whatever manner as long as it can help to achieve a higher throughput. Denote $r_{ij}(f)$ as the data rate (in DoFs) on link (i,j) that is attributed to session $f\in\mathcal{F}$, where $i\in\mathcal{N}$ and $j\in\mathcal{T}_i$. Then we have the following flow balance constraints:

- If node i is the source node of session f (i.e., $i=s(f)$), then

$$\sum_{j\in\mathcal{T}_i} r_{ij}(f) = r(f) \ , \quad (f\in\mathcal{F}) \ . \tag{2.8}$$

- If node i is an intermediate relay node for session f (i.e., $i\neq s(f), i\neq d(f)$), then

$$\sum_{j\in\mathcal{T}_i}^{j\neq s(f)} r_{ij}(f) = \sum_{k\in\mathcal{T}_i}^{k\neq d(f)} r_{ki}(f) \ , \quad (f\in\mathcal{F}, i\in\mathcal{N}) \ . \tag{2.9}$$

- If node i is the destination node for session f (i.e., $i=d(f)$), then

$$\sum_{j\in\mathcal{T}_i} r_{ji}(f) = r(f) \ , \quad (f\in\mathcal{F}) \ . \tag{2.10}$$

It can be easily verified that once (3.3) and (3.4) are satisfied, (3.5) must also be satisfied. As a result, it is sufficient to just include (3.3) and (3.4) in the formulation.

We assume that fixed modulation and coding scheme (MCS) is used for each data stream and each data stream corresponds to one unit data rate. For each link (i, j), the sum of the data rates over all sessions that traverse this link cannot exceed the average data rate that can be supported by link (i, j) over T time slots. Then on each link (i, j), we have:

$$\sum_{f \in \mathcal{F}} r_{ij}(f) \leq \frac{1}{T} \sum_{t=1}^{T} z_{ij}[t] \ , \quad (i \in \mathcal{N}, j \in \mathcal{T}_i) \ , \tag{2.11}$$

where T is the number of time slots in a frame.

2.2.2 Formulation

In multi-hop networks, a fundamental problem is to maximize the achievable end-to-end session throughput. Given that there are multiple sessions in the network, we also need to consider fairness issue when maximizing throughput among competing sessions. Fairness can be defined in different ways. In this study, we choose the objective of maximizing the minimum throughput among all the active sessions in the network ($r_{\min}$) . This objective aims to achieve both throughput maximization and fairness among the sessions. The problem can be formulated as follows:

OPT-raw

max	$r_{\min}$
s.t	$r_{\min} \leq r(f), \quad f \in \mathcal{F}$;
	half duplex constraints: (2.1) - (2.3);
	ordering and DoF constraints: (2.4) - (2.7);
	flow balance constraints: (3.3), (3.4), (2.11);

In this formulation, A_i, N and T are constants, $x_i[t], y_i[t], z_{ij}[t], \pi_i[t]$ and $\theta_{ji}[t]$ are integer variables, and $r_{\min}, r(f)$ and $r_{ij}(f)$ are continuous variables. Note that the two sets of constraints in (2.6) and (2.7) are stated in the form of sufficient conditions rather than in the form of mathematical programming. Therefore, a reformulation of (2.6) and (2.7) is needed.

Reformulation For constraint (2.6), if $x_i[t] = 1$, then we have $\sum_{j \in \mathcal{T}_i} z_{ij}[t] + \sum_{j \in \mathcal{I}_i} (\theta_{ji}[t] \sum_{k \in \mathcal{T}_j}^{k \neq i} z_{kj}[t]) \leq A_i$. On the other hand, if $x_i[t] = 0$, then no DoF is consumed. Constraint (2.6) can be reformulated by incorporating binary variable $x_i[t]$ into the expression as follows:

$$\sum_{j\in\mathcal{T}_i} z_{ij}[t] + \sum_{j\in\mathcal{I}_i}\left(\theta_{ji}[t]\sum_{k\in\mathcal{T}_j}^{k\neq i} z_{kj}[t]\right) \leq A_i x_i[t] + (1-x_i[t])B_i$$

$$(i\in\mathcal{N}, 1\leq t\leq T)\ , \qquad (2.12)$$

where $B_i = \sum_{j\in\mathcal{I}_i} A_j$ is an upper bound of $\sum_{j\in\mathcal{I}_i}(\theta_{ji}[t]\sum_{k\in\mathcal{T}_j}^{k\neq i} z_{kj}[t])$.

Similarly, constraint (2.7) can be reformulated as follows:

$$\sum_{j\in\mathcal{T}_i} z_{ji}[t] + \sum_{j\in\mathcal{I}_i}\left(\theta_{ji}[t]\sum_{k\in\mathcal{T}_j}^{k\neq i} z_{jk}[t]\right) \leq A_i y_i[t] + (1-y_i[t])B_i$$

$$(i\in\mathcal{N}, 1\leq t\leq T)\ . \qquad (2.13)$$

Note that constraints (2.12) and (2.13) have nonlinear terms $\sum_{j\in\mathcal{I}_i}(\theta_{ji}[t]\sum_{k\in\mathcal{T}_j}^{k\neq i} z_{kj}[t])$ and $\sum_{j\in\mathcal{I}_i}(\theta_{ji}[t]\sum_{k\in\mathcal{T}_j}^{k\neq i} z_{jk}[t])$, respectively. To remove the nonlinear terms in the formulation, we employ the *Reformulated-Linearization Technique* (RLT) [39, Chapter 6]. For constraint (2.12), we introduce a new variable $\lambda_{ji}[t] = \theta_{ji}[t]\sum_{k\in\mathcal{T}_j}^{k\neq i} z_{kj}[t]$. Then constraint (2.6) can be replaced by the following linear constraint:

$$\sum_{j\in\mathcal{T}_i} z_{ij}[t] + \sum_{j\in\mathcal{I}_i} \lambda_{ji}[t] \leq A_i x_i[t] + (1-x_i[t])B_i\ , \qquad (i\in\mathcal{N}, 1\leq t\leq T)\ . \qquad (2.14)$$

Now we need to add constraints for $\lambda_{ji}[t]$. Since $\theta_{ji}[t]$ is a binary variable (therefore $0\leq\theta_{ji}[t]\leq 1$) and $0\leq\sum_{k\in\mathcal{T}_j}^{k\neq i} z_{kj}[t]\leq A_j$, then the following constraints must hold:

$$[\theta_{ji}[t]-0]\cdot[\sum_{k\in\mathcal{T}_j}^{k\neq i} z_{kj}[t]-0]\geq 0\ , \qquad (i\in\mathcal{N}, j\in\mathcal{I}_i, 1\leq t\leq T)\ , \qquad (2.15)$$

$$[\theta_{ji}[t]-0]\cdot[A_j-\sum_{k\in\mathcal{T}_j}^{k\neq i} z_{kj}[t]]\geq 0\ , \qquad (i\in\mathcal{N}, j\in\mathcal{I}_i, 1\leq t\leq T)\ , \qquad (2.16)$$

$$[1-\theta_{ji}[t]]\cdot[\sum_{k\in\mathcal{T}_j}^{k\neq i} z_{kj}[t]-0]\geq 0\ , \qquad (i\in\mathcal{N}, j\in\mathcal{I}_i, 1\leq t\leq T)\ , \qquad (2.17)$$

$$[1-\theta_{ji}[t]]\cdot[A_j-\sum_{k\in\mathcal{T}_j}^{k\neq i} z_{kj}[t]]\geq 0\ , \qquad (i\in\mathcal{N}, j\in\mathcal{I}_i, 1\leq t\leq T)\ . \tag{2.18}$$

Substituting $\lambda_{ji}[t]$ for $\theta_{ji}[t]\sum_{k\in\mathcal{T}_j}^{k\neq i} z_{kj}[t]$ in the above constraints, we obtain

$$\lambda_{ji}[t]\geq 0\ , \qquad (i\in\mathcal{N}, j\in\mathcal{I}_i, 1\leq t\leq T)\ , \tag{2.19}$$

$$\lambda_{ji}[t]\leq A_j\cdot\theta_{ji}[t]\ , \qquad (i\in\mathcal{N}, j\in\mathcal{I}_i, 1\leq t\leq T)\ , \tag{2.20}$$

$$\lambda_{ji}[t]\leq \sum_{k\in\mathcal{T}_j}^{k\neq i} z_{kj}[t]\ , \qquad (i\in\mathcal{N}, j\in\mathcal{I}_i, 1\leq t\leq T)\ , \tag{2.21}$$

$$\lambda_{ji}[t]\geq A_j\cdot\theta_{ji}[t]+\sum_{k\in\mathcal{T}_j}^{k\neq i} z_{kj}[t]-A_j\ , \qquad (i\in\mathcal{N}, j\in\mathcal{I}_i, 1\leq t\leq T)\ . \tag{2.22}$$

Similarly, for constraint (2.7), we introduce a new variable $u_{ji}[t]=\theta_{ji}[t]\cdot\sum_{k\in\mathcal{T}_j}^{k\neq i} z_{jk}[t]$. Then, constraint (2.7) can be replaced by the following set of linear constraints:

$$\sum_{j\in\mathcal{T}_i} z_{ji}[t]+\sum_{j\in\mathcal{I}_i} u_{ji}[t]\leq A_i y_i[t]+(1-y_i[t])B_i\ , \qquad (i\in\mathcal{N}, 1\leq t\leq T)\ , \tag{2.23}$$

$$u_{ji}[t]\geq 0\ , \qquad (i\in\mathcal{N}, j\in\mathcal{I}_i, 1\leq t\leq T)\ , \tag{2.24}$$

$$u_{ji}[t]\leq A_j\cdot\theta_{ji}[t]\ , \qquad (i\in\mathcal{N}, j\in\mathcal{I}_i, 1\leq t\leq T)\ , \tag{2.25}$$

$$u_{ji}[t]\leq \sum_{k\in\mathcal{T}_j}^{k\neq i} z_{jk}[t]\ , \qquad (i\in\mathcal{N}, j\in\mathcal{I}_i, 1\leq t\leq T)\ , \tag{2.26}$$

$$u_{ji}[t]\geq A_j\cdot\theta_{ji}[t]+\sum_{k\in\mathcal{T}_j}^{k\neq i} z_{jk}[t]-A_j\ , \qquad (i\in\mathcal{N}, j\in\mathcal{I}_i, 1\leq t\leq T)\ . \tag{2.27}$$

The problem is now reformulated as follows:

OPT

max $\quad r_{\min}$

s.t $\quad r_{\min}\leq r(f) \quad f\in\mathcal{F}$;

half duplex constraints: (2.1) - (2.3);

ordering and DoF constraints: (2.4), (2.5), (2.14), (2.19) - (2.27);

flow balance constraints: (3.3), (3.4),(2.11);

In this formulation, A_i, B_i, N and T are constants, $x_i[t], y_i[t], z_{ij}[t], \pi_i[t]$ and $\theta_{ji}[t]$ are integer variables, and $r_{\min}, r(f), r_{ij}(f), \lambda_{ji}[t]$ and $\mu_{ji}[t]$ are continuous variables. This optimization problem is in the form of a mixed-integer linear program (MILP), which is NP-hard in general. The theoretical worst-case complexity of solving a general MILP problem is exponential [54]. It cannot be solved by commercial solvers for moderate sized networks. Thus, we need to develop an efficient algorithm to find a competitive solution.

2.3 Construction of an Initial Feasible Solution

2.3.1 Overview

The proposed solution consists of two stages. In the first stage, we obtain an initial feasible solution through a customized algorithm. The algorithm is based on the so-called *sequential fixing* (SF) technique [39, Chapter 6]. Another technique to find an initial feasible solution is to first fix route of each session (e.g., by using *Shortest Path* routing). After sessions' routes are given, the number of integer variables in the original MILP is reduced. However, the reduced problem is still MILP, which is NP-hard in general. On the other hand, SF finds a feasible solution by solving a series of LPs, each of which has a polynomial-time complexity (see Section 2.5). Since our goal in this chapter is to develop a polynomial time algorithm to solve the MILP, we employ SF to find an initial feasible solution.

The idea of SF is as follows. For a mixed integer linear program, if the integer variables are fixed, then the MILP would be reduced to an LP, which can be solved by commercial softwares in polynomial time. Therefore, the key challenge here is how to fix the values for all integer variables. This can be done by solving the linear relaxation of the original problem, where all integer variables are relaxed to continuous variables. Then we can fix the values of integer variables in iterations. Once we fix some integer variables' values in an iteration, we build a new MILP for the remaining integer variables and solve the new relaxed LP. A naive application of SF will fix the values of the relaxed variables solely based on the closeness to integers. However, the performance of such an application may be an issue due to the fact that it does not consider the relationships among

the variables. In the first stage of our algorithm, the decision of which variables to fix is carefully designed according to the roles they play in the optimization problem. Figure 2.1 shows the flow chart of finding a good initial feasible solution.

To obtain an initial feasible solution, we need to fix the integer values for $\mathbf{x}[t]$, $\mathbf{y}[t]$, $\mathbf{z}[t]$, $\boldsymbol{\pi}[t]$ and $\boldsymbol{\theta}[t]$, where $\mathbf{x}[t]$, $\mathbf{y}[t]$ and $\boldsymbol{\pi}[t]$ represent vectors $[x_1[t],\ x_2[t], ...,\ x_N[t]]$, $[y_1[t],\ y_2[t], ...,\ y_N[t]]$ and $[\pi_1[t],\ \pi_2[t], ...,\ \pi_N[t]]$. $\mathbf{z}[t]$ and $\boldsymbol{\theta}[t]$ represent matrices $[z_{ij}[t]]_{N\times N}$ and $[\theta_{ji}[t]]_{N\times N}$. In our problem formulation, the ordering variables $\boldsymbol{\pi}[t]$ and $\boldsymbol{\theta}[t]$ play a key role in the solution since they determine how IC is performed at each node. Therefore, in Step 1, we first fix the values of $\boldsymbol{\pi}[t]$ and $\boldsymbol{\theta}[t]$.

After Step 1, we obtain an ordered node list for each time slot, which can be used to calculate DoF consumption for IC at each node. Therefore, in Step 2, we can determine whether a link can be active or not based on the DoF constraints (2.6) and (2.7). If link (i, j) can be active in time slot t, we set $z_{ij}(t) > 0$, otherwise, we set $z_{ij}(t) = 0$. Then, we can fix the values of corresponding scheduling variables $\mathbf{x}[t]$ and $\mathbf{y}[t]$. Note that in this step, we do not fix the exact integer value of $\mathbf{z}[t]$ for active links.

After Step 2, all active links are identified. Note that since the route for each session is not predetermined, there may exist some active links that do not contribute to session throughput. Although these links do not affect the feasibility of the solution, however, some other nodes in the network may have to consume their DoFs to do IC for the transmit and receive nodes of these links. To avoid wasting DoF resources, we release DoF allocation on such "dummy" links in Step 3. Then the route of each session is now identified by the remaining active links.

In Step 4, we fix the integer values of $\mathbf{z}[t]$ according to their closeness to integer floor. After Step 4, all integer variables are fixed and we obtain an initial feasible solution for the MILP.

2.3.2 Algorithm Details

Now we give detailed descriptions of each step in finding an initial feasible solution.

Step 1: Fixing π and θ variables. Since the ordering of a node determines its DoF consumption for IC, in Step 1, we fix the integer values of ordering variables $\boldsymbol{\pi}[t]$ and $\boldsymbol{\theta}[t]$. In each iteration, for

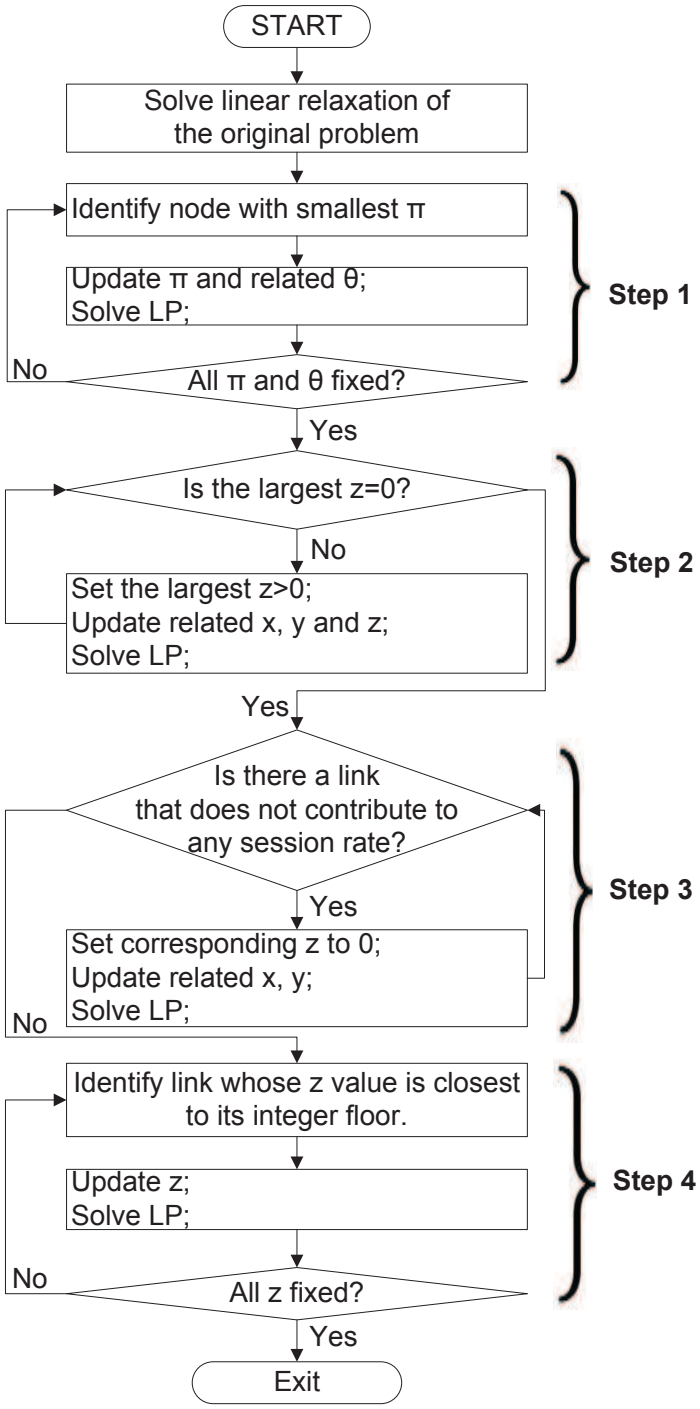


Figure 2.1: A flow chart for finding a good initial feasible solution.

$1 \leq t \leq T$, we fix one $\pi_i[t]$ variable and related $\theta_{ij}[t]$ variables. Since the values of $\pi_i[t]$ and $\theta_{ij}[t]$ variables in one time slot are independent of their values in another time slot, they may be fixed in parallel. Then we build a new MILP for the remaining integer variables and solve the new relaxed LP. Since there are N nodes in the network, we need $(N-1)$ iterations.

Specifically, as Fig. 1 shows, at the beginning of the first stage, we solve the linear relaxation of the original problem, where all integer variables in our model are relaxed to continuous variables. That is, before Step 1, we obtain a solution in which $\boldsymbol{\pi}[t]$ and $\boldsymbol{\theta}[t]$ are continuous variables. In the first iteration of Step 1, we identify node i with the smallest $\pi_i[t]$ value and set $\pi_i[t] = 1$, which means that we put node i at position 1 in the ordered node list. Once the position of node i is fixed, we know that all the other nodes are behind node i, and therefore, we set $\theta_{ji}[t] = 0$ and $\theta_{ij}[t] = 1$ for $j \in \mathcal{N}, j \neq i$ (according to Eq. (2.5)). With the integer values of $\pi_i[t]$ and corresponding $\boldsymbol{\theta}[t]$ fixed, we build a new MILP and solve its linear relaxation. In each iteration, we fix the π-value for one node. After $(N-1)$ iterations, all the nodes in the network have their positions fixed in the ordered node list.

Step 2: Fixing x and y variables. In Step 1, we determine an order value for each node so that IC can be performed according to this order. In Step 2, we determine the status of each link based on the DoF consumptions at its transmit and receive nodes. Then we fix the values of corresponding scheduling variables $\mathbf{x}[t]$ and $\mathbf{y}[t]$. Note that in this step, we do not fix the exact number of data streams on each active link ($\mathbf{z}[t]$). The reason behind this is that leaving the exact value of $\mathbf{z}[t]$ to be fixed later can help achieve a better solution when the algorithm terminates.

Specifically, in the first iteration, for $1 \leq t \leq T$, we identify link (i, j) that has the largest $z_{ij}[t]$ value and fix its status as active (by adding constraint $z_{ij}[t] > 0$). Once we set link (i, j) as active in time slot t, we can fix $x_i[t] = 1$ and $y_j[t] = 1$ (according to constraints (2.2) and (2.3)). To accelerate the algorithm, we identify all links that cannot be active simultaneously with link (i, j) and fix as many $\mathbf{x}[t]$, $\mathbf{y}[t]$ and $\mathbf{z}[t]$ values as possible by the following rules.

- **Links associated with nodes i and j.** Since we already identified link (i, j) as an active link in time slot t, according to constraint (2.1), we can fix $y_i[t] = 0$ for transmit node i and $x_j[t] = 0$ for receive node j. Then we can fix the status of all incoming links to node i as

inactive (according to constraint (2.3)) and the status of all outgoing links from node j as inactive (according to constraint (2.2)). That is, we set $z_{ki}[t] = 0$ for each node $k \in \mathcal{T}_i$ and $z_{jk}[t] = 0$ for each node $k \in \mathcal{T}_j$.

- **Links associated with those nodes that should perform IC for node i or node j.** Since link (i, j) is identified as active in time slot t, there is at least one data stream on it. For transmit node k, if its interference range covers receive node j and node k is after node j in the ordered node list (i.e., $\pi_k[t] > \pi_j[t]$), then node k has to consume DoFs to cancel its interference to node j. According to constraint (2.2) and (2.6), node k can be an active transmitter only if it has enough DoFs for IC and also at least one remaining DoF for SM. To check if node k meets the requirement, we first assume that there is only one data stream on link (i, j), then we calculate the minimum number of DoFs node k uses to perform IC by constraint (2.6). If the minimum DoF consumption for IC equals to or is larger than the number of DoFs at node k, then it cannot be an active transmitter (set $x_k[t] = 0$). Therefore, all outgoing links from node k cannot be active (set corresponding $\mathbf{z}[t]$ to zero) based on constraint (2.2). Similarly, for transmit node i, some nodes in the network cannot be active receivers in time slot t based on constraint (2.7). Therefore, all the incoming links to these nodes cannot be active in time slot t based on constraint (2.3). Thus, we can set the corresponding $\mathbf{y}[t]$ and $\mathbf{z}[t]$ to zero.

- **Links associated with those nodes that require node i or node j to perform IC.** Since link (i, j) is active in time slot t, transmit node i and receive node j both have to consume DoFs for IC. For receive node j, in addition to SM, it has to use its DoFs to cancel the interference from all unintended neighboring transmit nodes that are before itself in the ordered node list. We calculate the DoF consumption at node j. If it is equal to its total number of DoFs, then node j does not have any spare DoFs to perform IC for additional transmit nodes. As a result, an inactive node h, where $h \in \mathcal{I}_j$ and $\pi_h[t] < \pi_j[t]$, cannot be activated as a transmitter (set $x_h[t] = 0$). Moreover, all outgoing links from node h cannot be active (set corresponding $\mathbf{z}[t]$ to zero) based on constraint (2.2). Similarly, for transmit node i, some of the nodes in the network cannot be receivers. Moreover, based on constraint (2.3), all the incoming links

to these nodes cannot be active. Thus, we set the corresponding $\mathbf{y}[t]$ and $\mathbf{z}[t]$ to zero.

Based on the above discussions, we can identify all links that cannot be active simultaneously with link (i, j). After fixing the integer values of $\mathbf{x}[t]$ and $\mathbf{y}[t]$, we build a new MILP for the remaining integer variables and solve a new relaxed LP. If the largest $z_{ij}[t]$ value among the to-be-determined links is 0, it means that all possible active links have been fixed. Then we fix the remaining links as inactive links (i.e., set corresponding $\mathbf{z}[t]$ to zero). For the transmit and receive nodes of these links, if their $\mathbf{x}[t]$ (or $\mathbf{y}[t]$) values are not yet fixed, we set them to zero.

Step 3: Release DoF allocation on links that do not carry any data flow. In Step 2, we reserve DoF resources for all links that are allowed to be active. During each iteration, for link (i, j) with the largest $z_{ij}[t]$ value, we reserve DoF resources for it by adding constraint $z_{ij}[t] > 0$. However, since the route of each session is not predetermined, it may change throughout iterations in Step 2. Therefore, after Step 2, link (i, j) may not be used in any session's route. Although there are DoF resources allocated for SM on link (i, j), the link may not carry any data flow. Such DoF allocation is feasible, but it wastes DoF resources in the network since some other nodes in the network have to use their precious DoFs to perform IC due to these active links. These wasted DoF resources could have been used more productively for SM to achieve a higher session throughput. So it is necessary to release DoF allocation on links that do not carry any data flow.

Specifically, if we find link (i, j) with $r_{ij}(f) = 0$ for every session f traversing link (i, j), then we set $z_{ij}[t] = 0$ for every time slot t. If all the outgoing links from node i are inactive in time slot t, we set the status of node i to inactive (i.e., $x_i[t] = 0$) based on constraint (2.2). Similarly, if all incoming links to node j are inactive in time slot t, we set the status of node j as inactive (i.e., $y_j[t] = 0$) based on constraint (2.3). Then we update $\mathbf{x}[t]$, $\mathbf{y}[t]$ and $\mathbf{z}[t]$ and solve the LP. The process repeats until all such links are cleared.

Step 4: Fixing z variables. In steps 2 and 3, we set $\mathbf{z}[t]$ to 0 for inactive links. However, for the active links, the exact values of $\mathbf{z}[t]$ remain unknown. In this step, we fix the integer values of $\mathbf{z}[t]$ for active links. To do this, we run the following process in each iteration. For each time slot, among all the active links with unfixed $\mathbf{z}[t]$, we identify link (i, j) whose $z_{ij}[t]$ value is closest to its integer floor $\lfloor z_{ij}[t] \rfloor$ and set $z_{ij}[t] = \lfloor z_{ij}[t] \rfloor$. Then we update $z_{ij}[t]$ and solve the LP. The iteration

continues if there still exists fractional $\mathbf{z}[t]$ value. Eventually, all integer variables are fixed and we obtain an initial feasible solution.

Lemma 1. *A solution obtained by the first stage of our algorithm is feasible.*

Proof. In the first stage of our algorithm, we wish to obtain an initial feasible solution for the MILP. We first reduce the original MILP to LP by fixing the values of integer variables ($\boldsymbol{\pi}[t]$, $\boldsymbol{\theta}[t]$, $\mathbf{x}[t]$, $\mathbf{y}[t]$ and $\mathbf{z}[t]$) through four steps. Then we obtain the values of other variables by solving the LP to its optimality. The feasibility of the solution can be verified by checking whether all the integer variables are indeed integers, and whether all the constraints are satisfied after the integer variables are fixed. Specifically, in Step 1, we fix the integer values of ordering variables $\boldsymbol{\pi}[t]$ and $\boldsymbol{\theta}[t]$ through iterations. In each iteration, we fix one $\pi_i[t]$ by constraint (2.4) and fix its related $\boldsymbol{\theta}[t]$ by constraint (2.5). After these variables are fixed, we obtain values of other variables by solving a relaxed LP to its optimality. Therefore, after Step 1, $\boldsymbol{\pi}[t]$ and $\boldsymbol{\theta}[t]$ are fixed and all the constraints remain satisfied.

In Step 2, we identify the status of each link and fix all the scheduling variables $\mathbf{x}[t]$ and $\mathbf{y}[t]$. In each iteration, we identify one link (i, j) to be active (set $z_{ij}[t] > 0$) only if constraints (2.6) and (2.7) are satisfied. Then we fix corresponding scheduling variables $\mathbf{x}[t]$ and $\mathbf{y}[t]$ according to constraints (2.1) – (2.3). The values of other variables are obtained by solving a relaxed LP to its optimality. After Step 2, $\mathbf{x}[t]$ and $\mathbf{y}[t]$ are fixed and all the constraints remain satisfied.

In Step 3, we set $\mathbf{z}[t] = 0$ for "dummy" links and update corresponding scheduling variables $\mathbf{x}[t]$ and $\mathbf{y}[t]$ so that constraints (2.2) and (2.3) are satisfied. Since constraints (2.6) and (2.7) set an upper bound for DoF consumption at each active node, they remain satisfied. Since these links do not contribute to any session's throughput, the flow balance constraints (3.3), (3.4) and (2.11) are satisfied.

In Step 4, we fix the values of $\mathbf{z}[t]$ for active links. We set $\mathbf{z}[t]$ to their integer floors so that the sum of DoF consumption for SM and IC at each node will not exceed its available DoFs. That is, constraints (2.6) and (2.7) are satisfied at each node. After Step 4, $\mathbf{z}[t]$ are fixed and all the constraints remain satisfied.

As a result, all integer variables in our original MILP are fixed. The MILP is reduced to a

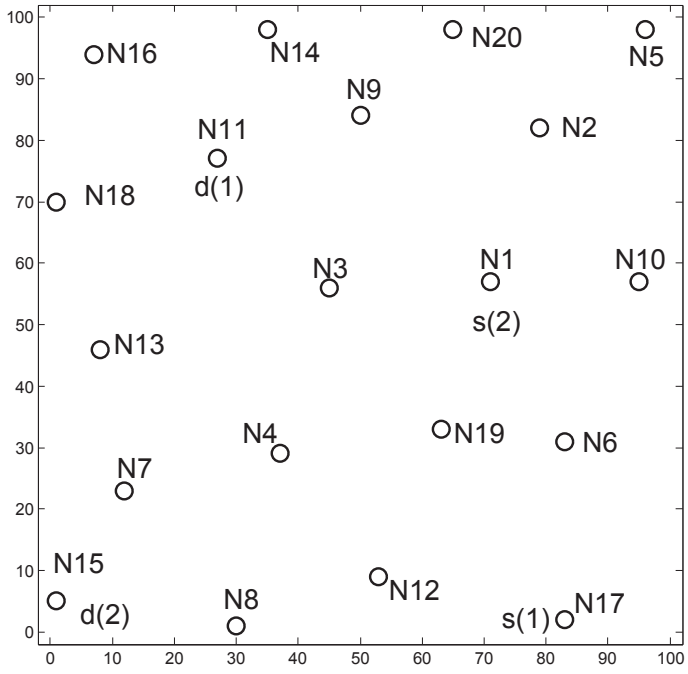


Figure 2.2: An example of a 20-node network.

LP, and values of other variables are obtained by solving this LP to its optimality. Therefore, all constraints in the original MILP are satisfied under the obtained solution. □

2.3.3 An Example

As discussed in Section 2.3.1, compared to other approaches such as setting sessions' routes first (e.g., with shortest path routing), SF is able to find a feasible solution in polynomial time. In addition, since SF technique preserves the flexibility of multipath routing for each session, the throughput for each session in the solution tends to be larger. We use an example to illustrate this point. Consider an example network with 20 nodes and 2 sessions in Figure 2.2, where $s(1)$, $d(1)$, $s(2)$ and $d(2)$ represent the source and destination nodes for sessions 1 and 2, respectively. We assume that each node in the network is equipped with four antennas and a node's transmission and interference ranges are 30 and 50, respectively. For scalability, we normalize all units for distance and time with appropriate dimensions. There are four time slots in a frame. The objective value (r_{min}) found by SF is 0.75, while the objective value found by first setting each session's route with

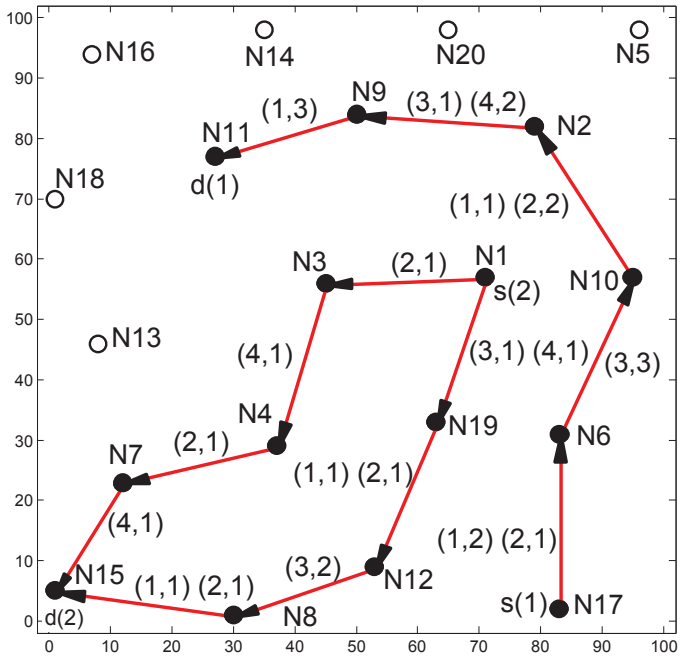


Figure 2.3: An initial feasible solution for flow routing and scheduling through SF.

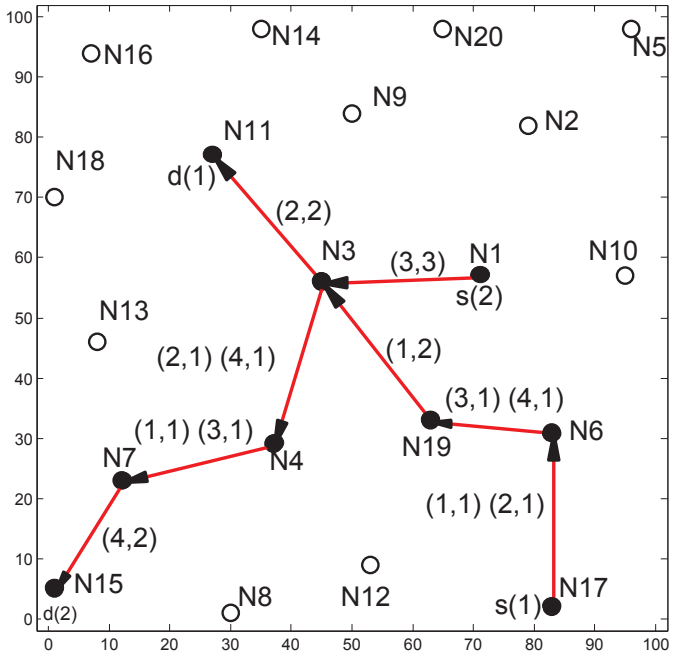


Figure 2.4: An initial feasible solution for flow routing and scheduling by first setting the route of each session to be shortest path.

shortest path is 0.5. Figure 2.3 and Figure 2.4 show the routing topologies and scheduling for each session by SF and shortest path, respectively. The tuple next to each link represents the time slot index of a frame in which the number of data streams are transmitted. For example, in Figure 2.3, $(1, 3)$ next to link (N_9, N_{11}) denotes that in time slot 1, there are 3 data streams on this link. In the case when there are multiple such tuples next to a link, it means that this link is active over multiple time slots in a frame. Comparing Figure 2.3 and Figure 2.4, SF has the following two advantages. First, SF allows flow splitting. In Figure 2.3, the route of session 2 splits from the source node N_1 into two branches and then merge at the destination node N_{15}. On the other hand, in Figure 2.4, the route of session 2 obtained by shortest path is only one of the two branches in Figure 2.3. SF uses DoF resources at more nodes, and therefore has a larger session throughput. Second, SF can make more efficient use of the DoF resources in the network. In Figure 2.4, node N_3 is the intersection of two sessions. Since both of these two sessions need to use the DoF resources at node N_3, it becomes the bottleneck node. However, in Figure 2.3, SF finds another route for session 1 to avoid intersection with session 2. Therefore, the initial feasible solution obtained by SF can utilize DoF resources among the nodes in the network more efficiently than the technique with shortest path routing.

2.4 Improving the Initial Feasible Solution

2.4.1 Overview

After the first stage, we have an initial feasible solution. In the second stage of our algorithm, we identify bottleneck link and improve throughput in each iteration until no further increment is possible. In essence, the second stage of the proposed solution is an iterative greedy algorithm. Figure 2.5 shows its flow chart. In Step 5, we find the current $r_{\min}$ and identify the session with this bottleneck throughput. For the session with this minimum throughput,[2] we identify the link

[2]In case of a tie, we choose the session with the smallest session number (index). This will eliminate any randomness in choosing a session.

associated with this bottleneck throughput among all links traversed by this session.[3]

In Step 6, we try to enlarge the "pipe" of the bottleneck link by increasing the number of data streams (DoFs) on this link. This is done by examining: (i) at the transmit and receive nodes of this link, if there is any remaining DoFs (in any time slot); and (ii) for nodes that are behind these two nodes in the ordered node list, if there are enough remaining DoFs for IC should a new data stream is added to the bottleneck link.

If Step 6 is not successful, then in Step 7, we try to see if any change in node ordering can change the DoF consumption at a node. This step is motivated by the fact that the relative position of a node in the ordered node list affects its IC behavior and DoF consumption.

If Step 7 is not successful, we try to modify the current routes for the bottleneck session. The design space here is large and will have different complexity and performance trade-off. We adopt a simple approach in this step. Specifically, for the underlying bottleneck link (for which steps 6 and 7 fail), we check if we can find a nearby relay node so that we can add a new route in parallel to this link.

The algorithm continues as long as any of steps 6 to 8 is successful, in which case we update $\mathbf{z}[t]$ and corresponding $\mathbf{x}[t]$, $\mathbf{y}[t]$, $\boldsymbol{\pi}[t]$ and $\boldsymbol{\theta}[t]$ variables as needed. The algorithm terminates when none of steps 6 to 8 is successful.

2.4.2 Algorithm Details

Now we give detailed descriptions of each step in the second part of our algorithm.

Step 5: Finding bottleneck link. In this step, we are given a feasible solution which gives values for $\mathbf{x}[t]$, $\mathbf{y}[t]$, $\mathbf{z}[t]$ and $\boldsymbol{\theta}[t]$. With these fixed values for $\mathbf{x}[t]$, $\mathbf{y}[t]$, $\mathbf{z}[t]$, problem OPT degenerates into a LP. By solving this LP, we can find the current $r_{\min}$. Subsequently, we identify a session with rate $r_{\min}$, which we denote as f. A tie is broken by choosing the session with the smallest session number (index). Such deterministic tie-breaking mechanism ensures that we keep working on the same session before moving on to the next one. For the chosen session, we find a bottleneck link, i.e., constraint (2.11) is binding. A tie among multiple bottleneck links may be broken arbitrarily.

[3] In case of a tie, we may randomly choose one from the bottleneck links.

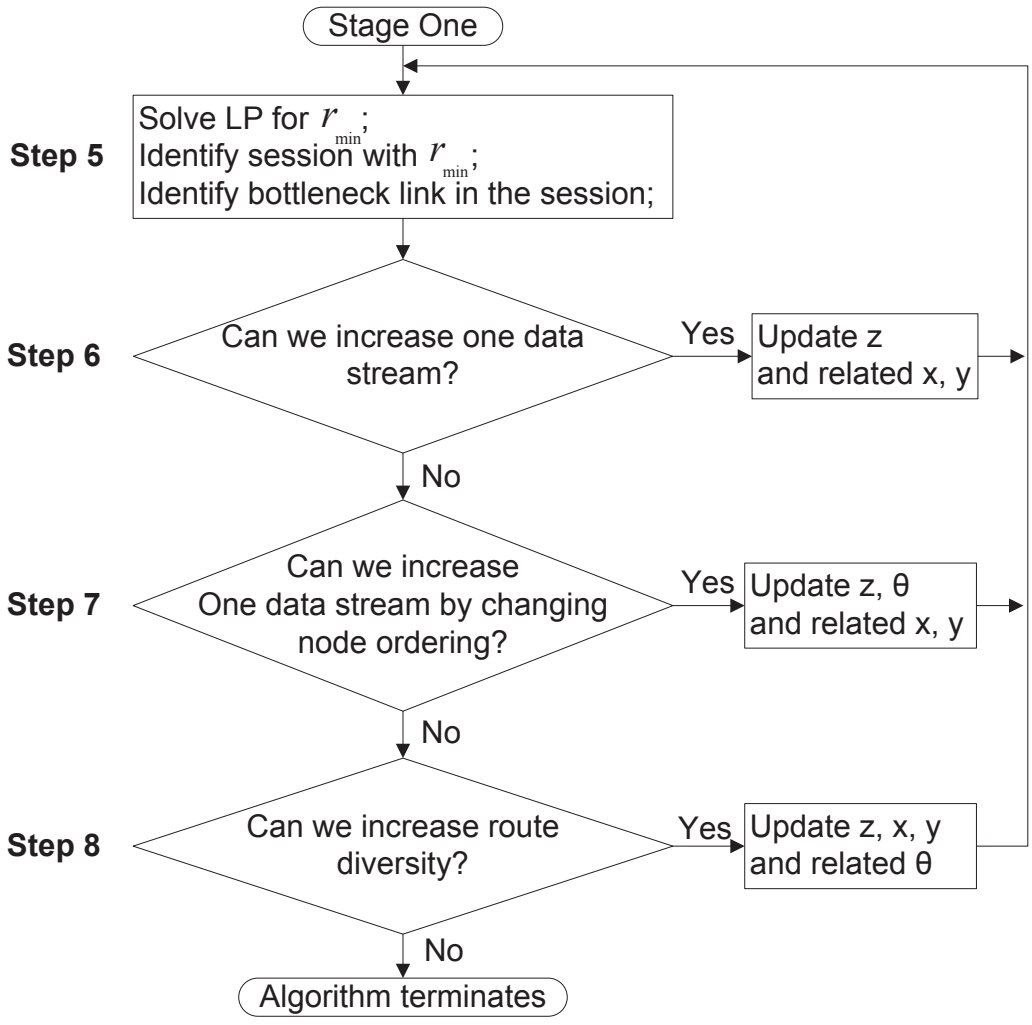


Figure 2.5: A flow chart for the second stage of our proposed algorithm.

Step 6: Adding a data stream. Denote (i, j) as the bottleneck link that we have identified for session f in Step 5, where i is the transmit node and j is the receive node. In Step 6, we try to increase one data stream on this link in some time slot over a frame. This increment is a successful one if the following conditions are satisfied:

- **(C-1):** Both transmit and receive nodes i and j have at least one remaining DoF.
- **(C-2):** For receive nodes after i (and within i's interference range) in the ordered node list, there is at least one DoF available for IC. Likewise, for transmit nodes after j (and have j in their interference range) in the ordered list, there is at least one DoF available for IC.

In the case when the above increment is successful, then $z_{ij}[t]$ is incremented by 1, and $x_i[t]$ and $y_j[t]$ are updated to 1. Since we have T time slots in a frame, we will check each time slot in Step 6.

Step 7: Adjusting node ordering. Step 6 will fail if condition (C-1) or (C-2) cannot be satisfied in the same time slot. Note that based on our DOF IC model, the ordering of nodes has a profound impact on each node's DoF consumption for IC. Therefore, in Step 7, we will try to adjust the node ordering $\pi[t]$ in each time slot to see if both conditions (C-1) and (C-2) can be satisfied. We propose a two-phase ordering change, denoted (A-1) and (A-2), to address the requirements in (C-1) and (C-2), respectively.

(A-1): Since condition (C-1) is not satisfied, we have that either transmit node i or receive node j does not have any remaining DoF. We consider the case for transmit node i first. The case for receive node j is similar.

The ordered list L_1 in Figure 2.6 shows the current ordering of nodes in the network, in which we have shown the position of transmit node i in this order as well as those nodes (i.e., p, m and k in this example) that are receive nodes before node i in L_1 and are within node i's interference range. We do not identify other receive nodes (except p, m and k) before node i in L_1 because they are outside the interference range of i. Among receive nodes p, m and k, k is closest to node i in L_1. Our idea of adjusting node ordering for node i is as follows. Since transmit node i has run out of its DoFs, it is likely that it is using some of its DoFs for IC to nodes p, m, and k. If we could

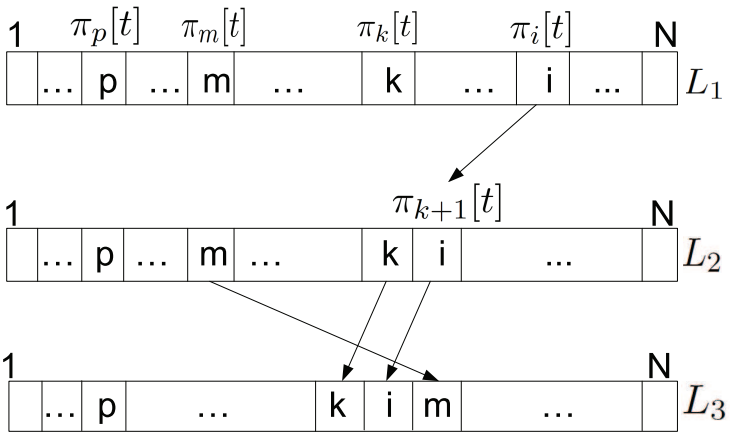


Figure 2.6: A schematic illustrating the process of adjusting node ordering.

move node i before one of these nodes, then the IC burden on node i will be reduced, allowing some DoFs to be freed up for SM of one more data stream. The outcome of such reordering (successful or not) depends on the DoF consumption on each node (after reordering) and whether the DoF constraints (2.6) and (2.7) can be met. There are many ways to move up node i (before p, m or k) in the ordered list. In the following, we present an algorithm that we have designed for this purpose.

To reduce the IC burden on node i, we want to choose a receive node (i.e., p, m and k in this example) and put it after node i. This move will add IC burden on the chosen node (p, m or k) as it will be responsible for IC for more transmit nodes (transmit nodes that are among the nodes between itself and node i in L_1). To reduce the number of these new transmit nodes for IC, we first move node i to position $\pi_k[t]+1$. This will cause the set of nodes between node k and node i in L_1 to be shifted to the right by one position, as shown in L_2 in Figure 2.6. Note that this operation will not change DoF consumption at any node in the network and DoF constraints at all nodes remain satisfied.

Now we need to choose a receive node among p, m and k in L_2 and move it after node i. A receive node is eligible for selection if it has enough DoFs available to cancel interference from all the interfering transmit nodes before itself in the node list after it is moved behind node i. We

check nodes p, m, and k individually for its eligibility. Among the eligible nodes,[4] we choose the one (say m) that has the most remaining DoFs after this move.[5] L_3 in Figure 2.6 shows the ordered node list after m is moved after node i, where all the nodes between nodes m and i in L_2 have been shifted by one position to the left. In L_3, transmit node i is before receive node m and is no longer responsible for canceling its interference to node m. As a result, node i has at least one DoF available for SM to transmit one more data stream on the bottleneck link (i, j).

For receiver node j, it is not hard to see that a similar approach can be applied to increase its available DoFs. To conserve space, we omit its discussion here.

(A-2): Since condition (C-2) is not satisfied, we know that either (i) for a receive node after transmit node i (and within i's interference range) in the ordered node list, there is no DoF left for IC; or (ii) for a transmit node after receive node j (and has j in its interference range) in the ordered node list, there is no DoF left for IC.

For (i), denote h as such a receive node. Then we will try to change the order for node h so that it will have at least one DoF available. But this is precisely the same reordering problem that we would have done for receive node j in (A-1). Therefore, the same node reordering procedure can be applied to node h. For (ii), again the reordering problem is precisely the same as that for transmit node i in (A-1) and therefore the same node reordering procedure for i can be applied.

Both (A-1) and (A-2) are performed in each time slot until a data stream can be added or they fail in all time slots.

Step 8: Improving route diversity. In steps 6 and 7, we try to increase one data stream on the bottleneck link (i, j). When both steps fail, it suggests that it may be futile to add one more data stream on this bottleneck link (i, j). A plausible approach is to open up some other routes (i.e., multiple parallel paths) between nodes i and j so that the extra data stream can be diverted over the new path. There has been extensive research on finding multiple paths between two nodes [53, 59]. For the purpose of this chapter, we show one simple algorithm that only employs one extra relay node to create a second path between nodes i and j.

A node k can be considered as a relay node only if k can serve as node i's receive node in one

[4] When there is no eligible node, we move on to the next time slot.

[5] In case of a tie, we choose the node with the smallest node index.

time slot and node j's transmit node in a different time slot. For node k, we need to check whether both links (i,k) and (k,j) can support one more data stream. For either link (i,k) or link (k,j), we are addressing the same problem for link (i,j) in steps 6 and 7. Therefore, procedures in steps 6 and 7 can be applied. If both links (i,k) and (k,j) can support one more data stream, then we update $\mathbf{z}[t]$, $\mathbf{x}[t]$, $\mathbf{y}[t]$ and $\boldsymbol{\theta}[t]$, and return to Step 5. Otherwise, the algorithm terminates.

Lemma 2. *A solution following the successful outcome of Step 6, 7, or 8 is feasible.*

Proof. The feasibility of the solution following the successful exit of Step 6, 7, or 8 can be verified by checking whether the DoF constraints (2.6) and (2.7) are satisfied at each node. Specifically, during Step 6, one more data stream can be added to the bottleneck link only if (C-1) and (C-2) are satisfied. If (C-1) and (C-2) are satisfied, then the DoF constraints (2.6) and (2.7) must remain satisfied at each node after the extra data stream is added to the bottleneck link. Therefore, if Step 6 is successful, then the DoF constraints (2.6) and (2.7) must be satisfied at each node.

If Step 6 fails, it indicates that either (C-1) or (C-2) cannot be satisfied under current node ordering. Then in Step 7, we try to alter the node ordering by using (A-1) and (A-2). (A-1) and (A-2) address the requirements in (C-1) and (C-2), respectively. If operations in Step 7 are successful, (C-1) and (C-2) are satisfied and therefore the DoF constraints (2.6) and (2.7) must remain satisfied at each node.

If Step 7 fails, it indicates that either (C-1) or (C-2) cannot be satisfied. In Step 8, we try to employ a relay node to create a second path between the transmit and receive nodes of the bottleneck link. A node can be chosen as a relay node only if constraint (2.1) is satisfied. Then the same algorithms in Step 6 and Step 7 are applied to these two new links. Thus, Step 8 is successful only if (C-1) and (C-2) are satisfied and therefore the DoF constraints (2.6) and (2.7) must remain satisfied at each node. □

2.4.3 An Example

In our DoF IC model, the ordering of nodes has a profound impact on each node's DoF consumption for IC. In Step 7, we exploit this unique property of our model to improve a solution.

Now we use an example to give the details of this process. Consider an example network with 20 nodes and 2 sessions shown in Figure 2.7, where $s(1)$, $d(1)$, $s(2)$ and $d(2)$ represent the source and destination nodes for sessions 1 and 2, respectively. We assume that each node in the network is equipped with four antennas and a node's transmission and interference ranges are 30 and 50, respectively. There are four time slots in a frame. The objective value (r_{min}) obtained after the first stage is 0.25, and is increased to 0.5 after the second stage. Figures 2.8 and 2.9 show the routing topologies and scheduling for each session at the end of the first and second stages, respectively. Note that the flow routing is not changed in the second stage, but the scheduling behavior is changed significantly.

After the first stage, in Figure 2.8, we can see that the first link of session 2 (link (N_8, N_{18})) is active in time slot 1 with one data stream. Since there are four time slots and the minimum session throughput is 0.25, link (N_8, N_{18}) is a bottleneck link (constraint (2.11) is binding). After the second stage, as shown in Figure 2.9, link (N_8, N_{18}) can transmit one more data stream in time slot 2. Therefore, its link throughput is increased to 0.5 and is no longer a bottleneck link. This increment is made possible by adjusting the node ordering list in time slot 2. We now show the details of this adjustment. Figure 2.10 shows the set of active nodes and their DoF allocation in time slot 2 after the first stage of our algoirthm (SF). Since link (N_8, N_{18}) is not active in time slot 2, we first check if we can activate it without changing the node ordering. The current node ordering list in time slot 2 is $\{N_1, N_6, N_8, N_{13}, N_{12}, N_{16}, N_9, N_{10}, N_{14}, N_{15}, N_2, N_3, N_4, N_5, N_7, N_{11}, N_{17}, N_{18}, N_{19}, N_{20}\}$ (see Fig. 2.11). For transmit node N_8, if it is active, it does not need to consume any DoF for IC since there is no active receive node before it. Therefore, it has 4 DoFs for SM. For receive node N_{18}, if it is active, it has to consume four DoFs to cancel interference from nodes N_9, N_3 and N_4. Then it does not have remaining DoFs to receive any data streams from node N_8. Therefore, to activate link (N_8, N_{18}), we move node N_{18} before one of the three nodes (N_9, N_3 or N_4) so that node N_{18} no longer needs to use its DoFs to cancel interference from that node. Figure 2.11 shows the details of this move. First, We move node N_{18} from position 18 to position 14 (after transmit node N_4). Note that this operation will not change DoF consumption at any node. Then we choose a node from N_9, N_3 and N_4 and move it after node N_{18}. As described in Section 2.4.2, the selected node should

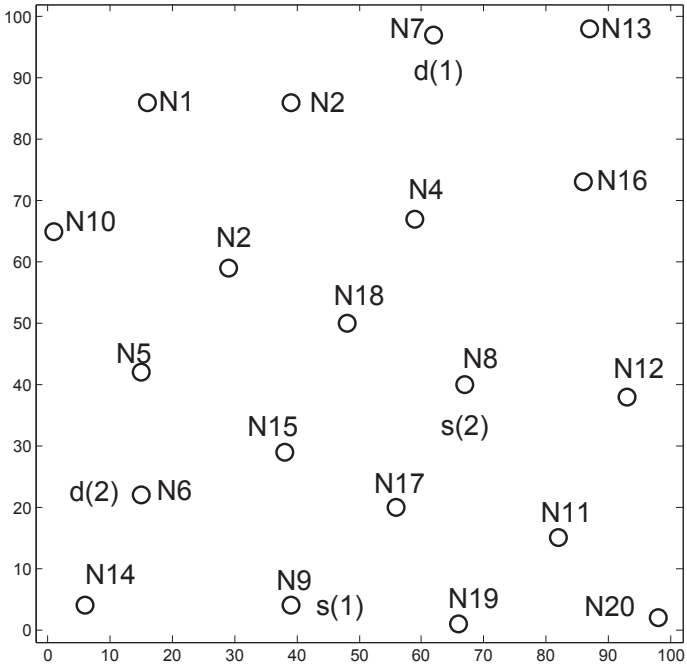


Figure 2.7: An example of a 20-node network.

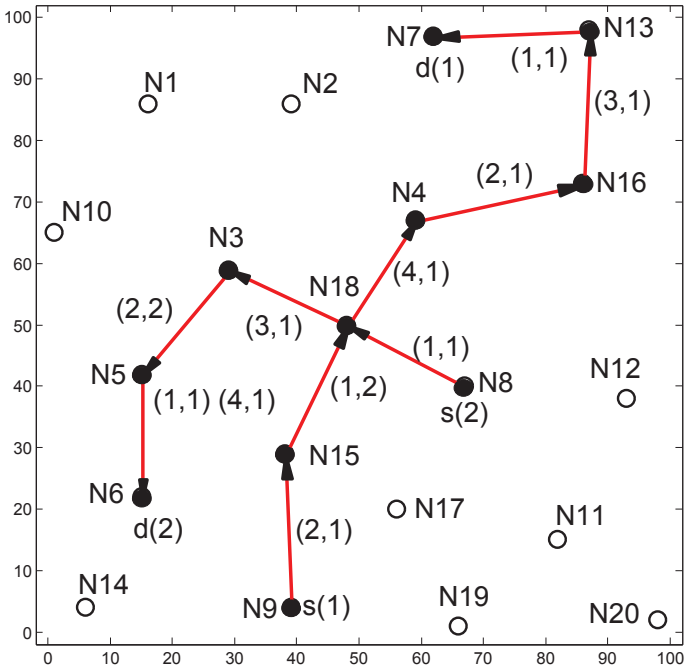


Figure 2.8: Flow routing and scheduling solution for the example network after the first stage of our algorithm.

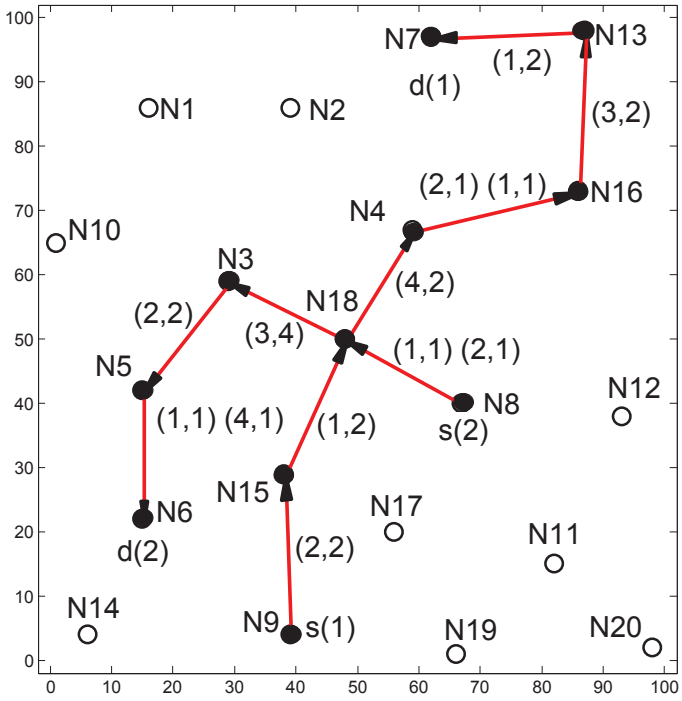


Figure 2.9: Flow routing and scheduling solution for the example network after the second stage of our algorithm.

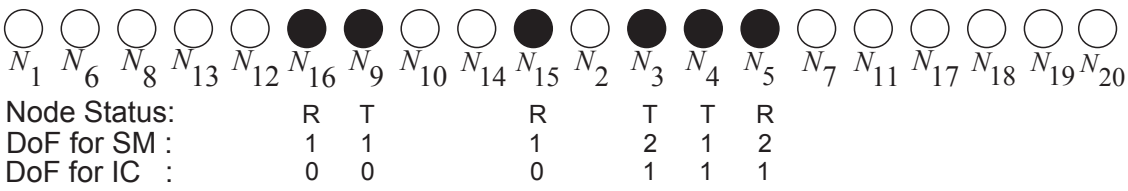


Figure 2.10: DoF allocation at each active node in time slot 2 at the end of the first stage of our algorithm.

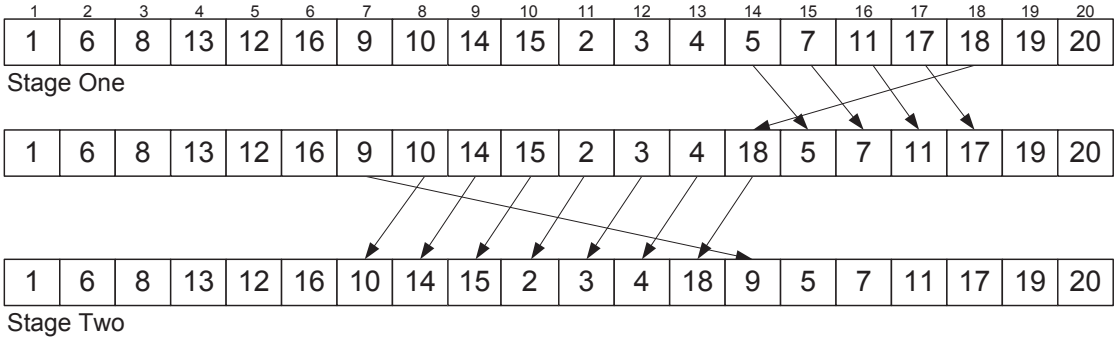


Figure 2.11: A schematic illustrating the process of adjusting the ordering for node 18.

have the most remaining DoFs after this move. According to Figure 2.10, in time slot 2, node N_9 consumes one DoF for SM, node N_3 consumes three DoFs (two for SM and one for IC) and node N_4 consumes two DoFs (one for SM and one for IC). Now we check how many additional DoFs these nodes need for IC if they are moved to position 14. For nodes N_3 or N_4, if it is moved to position 14, the number of receive nodes ahead them does not change (N_{16} and N_{15}). For node N_9, if it is moved to position 14, there is one more receive node ahead it (node N_{15}). However, node N_{15} is the intended receiver of node N_9, as shown in Figure 2.8, then there is no interference to be canceled. Therefore, if being moved to position 14, none of the three nodes needs to consume additional DoFs for IC and node N_9 has the most remaining DoFs (3). As Figure 2.11 shows, we move node N_9 from position 7 to position 14, which is the current position for node N_{18}. This move will shift nodes between positions 8 to 14 to the left by one position. After this move, node N_{18} is before N_9 in the node ordering list. If node N_{18} is active, it no longer needs to consume one DoF to cancel interference from node N_9. Instead, it can use this available DoF to receive one data stream from node N_8. Meanwhile, node N_9 can use one DoF to cancel its interference to node N_{18}. Now, link (N_8, N_{18}) can be active in time slot 2 with one data stream.

2.5 Complexity Analysis

We now show that the proposed algorithm has a polynomial time complexity. For each stage of the algorithm, we analyze the number of required iterations and the complexity of each iteration.

In the first stage, for each step, the complexity of each iteration involves solving an LP, searching for the largest/smallest values, and fixing the integer values for selected variables. The complexity of solving an LP is $O(V^3)$ [70], where V is the number of variables. It is not hard to see that the complexity of searching for the largest values and fixing integer variables are much lower than solving an LP. Therefore, the complexity of an iteration is $O(V^3)$.

We now analyze the total number of iterations in the first stage. Note that in each iteration, we fix some variables for all time slots. Since there are N nodes in the network, we need $(N-1)$ iterations for Step 1 as we discussed. Since there are $O(N^2)$ links, we need $O(N^2)$ iterations to determine status of each link (Step 2), identify wasted links (Step 3), and fix the z values for active links (Step 4). Therefore, we have no more than $O(N+3N^2) = O(N^2)$ iterations. As a result, the complexity of the first stage of our algorithm is $O(N^2V^3)$.

In the second stage, the complexity of each step involves solving an LP and identifying a bottleneck link (Step 5), increasing data streams on this link (Step 6 and 7) and adding a parallel path for this link (Step 8). It is not hard to see that solving an LP has the highest complexity among the four steps in an iteration. Therefore, the complexity of an iteration is $O(V^3)$.

We now analyze the total number of iterations in the second stage of our algorithm. Since the links in the network is upper bounded by $O(N^2)$, and for each link, we can increase its data streams by at most A_i times in each time slot. Therefore, the total number of iterations is $O(N^2AT)$, where $A = \max_{i\in\mathcal{N}}\{A_i\}$. As a result, the second stage of our algorithm has an overall complexity of $O(N^2AT \cdot V^3)$.

In our algorithm, three sets of variables, $\theta_{ji}[t]$, $z_{ij}[t]$ and $r_{ij}(f)$, dominate the total amount of variables. Both $\theta_{ji}[t]$ and $z_{ij}[t]$ have $O(N^2 \cdot T)$ variables while $r_{ij}(f)$ has $O(N^2 \cdot |\mathcal{F}|)$ variables. Therefore, $V = O(N^2 \cdot \max\{T, |\mathcal{F}|\})$. In summary, our algorithm has a polynomial time complexity of $O(N^2V^3) + O(N^2AT \cdot V^3) = O(N^2AT \cdot V^3)$, where $V = O(N^2 \cdot \max\{T, |\mathcal{F}|\})$.

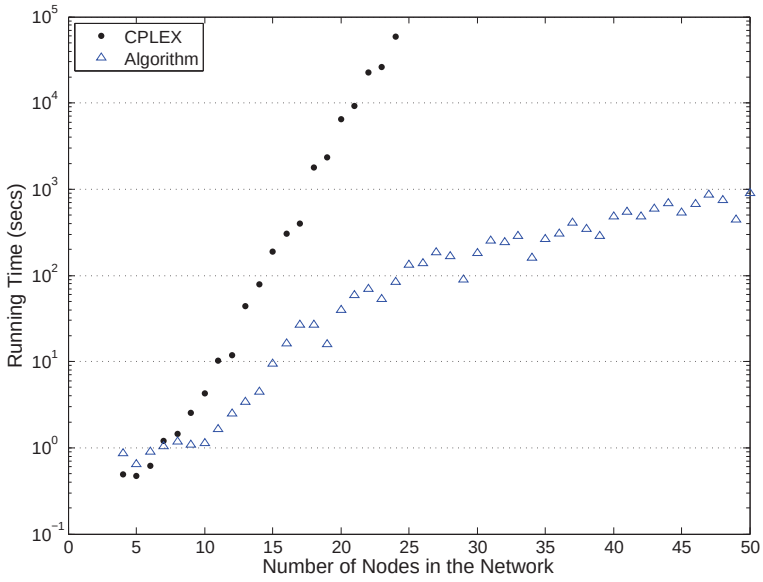


Figure 2.12: Running time required by the proposed algorithm and CPLEX.

2.6 Simulation Results

In this section, we present simulation results to demonstrate the performance and complexity of the proposed algorithm. We also use a case study to validate the feasibility of a final solution.

2.6.1 Simulation Setting

We consider a multi-hop ad hoc network, with nodes being randomly deployed in a 100×100 area. For scalability, we normalize all units for distance and time with appropriate dimensions. We assume that each node in the network is equipped with four antennas and a node's transmission and interference ranges are 30 and 50, respectively. There are four time slots in a frame.

2.6.2 Performance and Complexity

To demonstrate the complexity and performance of our proposed algorithm, we use a commercial optimization solver, CPLEX [95], as a benchmark. The computer we use to run the simulation results has 64GB of RAM and a E5-2687w CPU.

First, we compare the complexity (in terms of running time) of the proposed algorithm and CPLEX. We increase the number of nodes in this study. For each network setting, we randomly generate 50 network instances and obtain the average running time required by CPLEX and our proposed algorithm, respectively. Figure 2.12 shows the trend of average running time required by CPLEX and the proposed algorithm as the number of nodes in the network increases from 4 to 50. Note that the y-axis in Figure 2.12 is in log-scale, indicating *exponential* running time of CPLEX. On the other hand, the running time of the proposed algorithm is orders of magnitude smaller than the time needed by CPLEX. In Figure 2.12, in the case of 25 nodes, the average running time required by CPLEX exceeds 10^5 seconds, while it only requires 10^2 seconds by our algorithm. As shown, solving MILP by CPLEX has the disadvantage of requiring an exponentially growing run-time (due to the NP-hard nature of the problem), on the other hand, our proposed algorithm only requires to solve a series of LPs, each of which has polynomial time complexity.

For performance comparison, we want to demonstrate that the results obtained by the proposed scheme are competitive when compared with the optimal results from CPLEX. Such benchmark comparison is only meaningful when the size of the problem can be handled by CPLEX. Since the average running time required by CPLEX exceeds 24 hours when there are 25 nodes in the network as shown in Fig. 12, we choose network size to be 20 nodes for performance comparison. We present comparison study for 50 random network instances, each with 2 sessions. For each network instance, the node positions are randomly generated and the source and destination nodes for each session are randomly selected. Table 2.2 shows the following two set of results for 50 network instances: (i) the ratio between the objective values obtained after the first stage of our algorithm and those from CPLEX, (ii) the ratio between the objective values obtained after the second stage of our algorithm and those from CPLEX. As shown in Table 2.2, among 50 network instances, the objective values in 19 instances are improved by employing the second stage of our algorithm. With only the first stage algorithm, the average ratio is 75.3%, with a standard deviation of 0.19. When the second stage algorithm is employed, the average ratio is improved to 85.6%, with a standard deviation of 0.12. That is, the overall performance of our proposed algorithm is within 85.6% of the optimal solution that can be computed by CPLEX within a reasonable amount

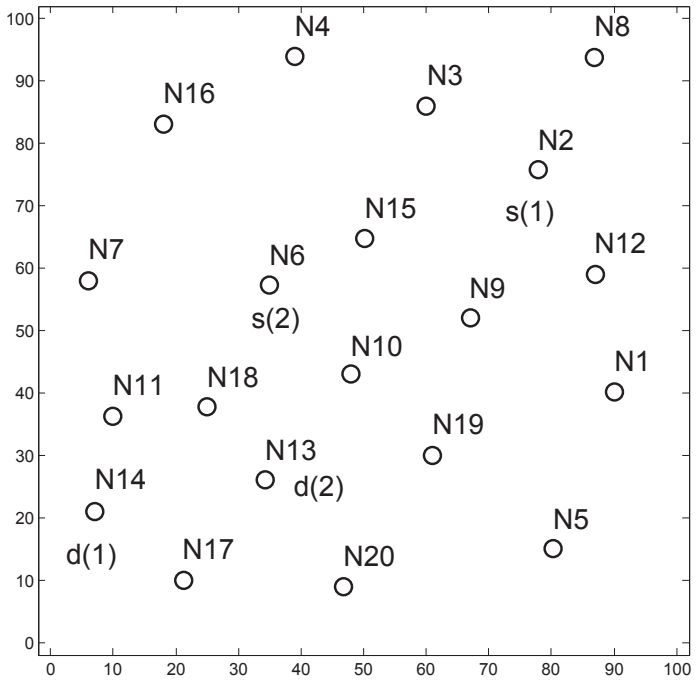


Figure 2.13: An instance of a 20-node network.

of time.

2.6.3 Feasibility of Our Solution

To validate the feasibility of the solution obtained by our proposed algorithm, we randomly pick a network instance (the 33-th) from the above 50 network instances and examine its solution details. Figure 2.13 shows the locations of the 20 nodes, where $s(1)$, $d(1)$, $s(2)$ and $d(2)$ represent the source and destination nodes for sessions 1 and 2. The objective values ($r_{\min}$) found by our algorithm and CPLEX are both 0.75, indicating the optimality of our solution for this network instance.

Although the two objective values by our algorithm and CPLEX coincide for this network instance, the flow routing and scheduling behavior under the two solutions are different. Figures 2.14 and 2.15 show the routing topologies and scheduling for each session by our algorithm and CPLEX, respectively, where the tuple next to each link represents the time slot index of a frame in which the number of data streams are transmitted. For example, in Figure 2.14, $(1, 3)$

Table 2.2: Ratio between objective values obtained by our algorithm and those from CPLEX for 50 network instances.

Network Instance	Objective Value Ratio		Network Instance	Objective Value Ratio	
	Stage 1 / CPLEX	Stage 2 / CPLEX		Stage 1 / CPLEX	Stage 2 / CPLEX
1	1	1	26	1	1
2	0.5	0.75	27	0.6	0.8
3	0.6	0.8	28	0.5	0.75
4	1	1	29	1	1
5	0.8	0.8	30	0.5	0.75
6	1	1	31	0.5	0.75
7	0.6	0.8	32	0.8	0.8
8	1	1	33	0.75	0.75
9	0.33	0.67	34	0.67	1
10	1	1	35	0.75	0.75
11	0.33	0.67	36	1	1
12	0.6	0.8	37	0.88	0.88
13	0.5	0.83	38	0.5	0.75
14	1	1	39	0.75	0.75
15	0.67	0.67	40	0.86	0.86
16	0.88	0.88	41	0.88	0.88
17	0.67	0.67	42	0.86	0.86
18	0.5	1	43	0.83	0.83
19	1	1	44	0.6	0.8
20	0.67	0.67	45	1	1
21	0.5	1	46	0.75	0.75
22	1	1	47	0.6	0.8
23	0.83	0.83	48	0.75	1
24	0.67	0.75	49	0.83	0.83
25	0.83	0.83	50	1	1

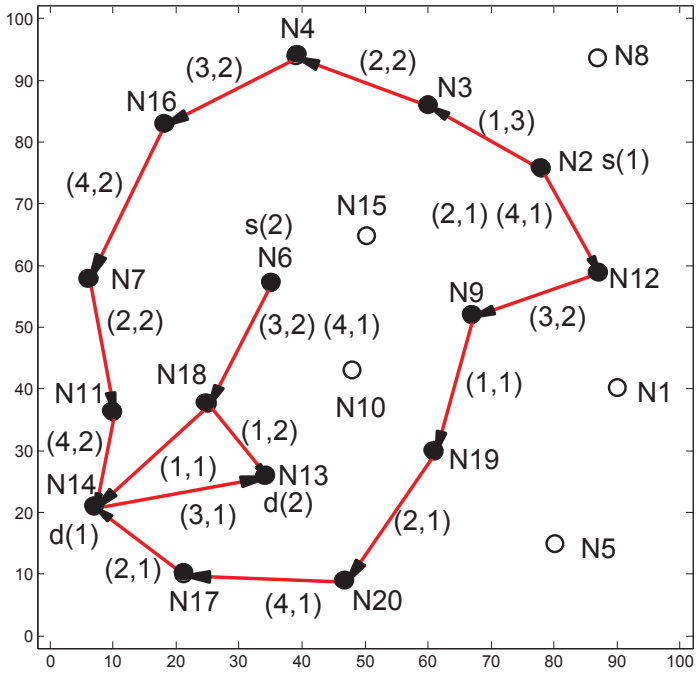


Figure 2.14: Solution for flow routing and scheduling by our algorithm.

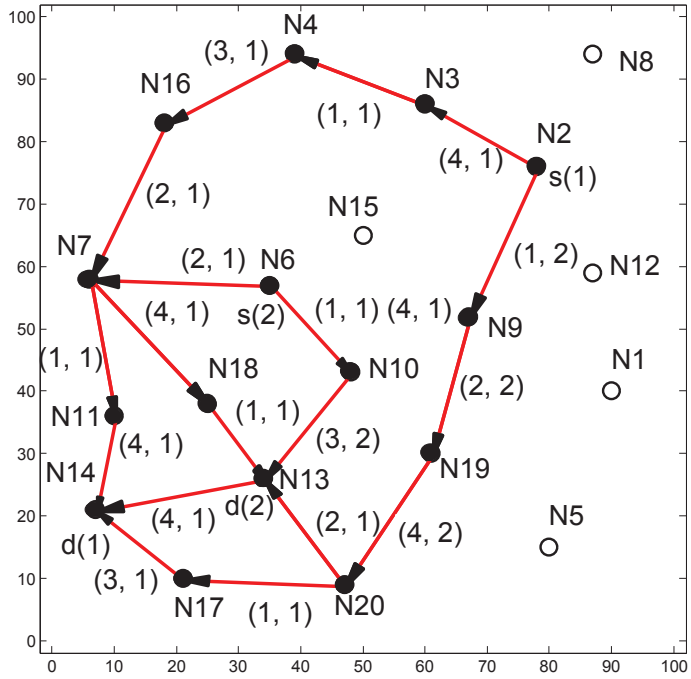


Figure 2.15: Solution for flow routing and scheduling by CPLEX.

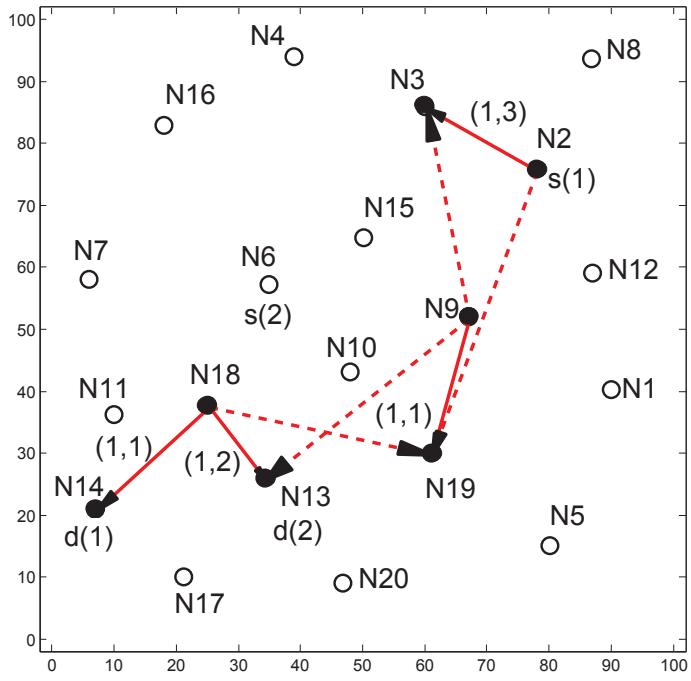


Figure 2.16: Scheduling in time slot 1 by our algorithm.

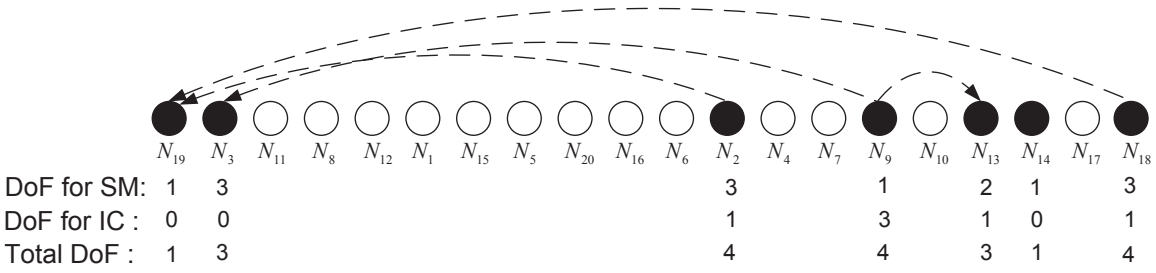


Figure 2.17: A schematic illustrating DoF consumption at each node in time slot 1.

next to link (N_2, N_3) denotes that in time slot 1, there are 3 data streams on this link. In the case when there are multiple such tuples next to a link, it means that this link is active over multiple time slots in a frame.

Now let's examine our solution. Table 2.3 shows the set of active nodes in each time slot and the DoF allocation for SM and IC at these nodes. As an example, consider the set of active nodes in time slot 1 in Table 2.3, which is shown in Figure 2.16. The interference relationships among these transmit and receive nodes are shown by the dashed arrows, i.e., node N_2 interferes N_{19}, node N_9 interferes nodes N_3 and N_{13}, and node N_{18} interferes N_{19}. It can be easily verified that by following the relative ordering of these 7 nodes in time slot 1, i.e., $N_{19}, N_3, N_2, N_9, N_{13}, N_{14}, N_{18}$, the DoF constraints in (2.6) and (2.7) are satisfied at each of these 7 nodes .

Now we discuss the details of DoF consumption at each active node following the order of their positions in the ordered list in time slot 1. Figure 2.17 shows the ordered node list, the interference relationship among active nodes (e.g., N_2 interferes N_{19}), and the summary of DoF consumption at each active node.

- N_{19}: the first active node in the ordered node list. It is a receive node. For SM, node N_{19} consumes one DoF to receive one data stream from node N_9. For IC, since it is the first node in the node list, it does not consume any DoF for IC.

- N_3: the second active node in the ordered node list. It is a receive node. For SM, node N_3 consumes three DoFs to receive three data streams from node N_2. For IC, since there is no active transmit node before node N_3 in the node list, it does not consume any DoF for IC.

- N_2: the third active node in the ordered node list. It is a transmit node. For SM, node N_2 consumes three DoFs to transmit three data streams to node N_3. For IC, only receive node N_{19} is within its interference range and before it in the ordered node list. So node N_2 needs to consume one DoF to cancel its interference to node N_{19}.

- N_9: the fourth active node in the ordered node list. It is a transmit node. For SM, node N_9 consumes one DoF to transmit one data stream to node N_{19}. For IC, only transmit node N_3 is

within its interference range and before it in the ordered list. So node N_9 needs to consume three DoFs to cancel its interference to node N_3.

- N_{13}: the fifth active node in the ordered node list. It is a receive node. For SM, node N_{13} consumes two DoFs to receive two data streams from node N_{18}. For IC, only transmit node N_9 is within its interference range and before it in the ordered node list. So node N_{13} needs to consume one DoF to cancel its interference to node N_9.

- N_{14}: the sixth active node in the ordered node list. It is a receive node. For SM, node N_{14} consumes one DoF to receive one data stream from node N_{18}. For IC, all the active transmit nodes before node N_{14} in the ordered node list are out of its interference range, it does not consume any DoF for IC.

- N_{18}: the seventh active node in the ordered node list. It is a transmit node. For SM, node N_{18} consumes two DoFs to transmit two data streams to node N_{13} and one DoF to transmit one data stream to node N_{14}. For IC, only receive node N_{19} is within its interference range and before it in the ordered node list. So node N_{18} needs to consume one DoF to cancel its interference to node N_{19}.

Note that the DoF constraints for SM and IC at each node are satisfied in time slot 1. Based on Table 2.3, the readers can easily verify that the DoF constraints at each node are also satisfied in time slot 2, 3, and 4.

2.7 Chapter Summary

DoF based IC model is a powerful tool to study network performance of multi-hop MIMO networks. In this chapter, we employed a new DoF IC model in the literature [71] to study a throughput maximization problem in a multi-hop MIMO network. Given the multi-hop network environment, the problem formulation involves joint consideration of multi-path flow routing at network layer and DoF allocation per node at link layer. Since the problem formulation is in the form of a mixed-integer linear program, we proposed to develop an efficient polynomial time algorithm to

solve it. Our algorithm design consists of two stages: the first stage is to find a quality initial feasible solution and the second stage is to improve the initial feasible solution. Specifically, in the first stage, we employed the sequential fixing technique to handle integer variables. We showed that this approach has polynomial time complexity and can offer a better initial feasible solution than some other approaches. In the second stage, we improved the initial feasible solution by exploiting the impact of node ordering on DoF consumption at a node and route diversity in the network. Simulation results showed that our solution offers competitive performance and polynomial time complexity.

Table 2.3: DoF allocation at each active node in each time slot for a 20-node network instance.

Time Slot 1				
Node Ordering	Active Node	Node Status	DoF for SM	DoF for IC
1	N_{19}	receive	1	0
2	N_3	receive	3	0
12	N_2	transmit	3	1
15	N_9	transmit	1	3
17	N_{13}	receive	2	1
18	N_{14}	receive	1	0
20	N_{18}	transmit	3	1
Time Slot 2				
Node Ordering	Active Node	Node Status	DoF for SM	DoF for IC
3	N_3	transmit	2	0
5	N_{20}	receive	1	0
6	N_{19}	transmit	1	0
8	N_2	transmit	1	0
9	N_4	receive	2	1
11	N_7	transmit	2	2
14	N_{11}	receive	2	0
15	N_{12}	receive	1	3
16	N_{14}	receive	1	2
19	N_{17}	transmit	1	3
Time Slot 3				
Node Ordering	Active Node	Node Status	DoF for SM	DoF for IC
4	N_4	transmit	2	0
7	N_{18}	receive	2	0
8	N_{12}	transmit	2	0
9	N_9	receive	2	0
10	N_6	transmit	2	2
15	N_{13}	receive	1	2
16	N_{14}	transmit	1	2
18	N_{16}	receive	2	2
Time Slot 4				
Node Ordering	Active Node	Node Status	DoF for SM	DoF for IC
3	N_{11}	transmit	2	0
4	N_{18}	receive	1	0
6	N_{20}	transmit	1	1
7	N_{17}	receive	1	2
8	N_{16}	transmit	2	1
10	N_7	receive	2	2
12	N_6	transmit	1	3
15	N_2	transmit	1	0
18	N_{12}	receive	1	0
19	N_{14}	receive	2	2

Chapter 3

Impact of Full Duplex Scheduling on End-to-end Throughput in Multi-hop Wireless Networks

3.1 Introduction

Half duplex (HD) has been the fundamental limitation in transceiver design since the beginning of wireless communications. Under HD, a transceiver can only tend to one task at a time: transmit or receive. Conceptually, HD cuts down potential throughput by half at a node. Recent breakthrough in full duplex (FD) transceiver design opens up new possibility in wireless communications. The main challenge in the design of a FD transceiver is how to cancel self-interference caused by simultaneous transmission and reception. Recent results have shown various transceiver designs [10, 11, 20, 41] that use a combination of antenna techniques, analog cancellation methods, and digital cancellation methods to suppress the self-interference.

As designs for FD transceivers continue to improve, there is a growing interest in the benefit of FD beyond physical layer, such as FD MAC protocol design [19, 30, 33, 45, 46, 68, 80, 84, 94] and cross-layer optimization problems [8, 29, 48, 50, 65, 81, 83, 85, 87]. The work in [85] presents a recent study on upper layer performance of FD in a multi-hop wireless network. Specifically,

in the first part of [85], the authors showed an asymptotic analysis of network capacity under the protocol model. The network capacity is defined as the maximum data rate that can be supported between every source-destination pair (per-flow capacity). The authors concluded that FD cannot double network capacity when the number of nodes goes to infinity. In the second part, the authors investigated the MAC-layer capacity for finite-sized network with one-hop bidirectional transmissions. The MAC-layer capacity is defined as the total capacity of all bi-directional links in the network. The results showed that with a CSMA-based MAC protocol, the capacity gain of FD over HD is below 2.

Different from [85], the goal of this chapter is to offer some fundamental understandings on the end-to-end throughput performance of FD in a finite-sized multi-hop network. In multi-hop newtorks, a fundamental problem is to compute the achievable end-to-end throughput for one or multiple active sessions. In this chapter, we study the impact of FD on end-to-end session throughput in multi-hop networks under *optimal* scheduling through rigorous mathematical formulation. Our results show that the achievable end-to-end session throughput in a FD network can be more than $2\times$ of that in a HD network within their schedulable region (i.e., traffic in both HD and FD networks can be scheduled with nonzero throughput). That is, from end-to-end throughput perspective, $2\times$ is not a fundamental performance barrier of FD in a multi-hop wireless network. This finding is enlightening and not entirely surprising, and can be explained by the much larger design space for optimal scheduling that is offered by removing the HD constraints in the problem formulation.

Our insights on why $2\times$ is not a fundamental limit of FD in multi-hop wireless networks can be explained as follows. For end-to-end session throughput in a multi-hop wireless network, the advantage of FD is not limited to achieving simultaneous transmission and reception at each node. Instead, FD also gives a much larger design space for scheduling compared to HD. Since a session's end-to-end throughput is determined by the capacity of the bottleneck link along its path, therefore, the larger scheduling space offered by FD can help to decrease the mutual interference for the bottleneck link (with fewer interfering links) and thus increase its capacity. Note that a node with FD capability does not have to exclude HD scheduling in the optimal solution, if it chooses to. As

we shall show in our numerical investigation, the end-to-end throughput ratio under FD over that under HD is no longer bounded by the $2\times$ factor that is associated with a FD node in isolation.

The remainder of this chapter is organized as follows. In Section 3.2, we review related work on FD in physical layer design, MAC layer protocol design, and network layer performance analysis, respectively. In Section 3.3, we present a mathematical model for a FD multi-hop network. Based on this model, we present a problem formulation for a throughput maximization problem. In Section 3.4, we reformulate the problem to a simpler mathematical form, which can be solved by a commercial solver such as CPLEX. In Section 3.5, we conduct numerical investigation on end-to-end throughput performance of FD in a multi-hop wireless network. Section 4.8 concludes this chapter.

3.2 Related Work

To demonstrate the feasibility of FD, various transceiver designs have been proposed. Note that simultaneous transmission and reception requires a wireless transceiver to have both the transmit and receive radio-frequency (RF) chains. There are two ways to separate these two RF chains [47]. One method is physically separating the transmit and receive antenna from one another in a separate-antenna system [20, 24, 68]. The other method is to use a circulator in a shared-antenna system [38]. However, the level of isolation provided by the above methods is not enough to extract the signal of interest from the self interference caused by signal leaking from its transmit RF chain to its receive RF chain. Therefore, the key challenge in enabling a transceiver to operate in FD mode is how to cancel self interference. Practical self-interference cancellation schemes usually use a combination of passive and active methods. Passive methods aim to suppress the self-interference before it enters the receive RF chain circuit by using various antenna techniques [6, 20, 24, 27, 28, 38, 43]. Acitve methods employ the knowledge of transmit signals and subtract it from the received signal, and can be classified into analog cancellation and digital cancellation. Analog cancellation aims to subtract the estimated self-interference signal in the analog receive circuit before the ADC [10, 11, 24–26, 41, 64]. Digital cancellation aims to subtract the estimate

self-interference signal after ADC in the digital domain [1, 26, 44, 67]. The current state-of-art FD transceiver design is the one in [10] by Bharadia *et al.* This design only requires a single antenna and is closest to what one would wish to have from a FD node. It can achieve 110dB of self-interference cancellation and thus meets the requirements of 802.11 standards.

To translate the benefit of FD in physical layer to upper layers, various MAC layer protocol designs have been proposed for both infrastructure networks and ad hoc networks. As for infrastructure networks, centralized MAC protocols are designed to manage the inter-user interference caused by simultaneous transmission and reception of AP and the hidden terminal problem caused by asymmetric data traffic [30, 41, 45, 68]. As for ad hoc networks, distributed MAC protocols are designed to support bi-directional links and two-hop relay transmissions [19, 33, 46, 84] as wells as multi-hop communications [80, 94].

Throughput performance analysis for various types of FD networks have been studied. In [81], Tong and Haenggi investigated the throughput performance for a FD wireless network consists of single-hop links. The authors quantified the impact of imperfect self-interference cancellation on the throughput gain of FD networks over HD networks under ALOHA protocol. In [48], Lee and Quek characterized the throughput of hybrid-duplex heterogeneous networks composed of multi-tier networks with access points (APs) operating in either FD bi-directional mode or HD downlink mode. The authors investigated the effect of AP spatial density and transmission power on the network throughput. In [83], Wang *et al.* established a spatial stochastic framework to investigate the network throughput gain of FD in multi-cell networks with contending single-hop links. The authors characterized the FD gain in terms of link distance, interference range network density, and carrier sensing schemes. Related work on cross-layer optimization for FD multi-hop networks includes [8, 29, 50, 65, 85, 87]. In [8], Baranwal *et al.* derived expressions for the outage probability in a multi-hop FD network as a function of the average channel fading conditions and interference. Then the optimal number of multi-hop FD relay nodes is examined so as to minimize outage probability. In [29], Fang *et al.* proposed distributed algorithms to solve two multi-path routing problems (user profit maximization problem and network power consumption minimization problem) in a FD multi-hop wireless network. In [50], Mahboobi *et al.* investigated

a joint power allocation and routing problem for a single source destination pair in a FD multi-hop network. Under optimal routing, the weighted sum of power at relay nodes is minimized while the end-to-end link outage probability is kept below a threshold. In [65], Ramirez *et al.* studied a joint routing and power allocation problem for a single source destination pair in a wireless FD network with imperfect self-interference cancellation. In [87], Yang and Shroff proposed a queue-length based CSMA-type scheduling algorithm for wireless networks with FD cut-through capability, where every node along the path of a traffic flow can receive and forward a packet simultaneously. In [85], Xie *et al.* showed an asymptotic analysis of network capacity under the protocol model for both HD and FD networks.

3.3 Modeling and Formulation

In this section, we present mathematical modeling and formulation to study end-to-end session throughput maximization problem in a multi-hop wireless network. Each node is assumed to have FD capability. For comparison, we also present modeling with HD on each node in Section 3.5.1. Table 4.1 lists the notation used in this chapter.

Consider a multi-hop wireless network with a set of nodes $\mathcal{N}$, where $N = |\mathcal{N}|$ is the number of nodes. We assume that each node in the network is equipped with a single antenna. Suppose that there is a set of $\mathcal{F}$ active sessions in the network, with $F = |\mathcal{F}|$. Denote $s(f)$ and $d(f)$ as the source and destination nodes of session f, respectively, where $f \in \mathcal{F}$. For each session f, we assume the route from $s(f)$ to $d(f)$ is known *a priori*, which can be found by some routing algorithm (e.g., AODV [61], DSR [42]). For fairness in comparison, we employ the same routing for networks with FD or HD. This allows us to focus on the impact of scheduling.

Scheduling Constraints. We assume time slot based scheduling, with T equal-length time slots in a frame. Within a time slot, a node may transmit and receive at the same time. When a node transmits, all the neighboring one-hop nodes will hear the signal. For unicast communication, among all the nodes that hear the signal, only one node belongs to the unicast session and will decode the signal. Consider a transmitter and its intended receiver as a (logical) link. Denote $\mathcal{L}$ as

Table 3.1: Notation in Chapter 3

Symbol	Definition
$\mathcal{N}$	Set of nodes in the network
$\mathcal{F}$	Set of sessions in the network
T	Total number of time slots in a frame
$\mathcal{L}$	Set of links in the network
$\mathcal{L}_i^{\mathrm{In}}$	The set of incoming links at node $i \in \mathcal{N}$
$\mathcal{L}_i^{\mathrm{Out}}$	The set of outgoing links at node $i \in \mathcal{N}$
$x_l(t)$	A binary variable to indicate whether or not link l is active in time slot t
$c_l(t)$	Capacity of link l in time slot t
$u_l(t)$	SINR of link l in time slot t
$r_l(f)$	The data rate on link l that is attributed to session $f \in \mathcal{F}$
$r(f)$	The throughput of session $f \in \mathcal{F}$
P	Transmit power
G	A large constant
M	Number of line segments in linear approximation
$z_{lm}(t)$	A binary variable to indicate the specific segment m that $u_l(t)$ falls within in linear approximation
$\lambda_{lm}(t)$	Weight factor in linear approximation
$g_{(\mathrm{Tx}(k),\,\mathrm{Rx}(l))}$	Path attenuation loss from $\mathrm{Tx}(k)$ to $\mathrm{Rx}(l)$
$n_{\mathrm{Rx}(l)}$	Ambient noise at node $\mathrm{Rx}(l)$
W	Network bandwidth
β	The capability of self-interference cancellation
ϵ	Approximation error
$s(f)$	The source node of session $f \in \mathcal{F}$
$d(f)$	The destination node of session $f \in \mathcal{F}$

the set of all possible links in the network and $x_l(t)$ as a binary variable to indicate whether or not link $l \in \mathcal{L}$ is active in time slot t, i.e., $x_l(t) = 1$ if link l is active and 0 otherwise. Then we have:

$$\sum_{l \in \mathcal{L}_i^{\text{Out}}} x_l(t) \leq 1 \ , \quad (i \in \mathcal{N}, 1 \leq t \leq T) \ , \tag{3.1}$$

$$\sum_{k \in \mathcal{L}_i^{\text{In}}} x_k(t) \leq 1 \ , \quad (i \in \mathcal{N}, 1 \leq t \leq T) \ . \tag{3.2}$$

where $\mathcal{L}_i^{\text{Out}}$ and $\mathcal{L}_i^{\text{In}}$ denote the sets of all possible outgoing and incoming links at node i, respectively.

Flow Balance Constraints. Denote $r(f)$ as the end-to-end throughput of session $f \in \mathcal{F}$ and $r_l(f)$ as the data rate on link l that is attributed to session f. Then we have the following flow balance constraints. If node i is the source node of a session, then

$$\sum_{l \in \mathcal{L}_i^{\text{Out}}} r_l(f) = r(f) \ , \quad (f \in \mathcal{F}, i = s(f)) \ . \tag{3.3}$$

If node i is an intermediate relay node for session f, then

$$\sum_{k \in \mathcal{L}_i^{\text{In}}} r_k(f) = \sum_{l \in \mathcal{L}_i^{\text{Out}}} r_l(f) \ , \quad (f \in \mathcal{F}, i \in \mathcal{N}, i \neq s(f), i \neq d(f)) \ . \tag{3.4}$$

If node i is the destination node for session f, then

$$\sum_{k \in \mathcal{L}_i^{\text{In}}} r_k(f) = r(f) \ , \quad (f \in \mathcal{F}, i = d(f)) \ . \tag{3.5}$$

It can be easily verified that once (3.3) and (3.4) are satisfied, (3.5) is also satisfied. Therefore, it is sufficient to include (3.3) and (3.4) in the formulation.

Link Rate Constraints. Denote $c_l(t)$ as the achievable capacity of link l in time slot t. Then $c_l(t)$ can be calculated by Shannon's capacity formula by treating both mutual interference (from other links) and self interference (due to FD) as noise. Note that under FD, there is still non-negligible self interference that cannot be completely canceled [10]. It is therefore important to incorporate such remaining self-interference when calculating link capacity. For each transmit node, we assume that its transmit power is P when it is "on" (i.e., $x_l(t) = 1$) and 0 when it is

"off". Denote $g_{(\mathrm{Tx}(k),\,\mathrm{Rx}(l))}$ as the path attenuation loss from transmit node of link k, $\mathrm{Tx}(k)$, to receive node of link l, $\mathrm{Rx}(l)$. Then we have,

$$c_l(t) = W \log_2 \left(1 + \frac{g_{(\mathrm{Tx}(l),\,\mathrm{Rx}(l))} \cdot P \cdot x_l(t)}{\sum_{k \in \mathcal{L}} g_{(\mathrm{Tx}(k),\,\mathrm{Rx}(l))} P x_k(t) + \sum_{m \in \mathcal{L}^{\mathrm{Out}}_{\mathrm{Rx}(l)}} \beta P x_m(t) + n_{\mathrm{Rx}(l)}} \right) , \quad (l \in \mathcal{L}, 1 \le t \le T) , \tag{3.6}$$

where W is the bandwidth. The first term in the denominator in (3.6) represents the interference signal power from other active links (mutual interference), the second term in the denominator represents residual self interference ($\mathrm{Rx}(l)$ is transmitting on its outgoing link m), and $n_{\mathrm{Rx}(l)}$ is the ambient noise at receive node $\mathrm{Rx}(l)$. The value of β characterizes the performance of self-interference cancellation – the smaller the value of β, the cleaner the cancellation of self interference.

For $t = 1, 2, \cdots, T$, the average link capacity is $\sum_{t=1}^{T} c_l(t)/T$. Since the sum of data rates over all sessions traversing a link cannot exceed the link's average capacity, We have

$$\sum_{f \in \mathcal{F}} r_l(f) \le \frac{1}{T} \sum_{t=1}^{T} c_l(t) , \quad (l \in \mathcal{L}) . \tag{3.7}$$

Problem Formulation. In this study, we are interested in end-to-end session throughput performance in a multi-hop network with FD at each node. Given that there are multiple sessions in the network, we also need to consider fairness issue when maximizing throughput among competing sessions. Fairness can be defined in different ways. In this study, we choose the objective of maximizing the minimum throughput among all active sessions in the network ($r_{\min}$) . This objective aims to achieve both throughput maximization and fairness among the sessions. The problem can be formulated as follows:

OPT-FD

max $\quad r_{\min}$

s.t $\quad r_{\min} \le r(f) \qquad (f \in \mathcal{F})$;

Scheduling constraints: (3.1), (3.2);

Flow balance constraints: (3.3), (3.4);

Link rate constraints: (3.6), (3.7);

In this formulation, W, $g_{(\text{Tx}(k),\,\text{Rx}(l))}$, β, $n_{\text{Rx}(l)}$ and P are constants; $c_l(t)$, $r(f)$ and $r_l(f)$ are continuous variables; and $x_l(t)$ is a binary variable. Note that constraints (3.6) are in the form of logarithmic functions and contain nonlinear terms inside the log functions. This optimization problem is in the form of a mixed-integer nonlinear program (MINLP), which is intractable. In the next section, we show how to linearize the nonlinear constraints in OPT-FD through reformulation and approximation.

3.4 Problem Reformulation

In this section, we linearize the nonlinear constraints (3.6) in OPT-FD. Our roadmap is as follows. First, we employ *Reformulation-Linearization Technique* (RLT) [39, Chapter 6] to remove the nonlinear term (SINR) inside the logarithmic functions without loss of optimality. After this step, we obtain a reformulated problem OPT-FD$_{\text{R}}$. Then, we approximate the logarithmic functions in (3.6) with a set of linear constraints using *piece-wise linear approximation* technique. After these two steps, we obtain a linearized problem OPT-FD$_{\text{L}}$. We show that an optimal solution to OPT-FD$_{\text{L}}$ is within $(1-\varepsilon)$-optimality of OPT-FD.

3.4.1 Reformulation of SINR Term

For (3.6), we replace the nonlinear SINR term inside the logarithmic function with a new variable $u_l(t)$. Then we obtain the following constraints:

$$c_l(t) = W \log_2(1+u_l(t)) \;, \qquad (l \in \mathcal{L}, 1 \le t \le T) \;, \tag{3.8}$$

$$u_l(t) = \frac{g_{(\text{Tx}(l),\,\text{Rx}(l))} P x_l(t)}{\sum_{k\in\mathcal{L}} g_{(\text{Tx}(k),\,\text{Rx}(l))} P x_k(t) + \sum_{m\in\mathcal{L}^{\text{Out}}_{\text{Rx}(l)}} \beta P x_m(t) + n_{\text{Rx}(l)}} \;, \qquad (l \in \mathcal{L}, 1 \le t \le T) \;. \tag{3.9}$$

(3.9) can be re-written as:

$$\sum_{k\in\mathcal{L}} g_{(\text{Tx}(k),\,\text{Rx}(l))} P x_k(t) u_l(t) + \sum_{m\in\mathcal{L}^{\text{Out}}_{\text{Rx}(l)}} \beta P x_m(t) u_l(t) + n_{\text{Rx}(l)} u_l(t) = g_{(\text{Tx}(l),\,\text{Rx}(l))} P x_l(t) \;,$$

$$(l \in \mathcal{L}, 1 \le t \le T) \;. \tag{3.10}$$

Note that (3.10) has nonlinear terms $x_k(t)u_l(t)$ and $x_m(t)u_l(t)$. To remove these nonlinear terms, we employ RLT by introducing a new variable, $a_{kl}(t) = x_k(t)u_l(t)$. Then (3.10) can be replaced by the following linear constraint:

$$\sum_{k\in\mathcal{L}} g_{(\text{Tx}(k),\,\text{Rx}(l))} P a_{kl}(t) + \sum_{m\in\mathcal{L}^{\text{Out}}_{\text{Rx}(l))}} \beta P a_{ml}(t) + n_{\text{Rx}(l)} u_l(t) = g_{(\text{Tx}(l),\,\text{Rx}(l))} P x_l(t) \ ,$$

$$(l \in \mathcal{L}, 1 \le t \le T) \ . \qquad (3.11)$$

Now we need to add constraints for $a_{kl}(t)$ and $a_{ml}(t)$. Denote G as an upper bound of $u_l(t)$ (SINR), which we can set to a large constant. Since $x_k(t)$ is a binary variable (therefore $0 \le x_k(t) \le 1$) and $0 \le u_l(t) \le G$, then the following constraints must hold:

$$[x_k(t) - 0] \cdot [u_l(t) - 0] \ge 0 \ , \quad (l \in \mathcal{L}, k \in \mathcal{L}, 1 \le t \le T) \ , \qquad (3.12)$$

$$[x_k(t) - 0] \cdot [G - u_l(t)] \ge 0 \ , \quad (l \in \mathcal{L}, k \in \mathcal{L}, 1 \le t \le T) \ , \qquad (3.13)$$

$$[1 - x_k(t)] \cdot [u_l(t) - 0] \ge 0 \ , \quad (l \in \mathcal{L}, k \in \mathcal{L}, 1 \le t \le T) \ , \qquad (3.14)$$

$$[1 - x_k(t)] \cdot [G - u_l(t)] \ge 0 \ , \quad (l \in \mathcal{L}, k \in \mathcal{L}, 1 \le t \le T) \ . \qquad (3.15)$$

Substituting $a_{kl}(t)$ for $x_k(t)u_l(t)$ in the above constraints, we obtain

$$a_{kl}(t) \ge 0 \ , \qquad (l \in \mathcal{L}, k \in \mathcal{L}, 1 \le t \le T) \ , \qquad (3.16)$$

$$G \cdot x_k(t) - a_{kl}(t) \ge 0 \ , \quad (l \in \mathcal{L}, k \in \mathcal{L}, 1 \le t \le T) \ , \qquad (3.17)$$

$$u_l(t) - a_{kl}(t) \ge 0 \ , \qquad (l \in \mathcal{L}, k \in \mathcal{L}, 1 \le t \le T) \ , \qquad (3.18)$$

$$u_l(t) + G x_k(t) - a_{kl}(t) \le G \ , \quad (l \in \mathcal{L}, k \in \mathcal{L}, 1 \le t \le T) \ . \qquad (3.19)$$

By adding the above new constraints for $a_{kl}(t)$, we have the following reformulated problem:

OPT-FD$_{\mathbf{R}}$

max	$r_{\min}$
s.t	$r_{\min} \le r(f)$ $\quad (f \in \mathcal{F})$;
	Scheduling constraints: (3.1), (3.2);
	Flow balance constraints: (3.3), (3.4).
	Link rate constraints: (3.7), (3.8), (3.11), (3.16)-(3.19).

Denote r^* and r^*_{R} as the optimal objective values of OPT-FD and OPT-FD$_{\mathrm{R}}$, respectively. Then we have the following lemma:

Lemma 3. The optimal objective values of OPT-FD and OPT-FD$_{\mathrm{R}}$ are identical, i.e, $r^* = r^*_{\mathrm{R}}$.

Proof. We show $r^* = r^*_{\mathrm{R}}$ by two steps. First, we show that an optimal solution to OPT-FD can be mapped to a feasible solution to OPT-FD$_{\mathrm{R}}$, that is, $r^* \leq r^*_{\mathrm{R}}$ holds. Then, we show that an optimal solution to OPT-FD$_{\mathrm{R}}$ can also be mapped to a feasible solution to OPT-FD, that is, $r^*_{\mathrm{R}} \leq r^*$ holds.

Step 1: In this step, we will show that $r^* \leq r^*_{\mathrm{R}}$. We prove this inequality holds by showing that if r^* is an optimal objective value of OPT-FD, then it is an achievable objective value of OPT-FD$_{\mathrm{R}}$. Suppose that r^* is an optimal objective value of OPT-FD, then there exits a solution $\psi = \big[r_{\min},\ x_l(t),\ c_l(t),\ r_l(f),\ r(f)\big]$ that satisfies all constraints in OPT-FD. Based on ψ, we can construct a solution $\hat{\psi}_R$ to OPT-FD$_{\mathrm{R}}$ as follows: $\hat{\psi}_R = \big[r_{\min},\ x_l(t),\ u_l(t),\ c_l(t),\ r_l(f),\ r(f),\ a_{kl}(t)\big]$, where $u_l(t)$ is obtained by constraint (3.9) and $a_{kl}(t) = x_k(t)u_l(t)$. We now show that the constructed solution $\hat{\psi}_R$ is a feasible solution to OPT-FD$_{\mathrm{R}}$. Since ψ satisfies constraints (3.1)– (3.4) and (3.7) in OPT-FD. it is obvious that $\hat{\psi}_R$ also satisfies these constraints in OPT-FD$_{\mathrm{R}}$. Since ψ satisfies constraint (3.6) in OPT-FD, it is easy to verify that $\hat{\psi}_R$ satisfies constraint (3.8) and (3.11) in OPT-FD$_{\mathrm{R}}$. Since $x_k(t)$ is a binary variable and $0 \leq u_l(t) \leq G$, then $a_{kl}(t) = 0$ when $x_k(t) = 0$, and $a_{kl}(t) = u_l(t)$ when $x_k(t) = 1$. Therefore, it is easy to verify that $\hat{\psi}_R$ satisfies constraints (3.16)-(3.19). Since $\hat{\psi}_R$ satisfies all constraints in OPT-FD$_{\mathrm{R}}$, it is a feasible solution to OPT-FD$_{\mathrm{R}}$. Note that the objective function in OPT-FD$_{\mathrm{R}}$ is the same as in OPT-FD, therefore, the feasible objective value corresponding to solution $\hat{\psi}_R$ is equal to r^*. Since OPT-FD$_{\mathrm{R}}$ is a maximization problem and its optimal objective value is r^*_R, then $r^* \leq r^*_R$ holds.

Step 2: In this step, we show that $r^*_R \leq r^*$. We show this inequality holds by arguing that if r^*_R is an optimal objective value of OPT-FD$_{\mathrm{R}}$, then it is also an achievable objective value of OPT-FD. Suppose that r^*_R is an optimal objective value of OPT-FD$_{\mathrm{R}}$. Then there exits a solution $\psi_R = \big[r_{\min}, x_l(t),\ u_l(t),\ c_l(t),\ r_l(f),\ r(f),\ a_{kl}(t)\big]$ that satisfies all constraints in OPT-FD$_{\mathrm{R}}$. Based on ψ_R, we can construct a solution $\hat{\psi}$ to OPT-FD as follows: $\hat{\psi} = \big[r_{\min}, x_l(t),\ c_l(t),\ r_l(f),\ r(f)\big]$. We now show that the constructed solution $\hat{\psi}$ is a feasible solution to OPT-FD. Since ψ_R satisfies constraints (3.1)– (3.4) and (3.7) in OPT-FD$_{\mathrm{R}}$, it is obvious that $\hat{\psi}$ also satisfies these constraints

in OPT-FD. Now we show that since ψ_R satisfies constraints (3.8), (3.11), and (3.16)-(3.19) in OPT-FD$_{\mathrm{R}}$, then $\hat{\psi}$ satisfies constraint (3.6) in OPT-FD. Note that constraint (3.6) in OPT-FD is identical to constraints (3.8) and (3.9) in OPT-FD$_{\mathrm{R}}$, then constraint (3.9) is replaced by constraints (3.11) and (3.16)-(3.19). Now we prove that once ψ_R satisfies constraints (3.11) and (3.16)-(3.19), it also satisfies constraint (3.9) by showing that $a_{kl}(t) = x_k(t)u_l(t)$ holds. Since $x_k(t)$ is a binary variable, according to constraints (3.16)-(3.19), $a_{kl}(t) = 0$ when $x_k(t) = 0$ and $a_{kl}(t) = u_l(t)$ when $x_k(t) = 1$, which indicates that $a_{kl}(t) = x_k(t)u_l(t)$ holds. Therefore, $\hat{\psi}$ satisfies constraint (3.6) in OPT-FD. Since $\hat{\psi}$ satisfies all constraints in OPT-FD, it is a feasible solution to OPT-FD and its corresponding objective value equals to r_R^*. Since r^* is the optimal objective value of OPT-FD, $r_R^* \leq r^*$ holds. This completes the proof. □

3.4.2 Piece-wise Linear Approximation of Log Function

In OPT-FD$_{\mathrm{R}}$, (3.8) is in the form of a logarithmic function. To linearize this logarithmic function, we employ *piece-wise linear approximation* technique. The idea is to use a finite number (say M) of line segments to approximate the logarithmic function, as shown in Fig. 3.1. First, we determine the range for $u_l(t)$. Denote $U_l^{\min}$ and $U_l^{\max}$ as the lower and upper bounds of $u_l(t)$ when link l is active ($x_l(t) = 1$), respectively. Then one lower bound could be that all interfering links are active at the same time, i.e., $U_l^{\min} = \frac{g_{(\mathrm{TX}(l),\,\mathrm{RX}(l))}P}{\sum_{k\in\mathcal{L}} g_{(\mathrm{TX}(k),\,\mathrm{RX}(l))}P + \sum_{m\in\mathcal{L}_{\mathrm{RX}(l)}^{\mathrm{Out}}} \beta P + n_{\mathrm{Rx}(l)}}$. Likewise, an upper bound could be that all interfering links (except l) are inactive at the same time, i.e., $U_l^{\max} = \frac{g_{(\mathrm{TX}(l),\,\mathrm{RX}(l))}P}{n_{\mathrm{Rx}(l)}}$. Then, for $h_l(t) = \log_2(1 + u_l(t))$, we construct a piecewise linear approximation $g_l(t)$ by connecting points $(U_l^{\min} + m\Delta, \log_2(1 + U_l^{\min} + m\Delta))$, where $\Delta = \frac{U_l^{\max} - U_l^{\min}}{M}$ and $m = 0, 1, ..., M$. We will determine the value of M to achieve the desired accuracy in Lemma 3.

Suppose $u_l(t)$ is within the mth segment between $(U_l^{\min}, U_l^{\max})$, i.e., $u_l(t) \in [U_l^{\min} + (m-1)\Delta, U_l^{\min} + m\Delta]$, then $u_l(t)$ can be represented by

$$u_l(t) = \lambda_{l,m-1}(t)[U_l^{\min} + (m-1)\Delta] + \lambda_{lm}(t)(U_l^{\min} + m\Delta)\ , \tag{3.20}$$

where $\lambda_{l,m-1}(t)$ and $\lambda_{l,m}(t)$ are two weights and satisfy:

$$\lambda_{l,m-1}(t) + \lambda_{lm}(t) = 1\ , \quad (\lambda_{l,m-1}(t) \geq 0, \lambda_{lm}(t) \geq 0)\ . \tag{3.21}$$

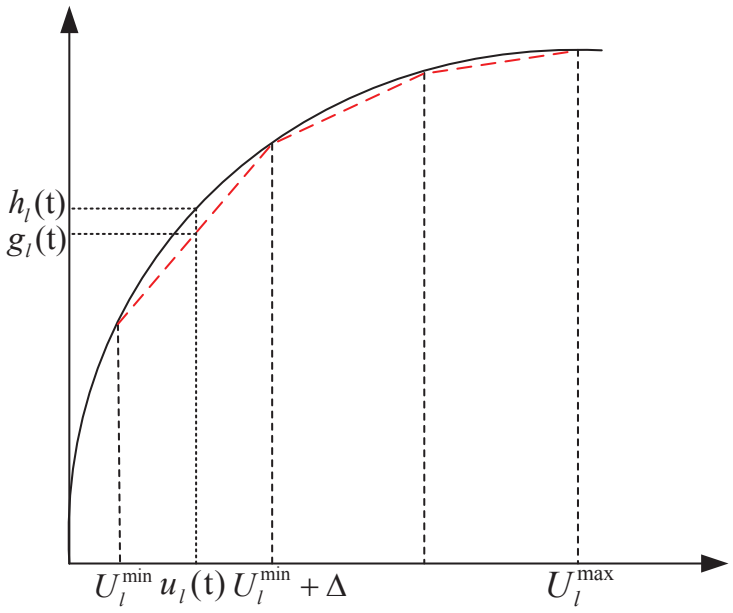


Figure 3.1: An illustration of piecewise linear approximation (with $M = 3$) for $h_l(t) = \log_2(1 + u_l(t))$.

With $u_l(t)$ in (3.20), its corresponding $g_l(t)$ value is:

$$g_l(t) = \lambda_{l,m-1}(t)\log_2(1 + U_l^{\min} + (m-1)\Delta) + \lambda_{lm}(t)\log_2(1 + U_l^{\min} + m\Delta)\ . \tag{3.22}$$

Note that (3.20) to (3.22) are for a given mth segment, $m = 1, 2, \cdots, M$. Now we extend the representation to the entire piecewise linear curve. First we need to find out the position of $u_l(t)$. We use a binary variable $z_{lm}(t)$, $1 \le m \le M$ to represent the specific segment that $u_l(t)$ falls within. If $u_l(t)$ falls within the mth segment (i.e., $U_l^{\min} + (m-1)\Delta \le u_l(t) \le U_l^{\min} + m\Delta$), then $z_{lm}(t) = 1$; otherwise, $z_{lm}(t) = 0$. Recall that $u_l(t)$ represents the SINR of link l. When link l is inactive (i.e., $x_l(t) = 0$), $u_l(t)$ is 0 and does not fall in any of the M segments (see Fig.(3.1)). Therefore, all of the $z_{lm}(t)$ value should be 0. When link l is active (i.e., $x_l(t) = 1$), $u_l(t)$ can only fall in one of the m segments, we have

$$\sum_{m=1}^{M} z_{lm}(t) = x_l(t)\ , \quad (l \in \mathcal{L}, 1 \le t \le T)\ . \tag{3.23}$$

Based on (3.20) and (3.21), when $u_l(t)$ falls within the mth segment, we can only have $\lambda_{l,m-1}(t)$ and $\lambda_{lm}(t)$ be positive and all other $\lambda_{lk}(t)$ be zero $(k \ne m-1, k \ne m)$. That is, if $\lambda_{l0}(t) > 0$ then

$z_{l1} = 1$. Since $\lambda_{l0}(t) \in [0, 1]$ and z_{l1} is a binary indicator, we have

$$\lambda_{l0}(t) \leq z_{l1}(t) \ , \qquad (l \in \mathcal{L}, 1 \leq t \leq T) \ . \tag{3.24}$$

Likewise, if $\lambda_{lm}(t) > 0$, then $z_{lm} = 1$ or $z_{l,m+1} = 1$, $1 \leq m < M$. We have

$$\lambda_{lm}(t) \leq z_{lm}(t) + z_{l,m+1}(t) \ , \quad (l \in \mathcal{L}, 1 \leq t \leq T, 1 \leq m \leq M-1) \ . \tag{3.25}$$

Finally, if $\lambda_{lM}(t) > 0$, then $z_{lM} = 1$. We have

$$\lambda_{lM}(t) \leq z_{lM}(t) \ , \qquad (l \in \mathcal{L}, 1 \leq t \leq T) \ . \tag{3.26}$$

The above three constraints ensure that there are at most two adjacent positive $\lambda_{lm}(t), m = 0, 1, \cdots M$ for each $u_l(t)$. Now (3.20) to (3.22) can be rewritten for the entire piecewise curve as follows:

$$u_l(t) = \sum_{m=0}^{M} \lambda_{lm}(t)(U_l^{\min} + m\Delta) \ , \qquad (l \in \mathcal{L}, 1 \leq t \leq T) \ , \tag{3.27}$$

where $\lambda_{lm}(t)$ satisfies:

$$\sum_{m=0}^{M} \lambda_{lm}(t) = x_l(t) \ , \qquad (l \in \mathcal{L}, 1 \leq t \leq T) \ , \tag{3.28}$$

$$g_l(t) = \sum_{m=0}^{M} \lambda_{lm}(t) \log_2(1 + U_l^{\min} + m\Delta) \ , \qquad (l \in \mathcal{L}, 1 \leq t \leq T) \ . \tag{3.29}$$

The following lemma characterizes the approximation error ratio $\frac{h_l(t) - g_l(t)}{h_l(t)}$ as a function of M.

Lemma 4. For a given $u_l(t)$, $l \in \mathcal{L}$, the ratio of linear approximation error $\frac{h_l(t) - g_l(t)}{h_l(t)}$ is upper bounded by $\frac{U_l^{\max} - U_l^{\min}}{M \cdot \ln(2) \cdot (1 + U_l^{\min}) \log_2(1 + U_l^{\min})}$.

Proof. Since the logarithm function $h_l(t)$ becomes closer to linear as the value of $u_l(t)$ increases, the maximum approximation error occurs within the first line segment, i.e., $u_l(t) \in [U_l^{\min}, U_l^{\min} + \Delta]$. Denote K as the gradient of logarithm function at $U_l^{\min}$. Then we have $K = \frac{1}{\ln(2)(1 + U_l^{\min})}$. Denote $E_{\max}$ as the maximum approximation error in the entire range $[U_l^{\min}, U_l^{\max}]$. Denote $\mathcal{R}$ as

$[U_l^{\min}, U_l^{\min} + \Delta]$. Since the maximum approximation error occurs within the first line segment, we have

$$\begin{aligned} E_{\max} &= \max_{u_l(t)\in\mathcal{R}} \frac{h_l(t) - g_l(t)}{h_l(t)} \\ &\leq \frac{\max_{u_l(t)\in\mathcal{R}}\{h_l(t) - g_l(t)\}}{\min_{u_l(t)\in\mathcal{R}}\{h_l(t)\}} \\ &\leq \frac{K\Delta}{\min_{u_l(t)\in\mathcal{R}}\{h_l(t)\}} \\ &\leq \frac{K\Delta}{\log_2(1+U_l^{\min})} \\ &= \frac{U_l^{\max} - U_l^{\min}}{M \cdot \ln(2) \cdot (1+U_l^{\min}) \log_2(1+U_l^{\min})} , \end{aligned}$$

where the last equation follows from the fact that $\Delta = \frac{U_l^{\max} - U_l^{\min}}{M}$ and $K = \frac{1}{\ln(2)(1+U_l^{\min})}$. □

Replacing $c_l(t)$ in (3.8), which is a logarithm function, with its piecewise linear approximation ($W \cdot g_l(t)$), we have:

$$c_l(t) = W \cdot \sum_{m=0}^{M} \lambda_{lm}(t) \log_2(1 + U_l^{\min} + m\Delta) , \qquad (l \in \mathcal{L}, 1 \leq t \leq T) . \tag{3.30}$$

With the piecewise linear approximation for $c_l(t)$, we have the following linear formulation:

OPT-FD$_{\mathbf{L}}$

max $r_{\min}$

s.t $r_{\min} \leq r(f)$ $(f \in \mathcal{F})$;

Scheduling constraints: (3.1), (3.2);

Flow balance constraints: (3.3), (3.4);

Link rate constraints:

(3.7), (3.11), (3.16)-(3.19); (3.23)-(3.28); (3.30).

In this formulation, W, G, $g_{(\mathrm{Tx}(k),\,\mathrm{Rx}(l))}$, β, $n_{\mathrm{Rx}(l)}$, P, M, $U_l^{\min}$ and $U_l^{\max}$ are constants; $x_l(t)$ and $z_{lm}(t)$ are integer variables; and $c_l(t)$, $r_l(f)$, $r(f)$, $u_l(t)$, $a_{kl}(t)$ and $\lambda_{lm}(t)$ are continuous variables. The optimization problem is in the form of a mixed-integer linear program (MILP). The

theoretical worst-case complexity of solving a general MILP problem is exponential [54]. There exists highly efficient optimal algorithms (e.g., branch-and-bound [39, Chapter 5]) and heuristic algorithms (e.g., sequential fixing algorithm [39, Chapter 10]). For most practical-sized networks, a commercial solver such as CPLEX [95] may also be effective. Since the goal of this chapter is to explore the performance limit of FD multi-hop network, we will employ CPLEX to solve OPT-FD$_{\mathrm{L}}$.

The linear approximation error in Lemma 4 can be made arbitrarily small when we increase the number of linear segments M. Denote r_{L}^* as the optimal objective value of OPT-FD$_{\mathrm{L}}$. Then the following lemma specifies the value of M so that r_{L}^* is within $(1-\epsilon)$ of r^*, the objective value of OPT-FD.

Lemma 5. For a given performance gap ϵ, $0 < \epsilon \ll 1$, if $M = \lceil \frac{U_l^{\max} - U_l^{\min}}{\epsilon \cdot \ln 2 \cdot (1+U_l^{\min}) \log_2 (1+U_l^{\min})} \rceil$, then we have $(1-\epsilon) r^* \leq r_L^* \leq r^*$

Proof. We prove this lemma by two steps. First, we show that $(1-\epsilon) r^* \leq r_L^*$ holds. Since r^* is the optimal objective value of OPT-FD, its corresponding solution $\psi = \big[x_l(t),\ c_l(t),\ r_l(f),\ r(f)\big]$ is clearly a feasible solution and satisfies all constraints in OPT-FD. Based on ψ, we can construct $\hat{\psi}_L = \big[x_l(t),\ u_l(t),\ \hat{c}_l(t),\ (1-\epsilon) r_l(f),\ (1-\epsilon) r(f),\ a_{kl}(t),\ z_{lm}(t),\ \lambda_{lm}(t)\big]$. It can be verified that $\hat{\psi}_L$ is a feasible solution to OPT-FD$_{\mathrm{L}}$ and its corresponding objective value is $(1-\epsilon) r^*$. Since OPT-FD$_{\mathrm{L}}$ is a maximization problem, its optimal objective value r_L^* must be greater or equal to its feasible objective value $(1-\epsilon) r^*$. Therefore, $(1-\epsilon) r^* \leq r_L^*$ holds. Following the same token, we can show that $r_L^* \leq r^*$ holds.

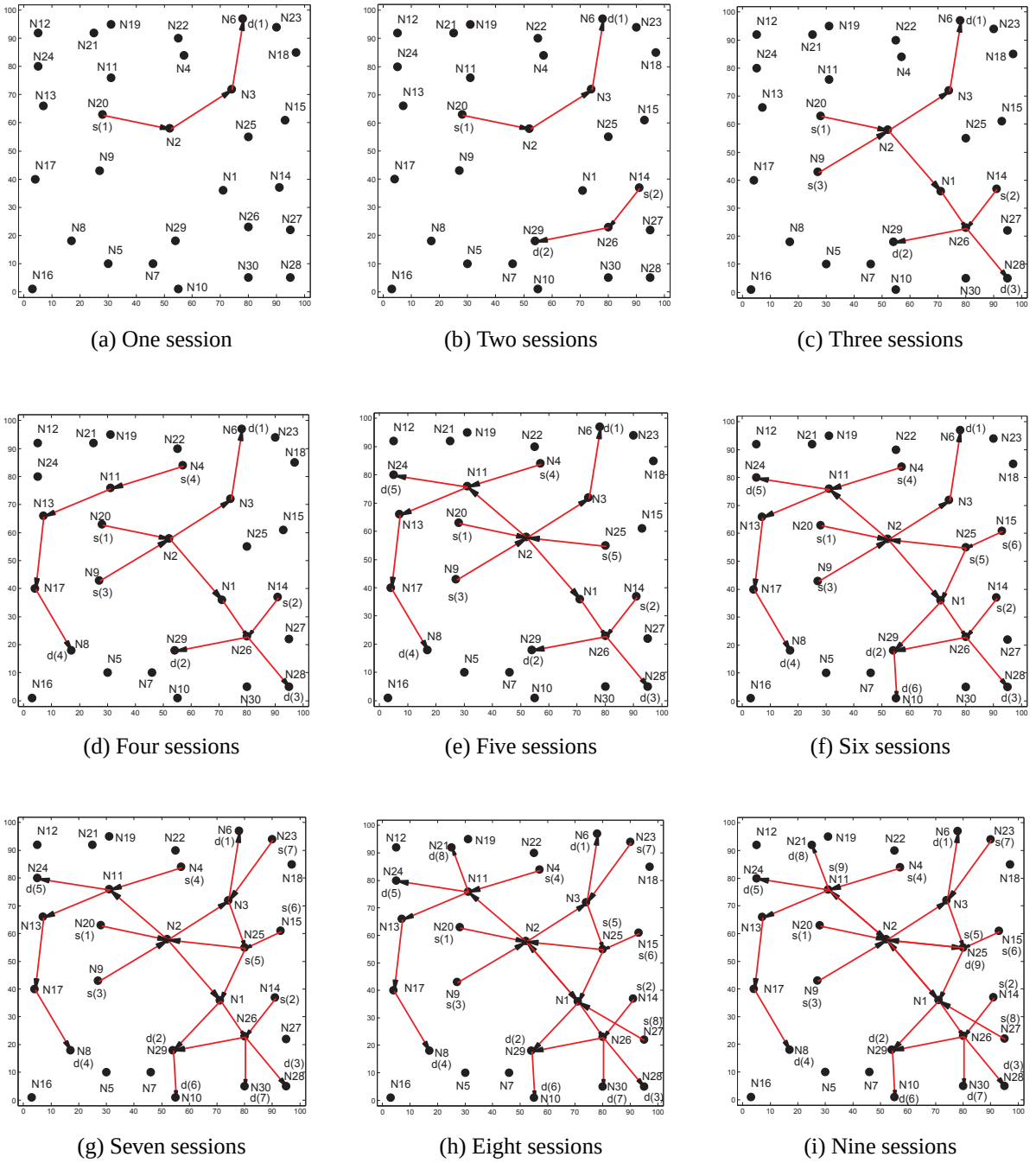
Step 1: In this step, we will show that $(1-\epsilon) r^* \leq r_L^*$. We prove this inequality holds by showing that if r^* is an optimal objective value of OPT-FD, then $(1-\epsilon) r^*$ is an achievable objective value of OPT-FD$_{\mathrm{L}}$. Suppose that r^* is an optimal objective value of OPT-FD. Then there exits a solution $\psi = \big[x_l(t),\ c_l(t),\ r_l(f),\ r(f)\big]$ that satisfies all constraints in OPT-FD. Based on ψ, we can construct a solution ψ_R to OPT-FD$_{\mathrm{R}}$ as follows: $\psi_R = \big[x_l(t),\ u_l(t),\ c_l(t),\ r_l(f),\ r(f),\ a_{kl}(t)\big]$, where $u_l(t)$ is obtained by constraint (3.9) and $a_{kl}(t) = x_k(t) u_l(t)$. Based on Lemma 1, we know that its corresponding objective value is equal to r^*. Then based on ψ_R, we can construct a solution $\hat{\psi}_L$ to OPT-FD$_{\mathrm{L}}$ as follows: $\hat{\psi}_L = \big[x_l(t),\ u_l(t),\ \hat{c}_l(t),\ (1-\epsilon) r_l(f),\ (1-$

$\epsilon)r(f),\ a_{kl}(t),\ z_{lm}(t),\ \lambda_{lm}(t)\big]$.

We now show that the constructed solution $\hat{\psi}_L$ is a feasible solution to OPT-FD$_{\rm L}$. For any given $u_l(t)$, we can always find $z_{lm}(t)$, $\lambda_{lm}(t)$ and $\hat{c}_l(t)$ by constraints (3.23)–(3.28) and (3.30). Since ψ_R satisfies constraints (3.1)-(3.4), (3.11), and (3.16)-(3.19) in OPT-FD$_{\rm R}$, it is easy to verify that $\hat{\psi}_L$ also satisfies these constraints in OPT-FD$_{\rm L}$. Based on Lemma 4, we know that we can find a M so that $\hat{c}_l(t) \geq (1-\epsilon)c_l(t)$. The corresponding $M = \lceil \frac{U_l^{\max} - U_l^{\min}}{\epsilon \cdot \ln 2 \cdot (1+U_l^{\min}) \log_2 (1+U_l^{\min})} \rceil$. Therefore, $\hat{\psi}_L$ satisfies constraint (3.7). Since $\hat{\psi}_L$ satisfies all constraints in OPT-FD$_{\rm L}$, it is a feasible solution to OPT-FD$_{\rm L}$. Since the objective function is a linear function of $r(f)$ in both OPT-FD and OPT-FD$_{\rm L}$, we know that the feasible objective value of OPT-FD$_{\rm L}$ corresponding to solution $\hat{\psi}_L$ is $(1-\epsilon)r^*$. Since OPT-FD$_{\rm L}$ is a maximization problem and its optimal objective value is r_L^*, $r_L^* \geq (1-\epsilon)r^*$ holds.

Step 2: In this step, we show that $r_L^* \leq r^*$. We show this inequality holds by arguing that if r_L^* is an optimal objective value of OPT-FD$_{\rm L}$, then r_L^* is also an achievable objective value of OPT-FD. Suppose that r_L^* is an optimal objective value of OPT-FD$_{\rm L}$. Then there exits a solution $\psi_L = \big[x_l(t),\ u_l(t),\ \tilde{c}_l(t),\ r_l(f),\ r(f),\ a_{kl}(t),\ z_{lm}(t),\ \lambda_{lm}(t)\big]$ that satisfies all constraints in OPT-FD$_{\rm L}$. Based on ψ_L, we can construct a solution $\hat{\psi}_R$ to OPT-FD$_{\rm R}$ as follows: $\hat{\psi}_R = \big[x_l(t),\ u_l(t),\ c_l(t),\ r_l(f),\ r(f),\ a_{kl}(t)\big]$, where $c_l(t)$ is obtained by constraint (3.8).

We now show that the constructed solution $\hat{\psi}_R$ is a feasible solution to OPT-FD$_{\rm R}$. Since ψ_L satisfies constraints (3.1)-(3.4), (3.11) and (3.16)–(3.19) in OPT-FD$_{\rm L}$, it is not difficult to see that $\hat{\psi}_R$ also satisfies these constraints in OPT-FD$_{\rm R}$. Based on Lemma 4, we know that the relationship between $c_l(t)$ and its piecewise linear approximation $\tilde{c}_l(t)$ is that $c_l(t) \geq \tilde{c}_l(t)$. Therefore, $\hat{\psi}_R$ satisfies constraint (3.7). Since $\hat{\psi}_R$ satisfies all constraints in OPT-FD$_{\rm R}$, it is a feasible solution to OPT-FD$_{\rm R}$ and its corresponding objective value equals to r_L^*. Based on $\hat{\psi}_R$, we can construct a solution $\hat{\psi} = \big[x_l(t),\ c_l(t),\ r_l(f),\ r(f)\big]$. Based on lemma 1, it is easy to see that $\hat{\psi}$ is a feasible solution to OPT-FD and its corresponding objective value equals to r_L^*. Since r^* is the optimal objective value of OPT-FD , $r_L^* \leq r^*$ holds. This completes the proof. □



(a) One session (b) Two sessions (c) Three sessions

(d) Four sessions (e) Five sessions (f) Six sessions

(g) Seven sessions (h) Eight sessions (i) Nine sessions

Figure 3.2: A 30-node network instance. The number of sessions in the network increases from 1 to 9.

3.5 Numerical Investigation

In this section, we conduct numerical study to compare end-to-end session throughput performance between FD and HD in multi-hop wireless networks. Our results show that the throughput ratio of FD over HD exceeds 2 in many cases. We also give insights on why $2\times$ is not the fundamental performance barrier for network level throughput.

3.5.1 Half Duplex Model

For comparison, we present mathematical modeling for a multi-hop wireless network with HD capability on each node. It is easy to see that constraints (3.3), (3.4) and (3.7) from OPT-FD still apply to HD network. But there are some new constraints for HD.

Half Duplex Constraints. Under HD, in one time slot, a node can operate in only one mode (transmit, receive, idle). Then we have

$$\sum_{k\in\mathcal{L}_i^{\mathrm{In}}} x_k(t) + \sum_{l\in\mathcal{L}_i^{\mathrm{Out}}} x_l(t) \leq 1\ , \quad (i\in\mathcal{N}, 1\leq t\leq T)\ . \tag{3.31}$$

Link Rate Constraints. The achievable capacity on link l can be calculated using Shannon's capacity formula by considering only the mutual interference (from other links). We have

$$c_l(t) = W\log_2\left(1+\frac{g_{(\mathrm{Tx}(l),\,\mathrm{Rx}(l))}\cdot P\cdot x_l(t)}{\sum_{k\in\mathcal{L}} g_{(\mathrm{Tx}(k),\,\mathrm{Rx}(l))}Px_k(t)+n_{\mathrm{Rx(l)}}}\right)\ , (l\in\mathcal{L}, 1\leq t\leq T)\ . \tag{3.32}$$

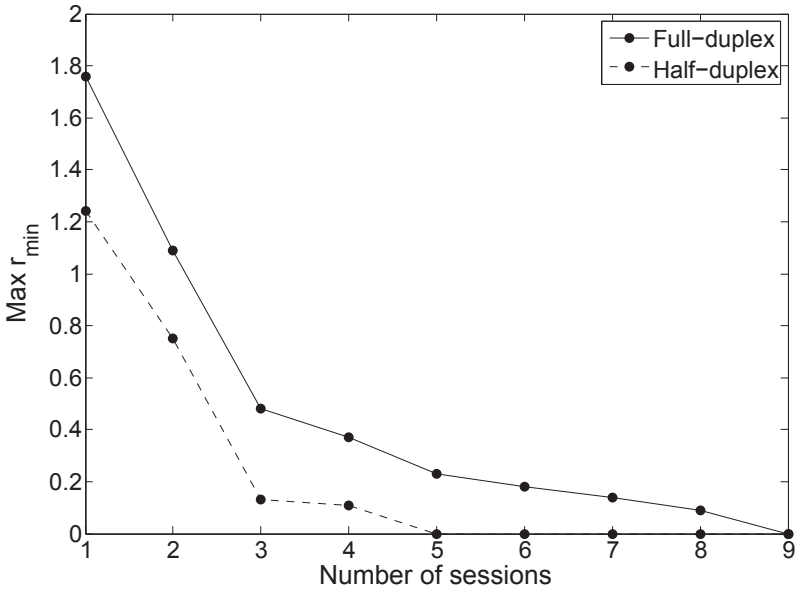
3.5.2 Simulation Setting

We consider a 30-node multi-hop wireless network, with nodes randomly deployed in a 100m $\times$ 100m area. The bandwidth is normalized to one unit (i.e., $W = 1$). The transmission power of each node is set to $P = 20dBm$. The path loss parameter $g_{(\mathrm{Tx}(k),\,\mathrm{Rx}(l))} = [20\log_{10}(d) + 38.25]$ (in dB) [99], where d is the distance between nodes $\mathrm{Tx}(k)$ and $\mathrm{Rx}(l)$. We assume the self-interference cancellation parameter $\beta = 110$dB [10] and ambient noise $n_{\mathrm{Rx}(l)} = -90$dBm. The number of time slots in a frame is $T = 4$. We set $\epsilon = 5\%$, which guarantees that the obtained solution is within 5% from the optimum.

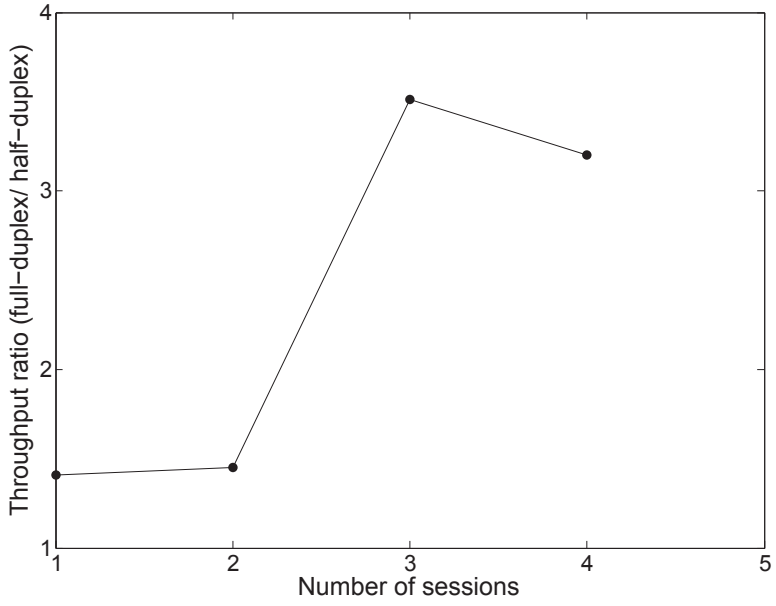
3.5.3 An Example

For the 30-node network instance in Fig. 3.2, we increase the number of sessions from 1 to 9. The source node and destination node of each session are randomly selected, with the route between the two being the shortest path route. Figure 3.3(a) shows the achievable throughput for all sessions in both FD and HD networks as the number of sessions grows. Note that the achievable throughput decreases (down to 0) under both FD and HD when the number of sessions increases. But for the same number of sessions, the objective value under FD is higher than that under HD. Figure 3.3(b) shows the ratio between the achievable throughput for FD and HD. When the number of sessions are 1, 2, 3, and 4, the ratios are 1.41, 1.45, 3.51 and 3.14, showing that FD can indeed exceed the $2\times$ performance barrier in a multi-hop network. When the number of session is 5, the throughput of HD network goes to zero, indicating that the schedulable region of HD network cannot support these five sessions (Fig. 3.2(e)). In this case, there are three sessions intersect at node N_2, with three incoming links ($N_9 \rightarrow N_2$, $N_{20} \rightarrow N_2$, $N_{25} \rightarrow N_2$) and three outgoing links ($N_2 \rightarrow N_1$, $N_2 \rightarrow N_3$, $N_2 \rightarrow N_{11}$). Since there are only four time slots in the system, HD network cannot find a feasible scheduling solution and thus renders an objective value of 0. For FD, it can offer a positive throughput until there are 9 sessions in the network (Fig. 3.2(i)). When there are 9 sessions, 5 sessions intersect at node N_2. With five incoming links ($N_1 \rightarrow N_2$, $N_9 \rightarrow N_2$, $N_{11} \rightarrow N_2$, $N_{20} \rightarrow N_2$, $N_{25} \rightarrow N_2$) and four outgoing links ($N_2 \rightarrow N_1$, $N_2 \rightarrow N_3$, $N_2 \rightarrow N_{11}$, $N_2 \rightarrow N_{25}$) at node N_2, this is beyond what can be scheduled with 4 time slots.

When the number of time slots increases (e.g., 6 or more), the schedulable regions for both FD and HD will increase. But the general shape for achievable throughput under both remains the same as that in Fig. 3.3(a), except that each curve intersects the x-axis with a higher session number. To better illustrate the idea, we increase the number of time slots to 6 for the same network instance. Fig. 3.4(a) shows the achievable throughput for all sessions in both FD and HD networks as the number of sessions grows. Compared with Fig. 3.3(a), the general shape for achievable throughput remains the same, except that the schedulable regions for HD and FD increase from 4 to 7 and 8 to 10, respectively. Note that when there are 4 time slots, the schedulable region of HD network cannot support the five sessions in Fig. 3.2(e) since there are 6 links involving node N_2. When the

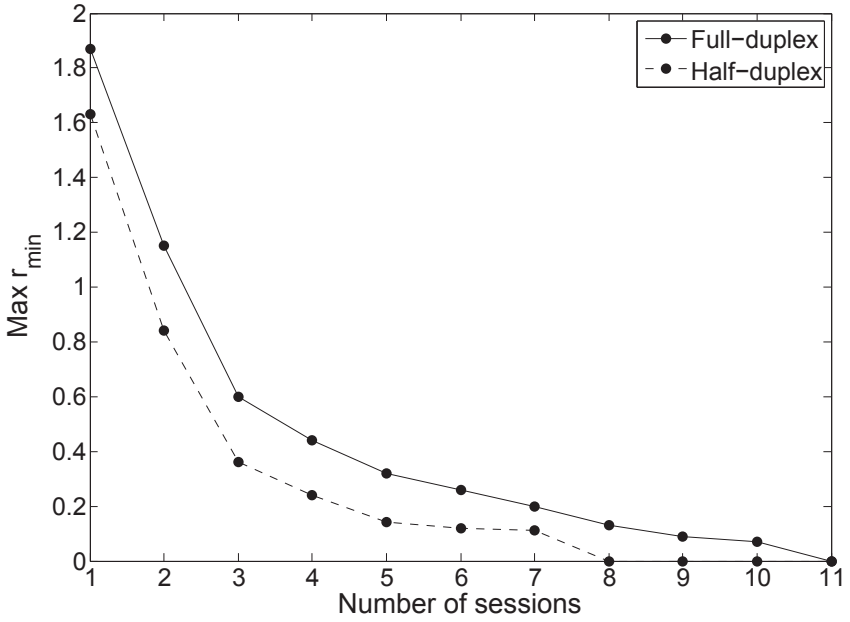


(a) Achievable throughput under FD and HD for a 30-node network as the number of sessions increases.

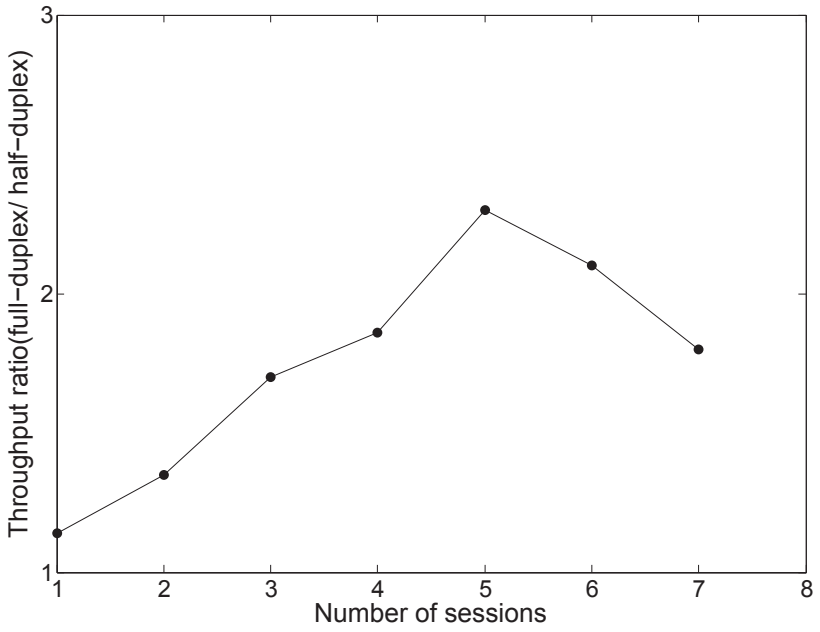


(b) Ratio between achievable throughput of FD and HD for a 30-node network as the number of sessions increases.

Figure 3.3: Throughput performance in a 30-node network with increasing number of sessions under FD and HD when $T = 4$ time slots.



(a) Achievable throughput under FD and HD for a 30-node network as the number of sessions increases.



(b) Ratio between achievable throughput of FD and HD for a 30-node network as the number of sessions increases.

Figure 3.4: Throughput performance in a 30-node network with increasing number of sessions under FD and HD when $T = 6$ time slots.

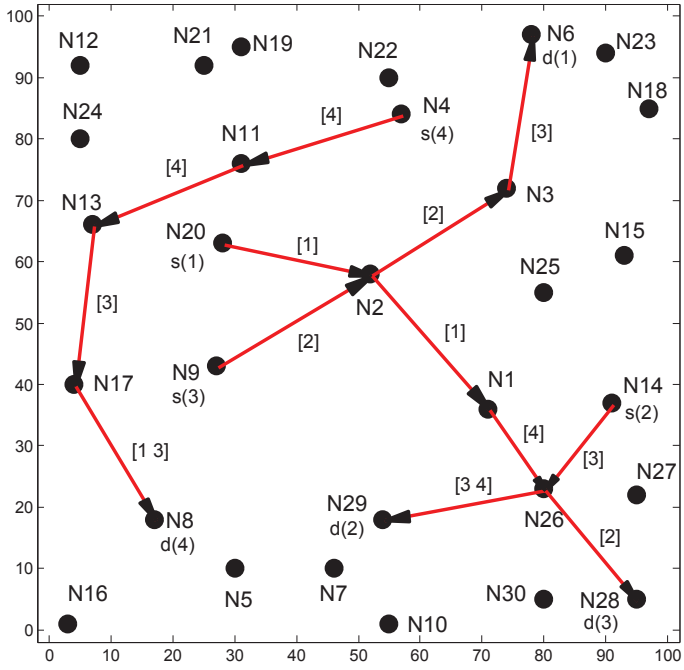
number of time slots is increased to 6, HD network can find a feasible scheduling solution among these six links and therefore has a positive throughput as shown in Fig. 3.4(a). When there are 8 sessions, the throughput of HD network goes to zero, indicating that the schedulable region of HD network cannot support these 8 sessions with 6 time slots. In this case, there are four sessions intersect at node N_2, with four incoming links ($N_1 \to N_2$, $N_9 \to N_2$, $N_{20} \to N_2$, $N_{25} \to N_2$) and three outgoing links ($N_2 \to N_1$, $N_2 \to N_3$, $N_2 \to N_{11}$) (Fig. 3.2(h)). Since there are six time slots in the system, HD network cannot find a feasible scheduling solution among these seven links and thus renders an objective value of 0. For FD, it can offer a positive throughput until there are 11 sessions in the network. Fig. 3.4(b) shows the ratio between the achievable throughput for FD and HD. When there are 5 and 6 sessions in the network, the ratio between achievable throughput of FD and HD exceeds 2.

3.5.4 A Closer Look Under the Hood

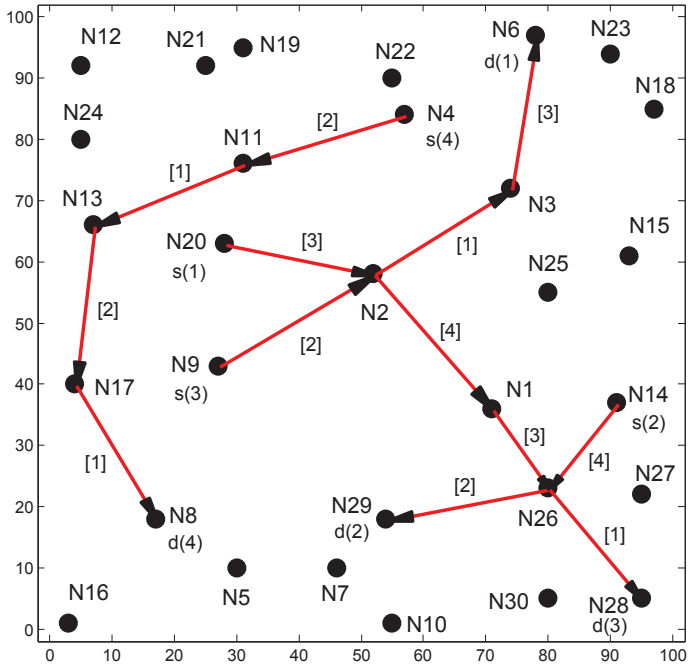
To see why $2\times$ is not a fundamental limit, let's take a closer look at the network instance in the last section when there are four sessions (Fig. 3.2(d)) in the network and $T = 4$ time slots. Figure 3.5(a) and (b) show the details of link scheduling for FD and HD, respectively. In both figures, $s(f)$ and $d(f)$ denote the source and destination nodes for session f, respectively. The number inside the bracket next to each link represents the time slot in which the link is active. Table 3.2 lists the scheduling details of each link, including the time slots in which it is active, the corresponding link capacity, and the average link capacity over T ($= 4$) time slots (i.e., $\frac{1}{T}\sum_{t=1}^{T} c_l(t)$).

We discuss HD and FD cases separately.

- *HD.* The achievable throughput of HD from our solution is 0.117. In Table 3.2, this shows that link ($N_9 \to N_2$) is the bottleneck link, which has an average throughput of 0.117. In Fig. (3.5)(b), we see that link ($N_9 \to N_2$) is active in time slot 2. Also in time slot 2, three links ($N_4 \to N_{11}$, $N_{13} \to N_{17}$, and $N_{26} \to N_{29}$) are also active and their transmitters produce strong interference to N_2. On the other hand, at node N_2, it has two incoming links and two outgoing links (for sessions 1 and 3). Due to HD, all four time slots are used by the four links involving N_2 and there is no extra time slot for bottleneck link ($N_9 \to N_2$).



(a) FD



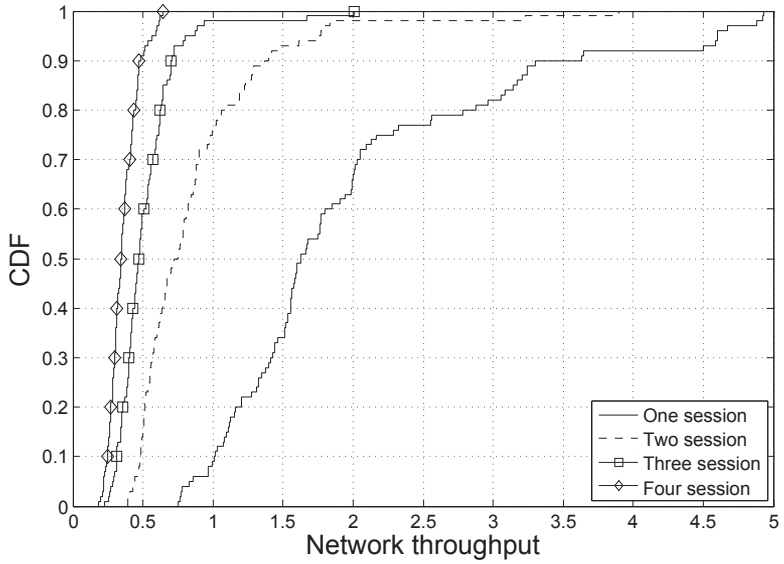
(b) HD

Figure 3.5: Link scheduling for a 30-node netowrk with 4 sessions: (a) FD, and (b) HD.

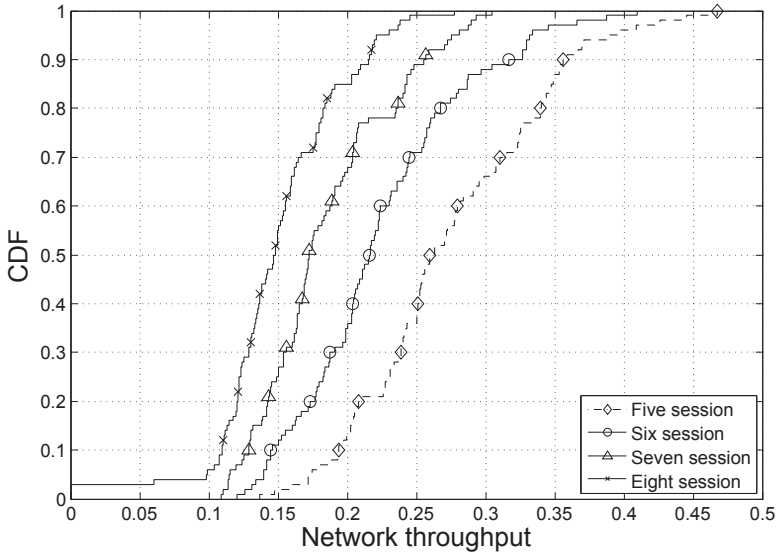
- *FD.* On the other hand, the achievable throughput of FD is 0.368. In Table 3.2, this shows that link ($N_9 \rightarrow N_2$) is again the bottleneck link, which has an average throughput of 0.368. In Fig. (3.5)(a), link ($N_9 \rightarrow N_2$) and ($N_2 \rightarrow N_3$) are both active in time slot 2 (due to FD at N_2). Except for these two links, only link ($N_{26} \rightarrow N_{28}$) is scheduled to be active in time slot 2. Therefore, there is only one interfering link for link ($N_9 \rightarrow N_2$). Note that such scheduling solution is infeasible under HD. It exploits FD capability to reduce the number of interfering links for bottleneck link $(N_9 \rightarrow N_2)$ to the minimum. Therefore, the throughput on link $(N_9 \rightarrow N_2)$ in FD can be much higher than in HD. In fact, the ratio between the achievable throughput of FD and HD is $0.368/0.117 = 3.14$, exceeding the $2\times$ barrier.

In Section 3.4.2, we employ *piece-wise linear approximation* technique to linearize the link capacity constraint (3.8). Now we present the approximation error between the original objective values and the approximated objective values. Table 3.3 lists the optimal objective values of the original problem (denoted as r^*) and linearized problem (denoted as r^*_{L}), and the normalized approximation error (i.e., $\frac{r^*-r^*_{\mathrm{L}}}{r^*}$) for both FD and HD networks. It shows that the actual approximation errors in the case study are smaller than the target approximation error $(\epsilon = 5\%)$. Therefore, the piece-wise linear approximation technique used in this chapter can indeed guarantee that the obtained objective value is within 5% from the optimum.

In summary, for end-to-end throughput of a session in a multi-hop wireless network, the advantage of FD is not limited to achieving simultaneous transmission and reception at each node. Rather, as shown in this case study, FD capability allows a much larger solution space for scheduling. Since a session's throughput at the network level is determined by the capacity of the bottleneck link along its path, a larger scheduling space will help to increase the bottleneck link rate by reducing the number of interfering links in the network. Therefore, for end-to-end throughput of a session, its throughput ratio under FD over that under HD is no longer limited by 2, as in the case for a single FD node in isolation.

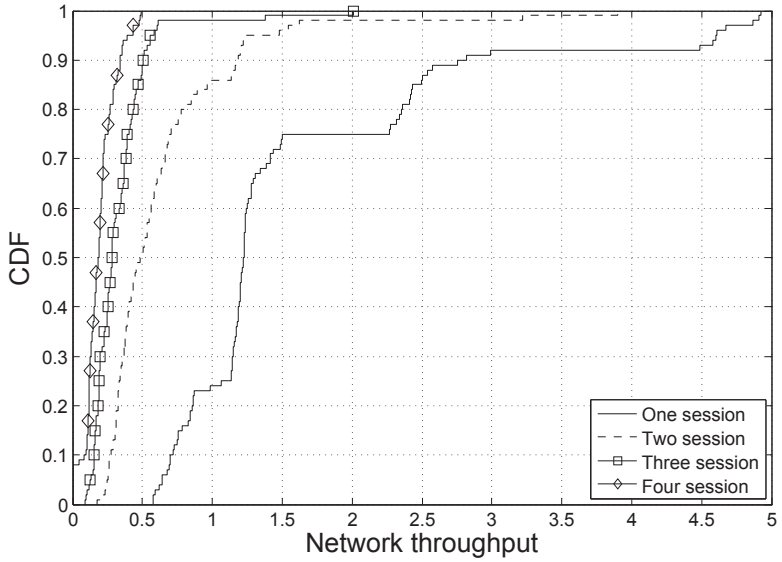


(a) The number of sessions increases from 1 to 4.

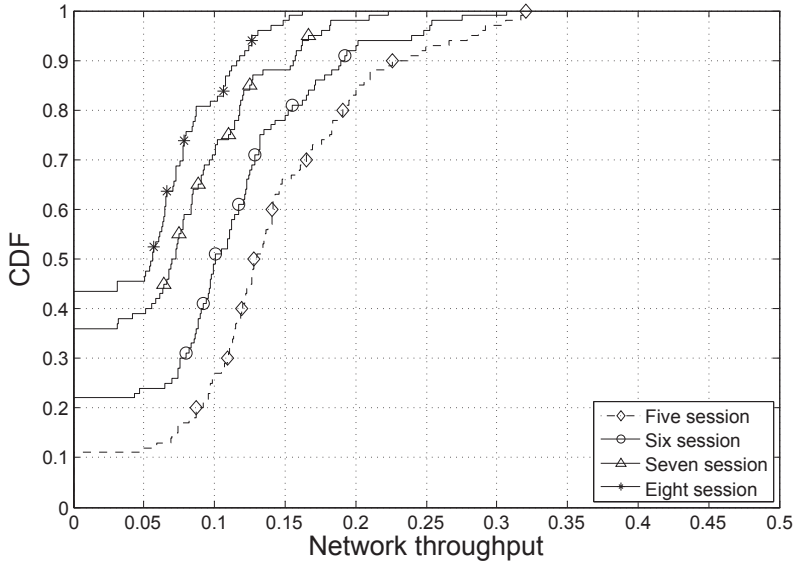


(b) The number of sessions increases from 5 to 8.

Figure 3.6: CDF of achievable throughput under FD when the number of sessions varies from 1 to 8. Each curve is based on 100 randomly generated network instances, each with 30-nodes.



(a) The number of sessions increases from 1 to 4.



(b) The number of sessions increases from 5 to 8.

Figure 3.7: CDF of achievable throughput under HD when the number of sessions varies from 1 to 8. Each curve is based on 100 randomly generated network instances, each with 30-nodes.

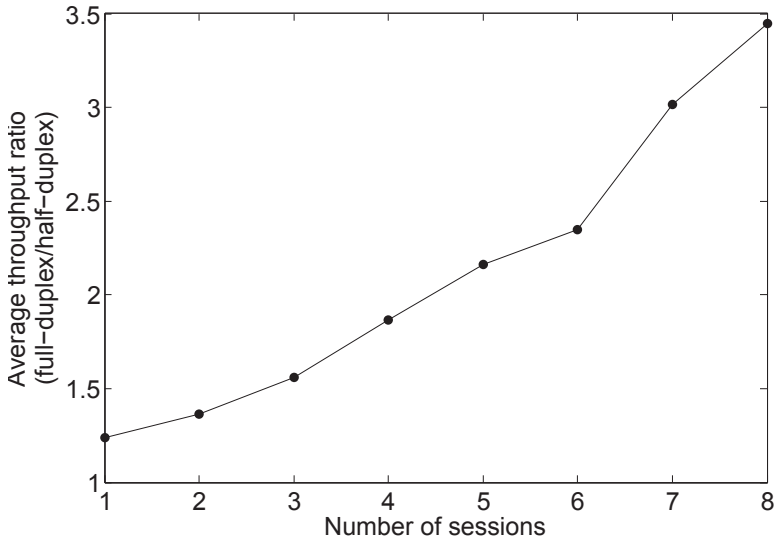


Figure 3.8: Ratio between average achievable throughput of FD and HD when the number of sessions varies from 1 to 8. The average achievable throughput (for either FD or HD) is obtained over 100 network instances, each with 30 nodes. $T = 4$ time slots.

3.5.5 Additional Results

The results in Section 3.5.3 are from one 30-node network instance with $T = 4$ time slots for increasing number of sessions. Now we perform the same study for 100 randomly generated 30-node network instances with $T = 4$ time slots. For each instance, we increase the number of sessions from 1 to 8.

Fig. 3.6 and Fig. 3.7 present the general trend of achievable throughput as the number of sessions increases from 1 to 8 under FD and HD, respectively. In both figures, we show the relative ordering relationship among the CDF curves (from 100 instances) for different number of sessions. In the case when the throughput is zero, it shows that we will have higher probability of zero throughput as the number of session increases, indicating an increasing level of difficulty to schedule the sessions with non-zero throughput.

Table. 3.4 shows the distribution of throughput ratio between FD and HD for 100 network instances as the number of sessions increases from 1 to 8. As shown in the third column, when the number of sessions increases from 2 to 8, there are 8, 27, 44, 52, 60, 66, and 71 instances that

achieve a throughput ratio ≥ 2.

Fig. 3.8 presents the ratio between the average achievable throughput of FD and HD (over 100 network instances) as the number of sessions increases from 1 to 8. It shows that the ratio between average throughput of FD and HD becomes more significant as the number of sessions increases.

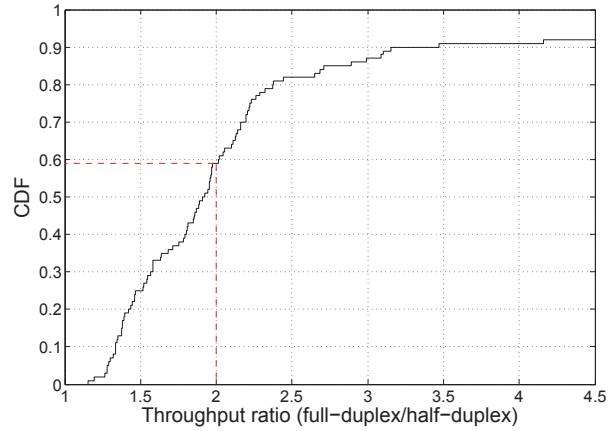
3.5.6 Impact of Various System Parameters

In Sec. 3.5.4, we present the solution details for a 30-node network instance. In Sec. 3.5.5, we perform numerical study for 100 randomly generated 30-node network instances. Through these results, we show that the achievable session throughput in a FD network can exceed $2\times$ of that in a HD network in many cases. To show the robustness of our results, we further perform numerical study under different system parameters, such as target approximation error, number of time slots, path loss parameter, and network size.

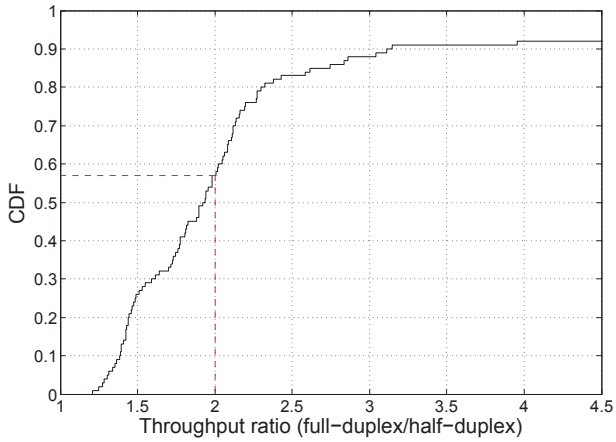
Target Approximation Error. We study the ratio between achievable throughput of FD and HD under different target approximation error. Fig. 3.9(a) to (c) present the distribution of the ratio between achievable throughput of FD and HD. Each curve is based on 100 randomly generated 30-node network instances under different estimation error (ϵ). The number of sessions in the network is set to 4 and there are 4 time slots in a time frame. As shown in the figures, when ϵ is 1%, 5% and 10%, there are 41%, 43%, and 46% of instances achieve a ratio greater than 2, respectively.

Number of Time Slots. We study the ratio between achievable throughput of FD and HD when the number of time slots is increased to 6. Fig. 3.10 presents the distribution of the ratio between achievable throughput of FD and HD based on 100 randomly generated 30-node network instances. The number of sessions in the network is set to 4. As shown in the figure, 17% of instances achieves a ratio greater than 2.

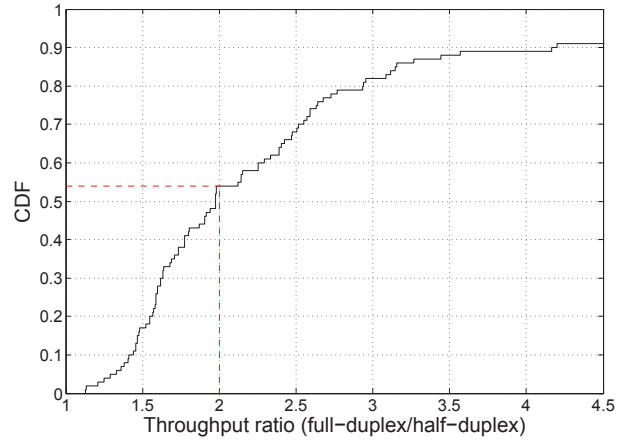
Network Size. We study the the ratio between achievable throughput of FD and HD when the number of nodes in the network is increased to 50. Fig. 3.11 presents the distribution of the ratio between achievable throughput of FD and HD based on 100 randomly generated 50-node network instances. The number of sessions in the network is set to 4 and there are 4 time slots in a time



(a) $\epsilon = 1\%$.



(b) $\epsilon = 5\%$.



(c) $\epsilon = 10\%$.

Figure 3.9: CDF of the ratio between achievable throughput of FD and HD with different target approximation error. Each curve is based on 100 network instances, each with 30 nodes and 4 sessions. $T = 4$ time slots.

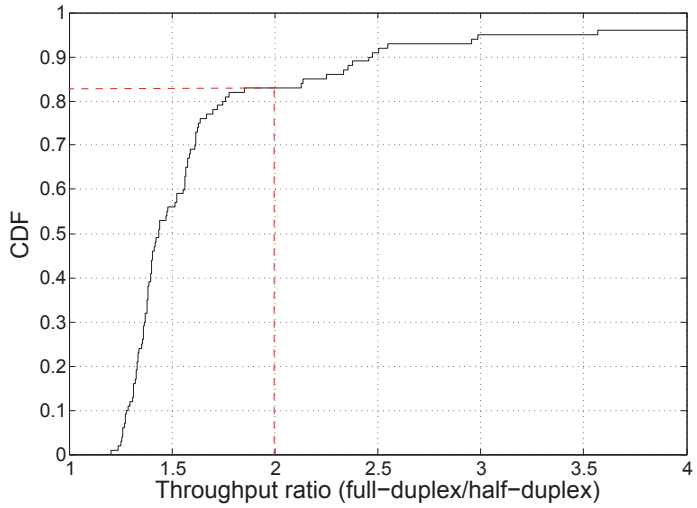


Figure 3.10: CDF of the ratio between achievable throughput of FD and HD based on 100 network instances, each with 30 nodes and 4 sessions. $T = 6$ time slots.

frame. As shown in the figure, 32% of instances achieves a ratio greater than 2.

Path Loss parameter. We study the the ratio between achievable throughput of FD and HD when the path loss parameter is set to $g_{(\text{Tx}(k),\,\text{Rx}(l))} = [16.9 \log_{10}(d) + 40.4]$ (in dB) [99]. Fig. 3.12 presents distribution of the ratio between achievable throughput of FD and HD based on 100 randomly generated 30-node network instances. The number of sessions in the network is set to 4 and there are 4 time slots in a time frame. As shown in the figure, 39% of instances achieves a ratio greater than 2

3.6 Chapter Summary

The benefits of FD for a single node in isolation were well understood. But the potential of FD on end-to-end throughput in a multi-hop wireless network was not well understood. This paper offered new results on this problem. Through rigorous mathematical modeling and formulation, we cast the end-to-end throughput problem under FD into an optimization problem. Using numerical results, we showed that session throughput in a FD network can exceed $2\times$ of that in a HD network in many cases. Our finding can be explained by the much larger scheduling space that is offered by

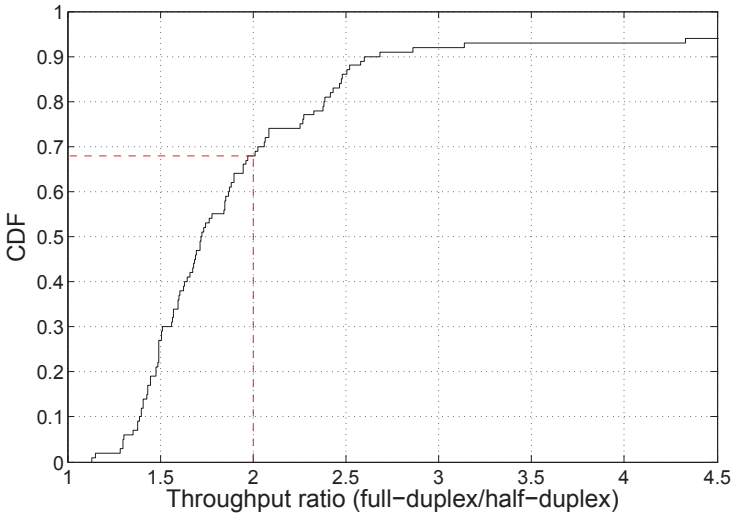


Figure 3.11: CDF of the ratio between achievable throughput of FD and HD based on 100 network instances, each with 50 nodes and 4 sessions. $T = 4$ time slots.

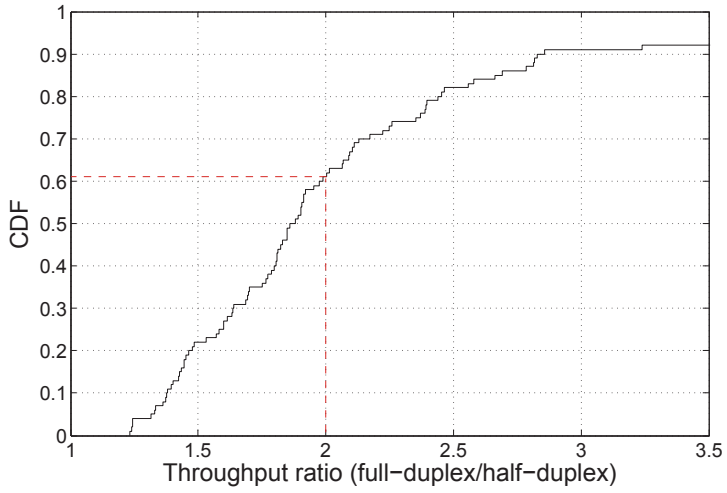


Figure 3.12: CDF of the ratio between achievable throughput of FD and HD based on 100 network instances, each with 30 nodes and 4 sessions. $T = 4$ time slots. The path loss parameter is $[16.9 \log_{10}(d) + 40.4]$.

removing HD constraints in the optimization problem. Our findings shed new light on the impact of node-level FD capability on network level end-to-end throughput performance.

Table 3.2: Details of link scheduling under FD and HD for a 30-node network with 4 sessions.

Link	FD		HD	
	(Time Slot, Link Capacity)	Average Link Capacity	(Time Slot, Link Capacity)	Average Link Capacity
$(N_{20} \rightarrow N_2)$	(1, 2.252)	0.563	(3, 0.598)	0.149
$(N_2 \rightarrow N_3)$	(2, 1.498)	0.375	(1, 0.572)	0.143
$(N_3 \rightarrow N_6)$	(3, 1.555)	0.389	(3, 1.579)	0.395
$(N_{14} \rightarrow N_{26})$	(3, 2.603)	0.651	(4, 2.728)	0.682
$(N_{26} \rightarrow N_{29})$	(3, 0.628) (4, 1.151)	0.445	(2, 1.075)	0.269
$(N_9 \rightarrow N_2)$	(2, 1.471)	0.368	(2, 0.468)	0.117
$(N_2 \rightarrow N_1)$	(1, 1.616)	0.404	(4, 0.697)	0.174
$(N_1 \rightarrow N_{26})$	(4, 2.907)	0.727	(3, 2.527)	0.632
$(N_{26} \rightarrow N_{28})$	(2, 2.349)	0.587	(1, 2.701)	0.675
$(N_4 \rightarrow N_{11})$	(4, 2.019)	0.505	(2, 0.717)	0.179
$(N_{11} \rightarrow N_{13})$	(4, 1.641)	0.410	(1, 1.087)	0.271
$(N_{13} \rightarrow N_{17})$	(3, 1.732)	0.433	(2, 1.026)	0.257
$(N_{17} \rightarrow N_8)$	(1, 1.157) (3, 1.454)	0.653	(1, 1.485)	0.371

Table 3.3: Approximation error between optimal objective values of original problem and linearized problem.

Network Type	Objective Value of Original Problem	Objective Value of Linearized Problem	Approximation Error
FD	0.374	0.368	0.017
HD	0.120	0.117	0.025

Table 3.4: Distribution of the throughput ratio between FD and HD based on 100 30-node network instances as the number of sessions increases from 1 to 8.

Session Number	Number of Network Instances with Throughput Ratio $\geq \gamma$			
	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$
1	100	0	0	0
2	100	8	1	0
3	100	27	2	1
4	100	44	12	8
5	100	52	14	11
6	100	60	24	22
7	100	66	39	37
8	100	71	48	44

Chapter 4

On AP Assignment and Transmission Scheduling for multi-AP 60 GHz WLAN

4.1 Introduction

Wi-Fi based WLAN has been widely regarded as a key technology to enable today's information technology based economy. But with the universal deployment of Wi-Fi and increasing traffic demand from existing and new applications, the radio spectrum allocated to Wi-Fi (i.e., 2.4 GHz and 5 GHz) has become overloaded. Recently, wireless communications on the mmWave have been explored to carry Wi-Fi traffic [104]. In particular, there is 7 GHz of unlicensed spectrum available at 60GHz, if utilized efficiently, can fundamentally change the spectrum shortage for Wi-Fi networks.

Comparing to 2.4 and 5GHz radio spectrum, radio propagation in the 60 GHz spectrum has some unique properties that must be carefully taken in to considerations. First, due to small wavelength, radio propagation in the 60 GHz regime suffers from significant loss in free space [66, 74, 78, 86]. To combat such severe attenuation, high-gain directional antennas at both transmitter and receiver are necessary for successful transmission [2, 35, 49]. Further, due to limited diffraction property, 60 GHz signals are vulnerable to blockage (e.g., furnitures, walls, and human body) [5, 73]. Therefore, a clear LOS path between a transmitter and a receiver is needed for

effective communication.

There have been a number of approaches that have been proposed to address the limitation of 60 GHz and to enable its application for WLAN (see Section 4.2 for more discussions). One approach that has been shown to be promising is to deploy multiple 60 GHz APs in the same area. There are a number of benefits with this approach. First and foremost, multiple APs offer more potential for a possible LOS path between a user and an AP, which can help address the blockage problem.[1] Second, multiple APs allow concurrent transmissions between users and APs (with directional transmission and reception), which can enable spatial reuse and improve network throughput.

This chapter considers multiple APs for 60 GHz WLAN and addresses two most important problems in this setting: *AP assignment* and *transmission scheduling.* When multiple APs are available, matching between APs and users is critical for network performance. When multiple transmissions can occur simultaneously in a multi-AP WLAN, transmission scheduling is needed for interference management and throughput maximization. This is particular important for an indoor environment as, unlike an outdoor open environment where interference between 60 GHz directional links is negligible [72], mutual interference between links in an indoor environement cannot be ignored due to higher user density and multiplicity of APs in a small area [37, 76]. There have been some separate studies on the AP assignment problem for 60GHz network [7, 91] and on the transmission scheduling problem for D2D communications in 60GHz WPAN [37, 57, 75, 76]. Theses prior efforts have offered some important understanding on 60 GHz wireless communications and we will review them in Section 4.2. Based on the results in the prior efforts, we believe that AP assignment and transmission scheduling are deeply intertwined together and an optimal performance for a 60 GHz multi-AP WLAN can only be achieved when they are jointly considered and optimized.

In this chapter, we study AP assignment and transmission scheduling for a multi-AP 60 GHz WLAN under a centralized control architecture. The centralized control architecture has been used in [34, 52, 55] and is shown to be feasible to coordinate multiple APs in practice. Under this

[1] In the case when a LOS does not exist between a user and any of the APs, the user will use legacy Wi-Fi bands (i.e., 2.4GHz or 5GHz) as a fallback [104].

architecture, all APs are connected to a centralized controller via high speed wired connection and the centralized controller makes the optimal AP assignment and transmission scheduling decisions based on input about active users in the network. We make the following contributions in this chapter:

- We study several AP assignment schemes of different complexity, namely per-time slot AP assignment, one-shot AP assignment, and strongest-signal AP assignment. For each AP assignment strategy, we consider downlink communication and develop an optimization problem with the objective of maximizing the minimium throughput among all users. In the cases of per-time slot AP assignment and one-shot AP assignment, the optimal AP assignment is jointly optimized with transmission scheduling in each time slot. Interference avoidance is considered as a constraint in all formulations by allowing only non-interfering links to be active simultaneously in a time slot.

- By analyzing simulation results, we find out that the strongest-signal AP assignment-based scheduling algorithm offers the worst performance. It affirms that a judicious assignment of AP to a user is critical in optimizing performance. On the other hand, per-time slot AP assignment-based scheduling algorithm only has marginal improvement in throughput performance compared with one-shot AP assignment-based scheduling algorithm. Since per-time slot AP assignment has substantially more control overhead than one-shot AP assignment, we recommend to use one-shot AP assignment-based scheduling.

- Since the one-shot AP assignment-based scheduling is an offline algorithm and requires possible AP re-assignment upon user arrival/departure, an online algorithm is needed. We design an online one-shot AP assignment by only optimizing the AP assignment for a new user while keeping the AP assingment for exisiting users intact. By analyzing the simulation results, we find that the performance of the proposed online algorithm is competitive compared with the offline algorithm. Moreover, it outperforms the strongest-signal AP assignment-based scheduling algorithm, which warrants the necessity of such an online algorithm.

The remainder of this chapter is organized as follows. In Sec. 4.2, we discuss related work to this research. In Sec. 4.3, we present a control architecture for multi-AP 60 GHz WLAN. In Sec. 4.4, we describe our throughput maximization problem for multi-AP 60 GHz WLAN. In Sec. 4.5,we present the mathematical modeling and formulation under several AP assignment strategies. In Sec. 4.6, we present numerical results and analyze the results. In Sec. 4.7, we propose an online algorithm for one-shot AP assignment and present its performance. Sec. 4.8 concludes this chapter.

4.2 Related Work

In this section, we review state-of-the-art on 60 GHz networks that is relevant to this research. These related works can be classified into three problem areas: blockage, AP assignment, and transmission scheduling.

Due to limited diffraction ability of 60 GHz signals, 60 GHz communication is vulnerable to blockage (e.g., furnitures, walls, human body) [5, 73]. To maintain network connectivity, a number of approaches has been proposed. In [4, 56, 79], a number of beam-steering algorithms have been proposed to steer the beam towards some ”good” directions so as to maximize the received signal strength at the intended receiver. To improve signal reflections from the environment for data transmission, the authors in [31, 88] propose to employ special-purpose reflectors and place them in the environment. In addition to the concern of cost in putting the relectors, the effectiveness of this approach depends heavily on a user’s position. Other than improving signal propagation for one-hop NLOS communications, another approach is to employ relay nodes and use multi-hop communications to get around the blockage [57, 73]. The effectiveness of this approach depends on the availability of relay nodes as well as the existence of LOS paths along the multi-hop communications. In a WLAN environment, to offer more LOS paths for the users, multiple APs are deployed at different positions in the same area [69, 91]. Under multiple APs, the possibility that a user has LOS path to at least one AP is increased. In the special case when a user does not have LOS to any AP, it is proposed in the 802.11ad standards [104] that the user falls back to legacy Wi-Fi bands for data communication [17, 62].

We adopt the multiple AP architecture in this research. When a user has LOS to multiple APs, AP assignment plays a critical role in throughput performance. In [91], Zhang *et al.* proposed to use multi-AP diversity for a user to overcome time-varying NLOS problem. Only a single user was considered in this study. In [7], Athanasiou *et al.* investigated AP assignment problem (called client association problem in the paper) where each user has a rate requirement and the goal is to match each user with an AP in such a way that the maximum load among all APs is minimized. The authors developed an algorithm based on lagrangian duality theory and subgradient methods to solve the problem efficiently. In this work, per time slot scheduling and potential inference among transmissions are not considered.

Transmission scheduling is necessary as only a small subset of users can be served by the APs in a time slot and potential interference must be avoided when the transmission beam of one node partially overlaps the reception beam of an unintended receiver. We consider TDMA-based scheduling since the new IEEE 802.11ad [104] and IEEE 802.15.3c [105] standards employ time division multiple access (TDMA). In [75], Son *et al.* proposed a frame-based scheduling algorithm to meet traffic demands of all users in a wireless personal area network (WPAN). The objective is to minimize the frame size. The authors proposed a graph coloring-based heuristic algorithm for tranmission scheduling for each time slot. In this work, interference among transmission pairs and blockage in the network are not considered. In a subsequent paper [37], the issues of mutual interference and blockage are considered. The authors proposed a column generation based heuristic algorithm to solve the scheduling problem for the same objective. Since the set of links in the network is pre-determined (D2D), an optimal matching between a user and its service provider (AP) is not part of the problem. In [76], Sum *et al.* proposed a scheduling algorithm for each time slot to maximize the aggregate user throughput in 60 GHz WPAN. Simultaneous transmission of interfering links is allowed if the accumulated interference at each receiver is below a threshold. Again, due to the D2D problem setting, a transmitter and its receiver is pre-determined. In [57], Niu *et al.* considered D2D communications in WPAN and explored the use of relay node to overcome blockage. They developed a scheduling algorithm by jointly considering relay selection and transmission scheduling with the objective of minimizing transmission time for a given traffic de-

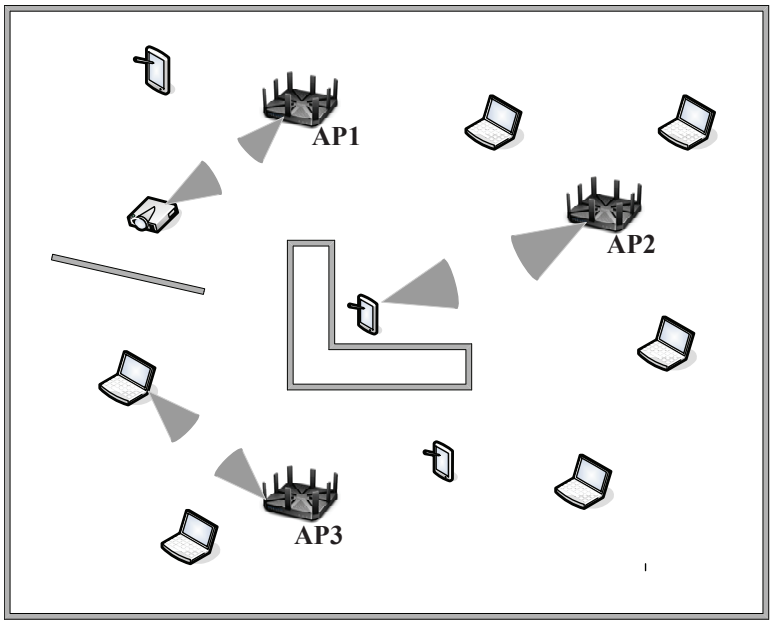


Figure 4.1: An indoor area that is being served by multiple APs.

mand. Due to significant differences in network setting (WPAN vs WLAN) and problem setting, the scheduling algorithm in [57] cannot be easily extended to address the problems that we study in this chapter.

4.3 Network Architecture

In this section, we present a system architecture for a 60 GHz WLAN with multiple APs. Consider an indoor 60 GHz WLAN as shown in Fig. 4.1, where there are multiple APs serving the users in the area. This is a typical indoor network setting for office cubicles, library, lab space, etc., where the APs are mounted on the ceilings of each floor. Given the LOS requirement for 60 GHz communications, we consider potential obstacles such as cubical walls, furniture, etc., as shown in Fig. 4.1. An AP can communicate with a user only when there exists a LOS path between them. As a result, a user may not be able to use 60 GHz communications to all APs in the area. In the case when 60 GHZ communications is not feasible for a user (due to the absence of LOS), the user can alway fall back to communicate with an AP on a lower frequency band (i.e., 2.4 GHz or 5

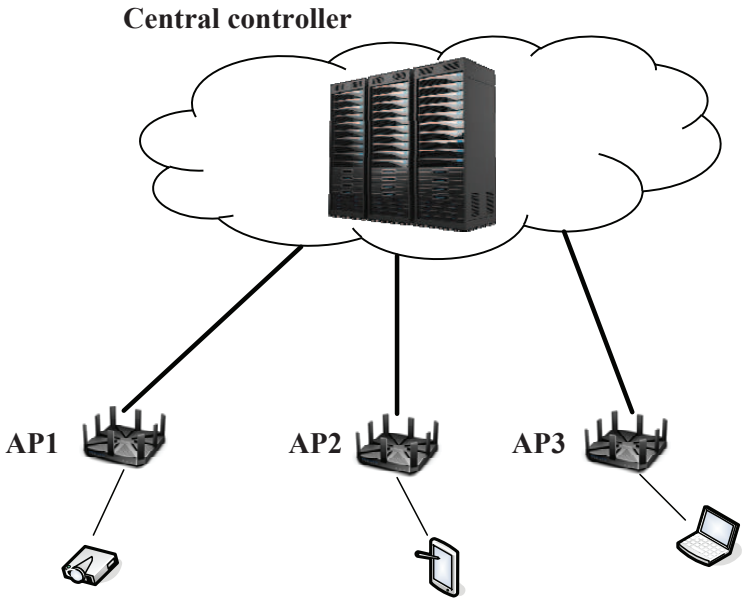


Figure 4.2: A centralized architecture to control the operation of multiple APs in the area.

GHz), which is being implemented in a multi-spectrum 60 GHz AP (see [104]). Therefore, in this chapter, we only consider those users who can access (with LOS) at least one AP in the area.

On the control plane, we employ a centralized controller in the back-end which has direct high speed connection with each of the APs in the area (see Fig. 4.1). Such centralized control of APs is becoming feasible and popular in recent years (see [34, 52, 55]) due to wide deployment of 10 GHz Ethernet as well as HPC I/O optimized servers. With the introduction of 100 GHz Ethernet and the emergence of petascale computing, it is expected that many existing applications buit on distributed computing may migrate into centralized cloud-based computing and benefit significantly from much higher computational speed.

During initial AP discovery phase, an AP transmits multiple beacon frames, each on a different sector to cover all possible directions. Then users that have LOS path toward the AP can receive the beacon and respond [104]. Subsequently, each AP (user) will have the knowledge of the LOS users (APs) for 60 GHz communications. Since the up-to-date 802.11ad standard for 60 GHz communication adopted TDMA MAC protocol for data transmission [104], we consider a time-slotted system. Under a centralized control architecture, it is not difficult for the central controller

to synchronize the clocks of all APs and users in the network. The central controller will perform all AP assignment and scheduling decisions, based on the input from the APs. Upon finding an optimal AP assignment and scheduling, the central controller will convey its solution to all APs, who will then notify all users in the area. We assume the physical (PHY) layer behavior is programmable (software-defined), each AP and user can be configured to the desired waveform on demand based on the decision from the central server.

4.4 Problem Statement

As discussed in Section 4.3, we will only study scheduling problems for those users who have LOS to at least one AP. There are some unique challenges for scheduling transmission in the 60 GHz regime.

First, due to strong absorption by atmospheric oxygen, propagation of radio signal in the 60 GHz range suffers significant greater loss than lower frequency ranges (e.g., 2.4 GHz and 5 GHz) and must rely on highly directional transmission (instead of omni-directional transmission) for effective communications. A clear LOS is also required at such wavelength to avoid blockage. Therefore, within each time slot, a scheduling solution must perform a matching between an AP and one of its LOS user.

Second, despite the directionality between 60 GHz transmitter and receiver, interference remains a concern in scheduling and must be addressed by a scheduling solution. This is because that a transmitter's beam may partially collide (overlap) an unintended receiver's reception beam towards another transmitter. Such interference, depending on the amount of beam alignment and overlap, may severely degrade the throughput of the unintended receiver. Therefore, a scheduling solution must ensure that matching between the APs and the users are free of interference.

Finally, a scheduling solution should be designed and optimized for some performance objective, in addition to offering an interference-free matching between the APs and the users in each time slot. In this chapter, we focus on throughput maximization. In each time slot, we can only schedule a small number of transmissions between the APs and their LOS users. So a user will

only be able to transmit in a subset of time slots in a frame and thus its throughput is taken as the transmitted data averaged over the total number of time slots. Therefore, a meaningful objective in our throughput maximization problem is to maximize the minimum average rate among all users (i.e., max-min).[2]

In this chapter, we explore scheduling algorithm under the following AP assignment strategy:

- Per-time slot AP assignment: This is the most complex (and best performing) approach. Under this approach, the scheduler decide the optimal matching between an AP a user in each time slot for transmission. A user may be matched to different AP in different time slot.

- One-shot AP assignment: As the name suggests, the "optimal" matching between a user and an AP is done in one-shot and their matching is permanent. That is, a user will communicate to the same AP during its communication and the user cannot choose a different AP in different time slot. This approach is simpler than the per-time slot AP assignment and is likely to incur some compromise in performance. So the question is how much performance fall-off it will have (comparing to the per-time slot approach).

- Strongest-signal AP assignment: This is a special case of the one-shot AP assignment, where each user is matched to the AP with the strongest signal among all the APs. Unlike the one-shot AP assignment, where the matching between a user and an AP requires the central controller to solve a one-shot optimization problem, the strongest-signal AP assignment only requires measurement and comparison of signal strengths between a user and the APs. Here, the scheduling problem only involves solving the transmission pattern (which set of users should be active) in each time slot. We will use this approach as the worst case benchmark.

[2]Other objectives such as maximizing the sum of log of average rates may also be considered following the same approach in this chapter.

Table 4.1: Notation in Chapter 4

Symbol	Definition
$\mathcal{M}$	Set of APs in the network
$\mathcal{N}$	Set of users in the network
$\mathcal{U}_i$	Set of LOS users within the transmission range of AP $i \in \mathcal{M}$
$\mathcal{A}_j$	Set of LOS APs within the transmission range of user $j \in \mathcal{N}$
T	Total number of time slots in a time frame
$x_{ij}(t)$	A binary variable to indicate whether or not AP $i \in \mathcal{M}$ transmits to user $j \in D_i^{\text{usr}}$ in time slot t
y_{ij}	A binary variable to indicate whether or not AP i is assigned to user j interferes with the reception at user k from AP m
c_{ij}	Achievable data rate from AP $i \in \mathcal{M}$ to user $j \in D_i^{\text{usr}}$
W	Channel bandwidth
P_t	Transmit power
P_{ij}	Received power at user $j \in \mathcal{N}$ from AP $i \in A_j^{\text{ap}}$
G_t	Antenna gain of transmit antenna
G_r	Antenna gain of receive antenna
n	Path loss exponent
N_0	Noise power spectral density
λ	Wavelength
d_{ij}	Distance between AP i and user j
r_{ij}	Data rate of flow from AP i to user j
r_j	Total data rate at user j
$r_{\min}$	The minimum data rate among all users in the network
ϕ	Antenna beamwidth
α_{ij}^k	Direction of the transmission beam from AP i to user j relative to user k
β_{km}^i	Direction of the reception beam at user k towards AP m relative to AP i

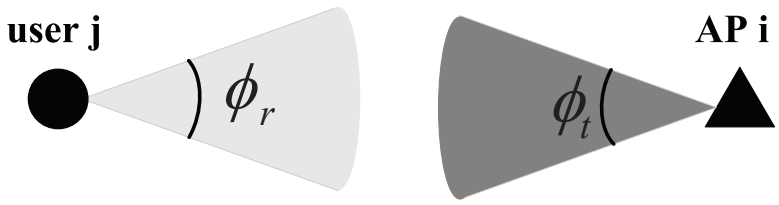


Figure 4.3: 2D radiation pattern for ideal flat-top antenna.

4.5 Mathematical Modeling and Problem Formulation

To understand how the three scheduling algorithms discussed in the last section compare with each other, we study a user rate maximization problem and develop mathematical model and problem formulation under each scheduling algorithm for the same problem. Denote $\mathcal{M}$ as the set of APs in the area and $M = |\mathcal{M}|$ is the number of APs. Denote $\mathcal{N}$ as the set of users in the network and $N = |\mathcal{N}|$ is the number of users. To model potential blockage between a user and the set of APs $\mathcal{M}$, we randomly generate blockage in the network so that each user j only has LOS path to a subset of APs $\mathcal{A}_j$, i.e., $\mathcal{A}_j \subseteq \mathcal{M}$. Thus, user j can communicate with AP i only when $i \in \mathcal{A}_j$. On the AP side, denote $\mathcal{U}_i$ as the subset of users to which AP i has LOS path. For our design of scheduling algorithms, we employ time-slot based scheduling as in [104], with T time slots in a frame. Table 4.1 lists notation in this chapter.

4.5.1 Per-time Slot AP Assignment

AP Selection Constraints We assume that a user can select at most one of its LOS APs for communication in each time slot. To model this scheduling behavior, we define binary variable $x_{ij}(t)$ to indicate whether or not AP i transmits to user j in time slot t. That is, $x_{ij}(t) = 1$ if AP i transmits to user j, and 0 otherwise. Under per-time slot AP assignment, a user can be served by different APs in different time slots. Then we have

$$\sum_{i \in \mathcal{A}_j} x_{ij}(t) \leq 1 \ , \quad (j \in \mathcal{N}, 1 \leq t \leq T) \ . \tag{4.1}$$

Transmission Scheduling Constraints Likewise, in each time slot, we assume that an AP can

transmit to at most one of its LOS users. Then we have:

$$\sum_{j \in \mathcal{U}_i} x_{ij}(t) \leq 1 \ , \quad (i \in \mathcal{M}, 1 \leq t \leq T) \ . \tag{4.2}$$

Interference Constraints For 60 GHz communication, to compensate the high path loss, we employ directional antennas at both the transmitter and receiver. One commonly used antenna model for directional antenna is the flat-top model [72], which has constant gain within its beam width and zero gain outside. As shown in Fig. 4.3, the dark gray beam (with a beamwidth of ϕ_t) represents the transmission beam from AP i; the light gray beam (with a beamwidth of ϕ_r) represents the reception beam from user j. The antenna gains are assumed to be constants inside the transmission and reception beams, and zero outside both beams.

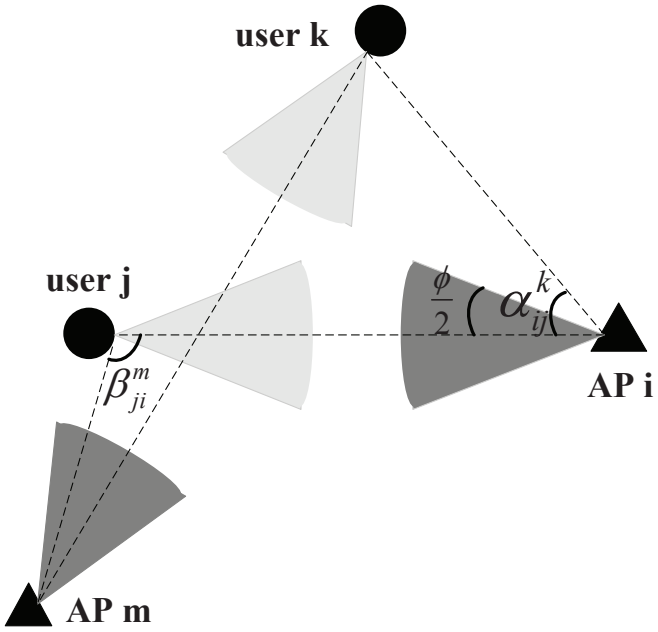
In 3-dimension, the flat-top antenna is assumed to be symmetric about its beam-axis. So the horizontal and vertical beam widths are equal. Denote G_t as antenna gain of the transmitter, then it can be approximated as [51]:

$$G_t = \frac{40000}{\triangle \phi_t^2}. \tag{4.3}$$

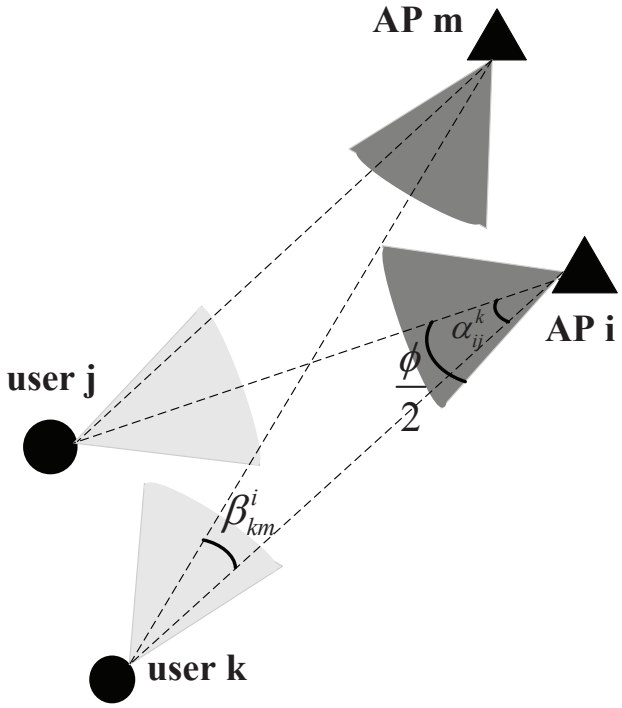
For a transmitter with a beamwidth of 30°, its antenna gain is about 16.5 dBi. Similarly, the antenna gain of a receiver (G_r) can be calculated by $G_r = 40000/\phi_r^2$.

Under the ideal flat-top antenna model that has constant transmit/receive gain within its beamwidth and zero gain elsewhere, interference exists only when the receiver and interferer are located within each other's boresight. That is, we say that AP i interferes with an unintended user k *if and only if* the following conditions are satisfied: (i) user k is within the boresight of AP i, *and* (ii) AP i is within the boresight of user k. On the other hand, we say that AP i does not interfere with user k if (i) user k is outside the boresight of AP i, *or* (ii) AP i is outside the boresight of user k. To illustrate such inteference relationships, consider the two examples in Fig. 4.4. In both examples, AP i transmits to user j and AP m transmits to user k. To simplify our discussion, we assume the beamwidths at transmit nodes (AP i and AP m) and receive nodes (user j and user k) are all ϕ.

- In Fig. 4.4(a), α_{ij}^k represents the angle between AP i to an unintended receiver k relative to the line connecting AP i and its intended receiver j. Likewise, β_{ji}^m represents the angle



(a) Two non-interfering links.



(b) Two interfering links.

Figure 4.4: An example illustrating interference relationship under directional transmission and directional reception.

between receiever j to an unintended transmitter AP m relative to the line connecting receiver j and its intended transmitter AP i. As shown in the figure, AP i does not interfere with user k since user k is outside boresight of AP i ($\alpha_{ij}^k > \frac{\phi}{2}$). Likewise, AP m does not interfere with user j since, even user j is within the boresight of transmitter m, AP m is still outside boresight of node j's reception beam (i.e., $\beta_{ji}^m > \frac{\phi}{2}$). Therefore, these two links are not interfering with each other and can be activated in the same time slot.

- In Fig. 4.4(b), AP i interferes with user k since (i) user k is within the boresight of AP i ($\alpha_{ij}^k \leq \frac{\phi}{2}$); and (ii) AP i is within the boresight of user k ($\beta_{km}^i \leq \frac{\phi}{2}$). Likewise, AP m's transmission also interferes with non-intended receiver j. Therefore, these two links should not be activated in the same time slot.

To model the interference relationship, we introduce two indicator variables, one from the transmit beam side and one from the receive beam side. Define $I(\alpha_{ij}^k)$ as a binary varible to indicate whether or not unindented receiver k is within the boresight of AP i when AP i transmits to user j. That is,

$$I(\alpha_{ij}^k) = \begin{cases} 1 & (\alpha_{ij}^k) \leq \frac{\phi}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, define $I(\beta_{km}^i)$ as a binary variable to indicate whether or not AP i is within the boresight of user k when user k receives from AP m. That is, $I(\beta_{km}^i) = 1$ if $\beta_{km}^i \leq \frac{\phi}{2}$ and 0 otherwise. By definition, AP i interferes with user k if and only if (i) user k is within the boresight of AP i, *and* (ii) AP i is within the boresight of user k. Then we have AP i interferes with user k if and only if $I(\alpha_{ij}^k)I(\beta_{km}^i) = 1$. Otherwise, AP i does not interfere with user k if and only if $I(\alpha_{ij}^k)I(\beta_{km}^i) = 0$.

To ensure interference-free scheduling between any two transmissions in the area (i.e., two interfering transmissions do not occur in the same time slot), we must have:

$$x_{ij}(t) + I(\alpha_{ij}^k)I(\beta_{km}^i)x_{mk}(t) \leq 1\ , \\ (i \in \mathcal{M}, j \in \mathcal{U}_i, k \in \mathcal{N}, m \in \mathcal{A}_k, m \neq i, k \neq j, 1 \leq t \leq T)\ . \tag{4.4}$$

That is, when $I(\alpha_{ij}^k)I(\beta_{km}^i) = 1$, only one link (either AP i to user j or AP m to user k is active), and when $I(\alpha_{ij}^k)I(\beta_{km}^i) = 0$, both links can be active in the same time slot.

Data Rate Constraints Denote c_{ij} as the achievable data flow rate from AP i to its LOS user $j \in \mathcal{N}$. Then we have:

$$c_{ij} = W \log_2 \left(1 + \frac{P_{ij}}{N_0 W}\right), \quad (i \in \mathcal{M}, j \in \mathcal{U}_i, 1 \leq t \leq T), \tag{4.5}$$

where W is the bandwidth, N_0 is noise power spectral density, and P_{ij} is power of directional transmission and can be obtained by the Friis transmission equation as follows:

$$P_{ij} = P_t G_t G_r (\frac{\lambda}{4\pi})^2 (d_{ij})^{-n}, \quad (i \in \mathcal{M}, j \in \mathcal{U}_i, 1 \leq t \leq T). \tag{4.6}$$

In (4.6), P_t is the transmission power, G_t and G_r are the antenna gain of diretional transmit and receive antennas in (4.3), respectively, λ is the wavelength, d_{ij} is the distance between AP i to user j, and n is the path loss exponent.

Under a time-slotted scheduling system, transmission from AP i to user j may be active only on a subset of time slots in a frame. So the average transmission rate from AP i and user j can be calculated as follows:

$$r_{ij} = \frac{1}{T} \sum_{t=1}^{T} c_{ij} x_{ij}(t), \quad (i \in \mathcal{M}, j \in \mathcal{U}_i, 1 \leq t \leq T). \tag{4.7}$$

Under per-time slot AP assignment, user j may be served by different APs in different time slots. So its data rate from all APs is,

$$r_j = \sum_{i \in \mathcal{A}_j} r_{ij}, \quad (j \in \mathcal{N}, 1 \leq t \leq T). \tag{4.8}$$

Problem Formulation Denote $r_{\min}$ as the minimum rate among all users in the network, then our problem can be formulated as follows:

OPT-P

max $r_{\min}$

s.t $r_{\min} \leq r_j \qquad (j \in \mathcal{N})$;

AP selection constraints: (4.1) ;

Transmission scheduling constraints: (4.2) ;

Interference constraints : (4.4) ;

Data rate constraints: (4.7), (4.8) .

In this formulation, $x_{ij}(t)$ are binary variables, r_{ij}, r_j and $r_{\min}$ are continuous variables. α_{ij}^k, β_{km}^i and c_{ij} are constants. We assume that all the APs and users in the network employ the same beamwidth ϕ, therefore, ϕ is a constant. Although the optimization problem is in the form of a mixed-integer linear program (MILP), a commercial solver such as CPLEX can solve such problem for the scale of a WLAN.

4.5.2 One-shot AP Assignment

Different from per-time slot AP assignment, under one-shot AP assignment scheme, the matching between a user and an AP is permanent and does not change on a per-time slot basis. This approach is much easier to implement in practice and has significant benefit on the upper layer as well (e.g., maintaining stability in a communication session).

Denote y_{ij} as a binary variable to indicate whether user j choose AP i for communication, i.e., $y_{ij} = 1$ if user j choose AP i for communication and 0 otherwise. Since user j is permanently assigned to one AP, we have

$$\sum_{i \in \mathcal{A}_j} y_{ij} \leq 1 \ , \quad (j \in \mathcal{N}) \ . \tag{4.9}$$

Even if $y_{ij} = 1$ (i.e., user j is assigned to AP i for communication), such transmission may not be active in each time slot t. We have

$$x_{ij}[t] \leq y_{ij} \ , \quad (i \in \mathcal{M}, j \in \mathcal{U}_i, 1 \leq t \leq T) \ . \tag{4.10}$$

That is, when $y_{ij} = 1$, $x_{ij}[t]$ may be 1 (active transmisson) or 0 (idle); when $y_{ij} = 0$ (i.e., user j is not assigned to AP i), then $x_{ij}[t]$ must be 0.

Based on these updates, the problem formulation under one-shot AP assignment can be put forth as follows:

OPT-O

max $r_{\min}$

s.t $r_{\min} \leq r_j \qquad (j \in \mathcal{N})$;

AP selection constraints: (4.9) ;

Transmission Scheduling constraints: (4.2), (4.10) ;

Interference constraints : (4.4) ;

Data rate constraints: (4.7), (4.8) .

In this formulation, $x_{ij}(t)$ and y_{ij} are binary variables, r_{ij}, r_j and $r_{\min}$ are continuous variables. α_{ij}^k, β_{km}^i and c_{ij} are constants. We assume that all the APs and users in the network employ the same beamwidth ϕ, therefore, ϕ is a constant. Although the optimization problem is in the form of a mixed-integer linear program (MILP), a commercial solver such as CPLEX can solve such problem for the scale of a WLAN.

4.5.3 Strongest-signal AP Assignment

The strongest-signal AP assignment scheme is a special case of the one-shot AP assignment and is what is typically done in the real world. Instead of finding the "optimal" matching between a user and an AP through global optimization, each user is matched to the AP which has the strongest signal. That is, the AP assignment variable y_{ij} in OPT-O is a pre-assigned constant here. More specifically, y_{ij} is 1 if AP i has the strongest signal to user j among all the APs, and 0 otherwise. Therefore, the scheduling problem here does not involve AP assignment, as in per-time slot AP assignment and one-shot AP assignment problems. Instead, it only needs to determine whether or not a user should be active in each time slot. Then the problem formulation under strongest-signal AP assignment can be put forth as follows:

OPT-S

max $r_{\min}$

s.t $r_{\min} \leq r_j \qquad (j \in \mathcal{N})$;

Transmission scheduling constraints: (4.2), (4.10) ;

Interference constraints : (4.4) ;

Data rate constraints: (4.7), (4.8) .

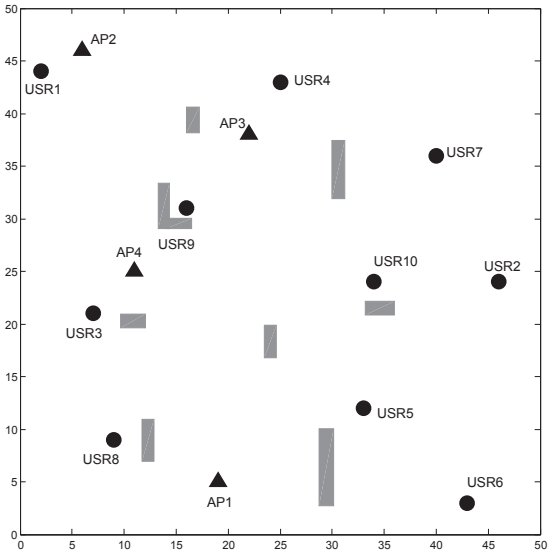
In this formulation, $x_{ij}(t)$ are binary variables, r_j and $r_{\min}$ are continuous variables. y_{ij}, α_{ij}^k, β_{km}^i and c_{ij} are constants. Again, we assume that all the APs and users in the network employ the same beamwidth ϕ, which is a constant.

4.6 Performance Evaluation

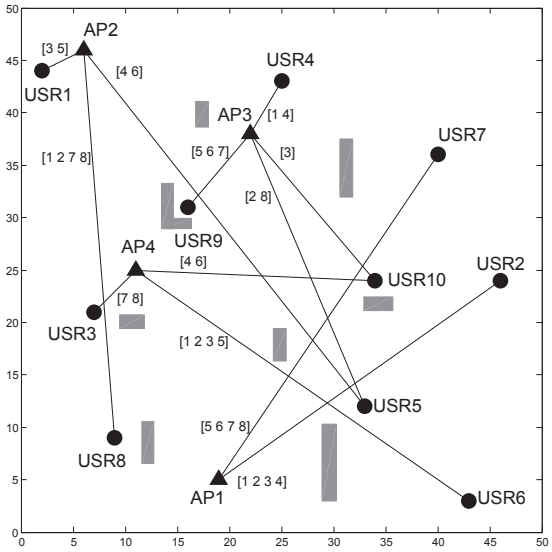
In this section, we compare the performance of three different algorithms for AP assignment and transmission scheduling under the same throughput objective outlined in the last section. We find out that the strongest-signal AP assignment-based scheduling algorithm offers the worst performance, confirming that such a naive (despite simple) approach is not plausible. On the other hand, both per-time slot AP assignment and one-shot AP assignment based scheduling algorithms have significant advantage over the strongest-signal AP assignment. This shows that some judicious matching between AP and user is critical in throughput performance. We also find that per-time slot AP assignment based scheduling does not have much advantage over one-shot AP assignment based scheduling.

4.6.1 Simulation Setting

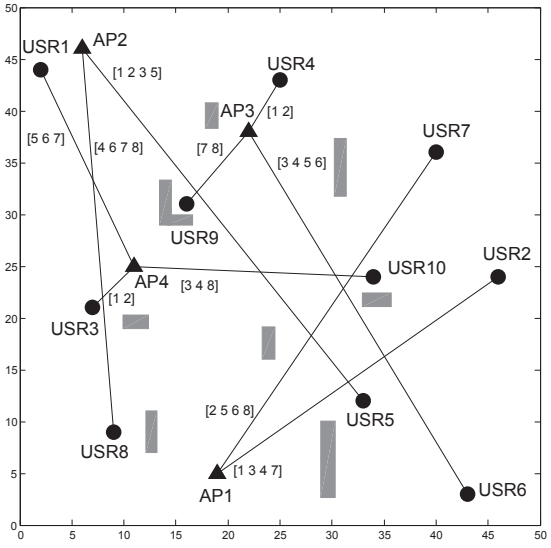
We consider an indoor environment where multiple 60 GHz APs are randomly deployed in a 50m $\times$ 50m area. We assume that the transmission range of each AP can cover this entire area. Following the parameters in [104], the transmission power at AP is 10 dBm, the bandwidth is 2.16 GHz, and the noise spectral density N_0 is -134 dbm/MHZ. The path loss exponent n is 2.3 [74]. We assume the beamwidth of all the APs and users are 30 degrees.



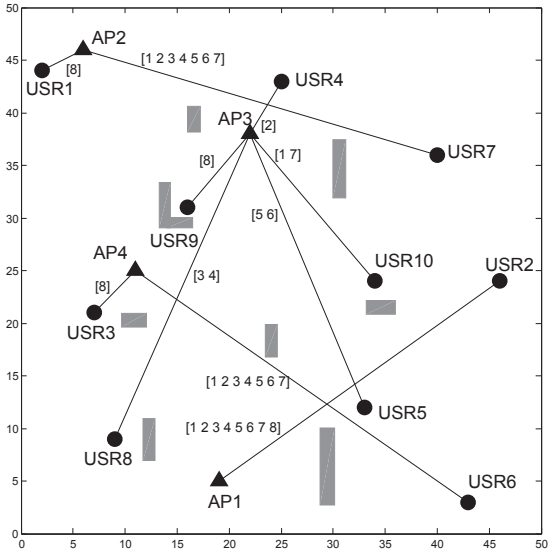
(a) Network typology.



(b) Per-time slot AP assignment.



(c) One-shot AP assignment.



(d) Strongest-signal AP assignment.

Figure 4.5: A network instance with 4 APs and 10 users.

4.6.2 A Case Study

Before we present complete results in the following section, we present a case study to show the details that we are interested in. In this case study, we have 4 APs and 10 users as shown in Fig. 4.5(a). In the figure, the grey blocks are randomly generated and represent the potential blockage in this area. Therefore, each user in the network may only have LOS access to a subset of APs. For example, user 1 has LOS path to AP 2 and AP 4 while does not have LOS path to AP 1 and AP 3 due to blockage.

Suppose there are 8 time slots in a time frame. By solving the three optimization problems in the last section, we can find the AP assignment and transmission scheduling solution to each problem. Figures 4.5(a), (b), and (c) show the solution details under per-time slot AP assignment, one-shot AP assignment, and strongest-signal AP assignment based scheduling algorithms, respectively. The number inside the bracket next to each link represents the time slot in which the link is active. Table 4.2 lists the scheduling details for each user under the three AP assignment schemes, including the time slots in which it is active, to which AP it is assigned to, and its average rate over a frame. Note that under per-time slot AP assignment, a user may be assigned to different AP in different time slot. This is shown by user 5 and user 10, where user 5 is assigned to AP 2 in time slots 4 and 6 and is assigned to AP 3 in time slots 2 and 8; user 10 is assigned to AP 3 in time slot 3 and is assigned to AP 4 in time slots 4 and 6. The average rate for each user under each algorithm is calculated based on (4.5), (4.7), and (4.8) in the optimal solutions. Note the significant differences in transmission scheduling (time slot), AP assignment, and throughput for each user under the three difference algorthms. The minimum user rates under per-time slot AP assignment, one-shot AP assignment, and strongest-signal AP assignment are 6.7 Gbps, 6.0 Gbps, and 2.9 Gbps, respectively. Now we look into the details in each figure.

- Figure 4.5(b) shows the AP assignment and scheduling details under per-time slot AP assignment. Under this scheme, the optimal solution decides the optimal matching between an AP and a user for each time slot. A user can be matched to different AP in different time slots. As shown in the figure, in different time slots, user 5 is matched to AP 2 and AP 3, while user 10 is matched to AP 3 and AP 4. Under 30 degrees directional transmission and

reception beams, the transmit beam from AP 2 to user 5 interferes with the reception beam at user 10 when it receives from AP 3. Therefore, these two links must not be scheduled in the same time slots. As shown in Table 4.2, the transmissions from AP 2 to user 5 are in time slots 4 and 6, while transmission from AP 3 to user 10 is in time slot 3. The minimum achievable average rate among all users is 6.7 Gbps, which is user 5.

- Figure 4.5(c) shows the AP assignment and scheduling details under one-shot AP assignment scheme. Under this scheme, the optimal solution decides the matching between a user and an AP and this matching is permanent in all time slots. As shown in the figure, each user is matched to only one AP. In contrast to the previous case, user 5 is matched to AP 2 while user 10 is matched to AP 4. The minimum achievable average throughput is 6.0 Gbps, which is again on user 5. To see why the rate for user 5 is smaller than previous case, we look into the details in Table 4.2. Under per-time slot AP assignment, user 5 is matched to AP 2 and AP 3 for 2 time slots each; under one-shot AP assignment, user 5 is matched to AP 2 only and is active for 4 time slots. Although the total number of time slots for transmission is the same, there is a difference in throughput when user 5 is matched to AP 2 and AP3 (AP3 is closer to user 5 than AP 2, as shown in Fig. 4.5(b)). By being permanently matched to one AP in all time slots, the average rate of user 5 becomes smaller.

- Figure 4.5(d) shows the scheduling details under strongest-signal AP assignment scheme. Under this scheme, a user is matched to an AP with the strongest signal among all its LOS APs. Hence, AP assignment is not part of the optimization problem, as it is in the one-shot AP assignment problem. AP assignment is done before we perform optimal scheduling in each time slot. The biggest potential problem with strongest-signal AP assignment is that it may overload certain AP when many users are assigned to this AP because they find this AP offers the strongest signal. As shown in Figure 4.5(d), the number of users assigned to each AP is unbalanced. Here, AP 1 serves only one user, AP 2 and AP 4 each serves two users, while AP 3 has to serve 5 users. Due to this unbalance in AP assignment, users assigned to AP 3 have fewer number of time slots in scheduling comparing to users assigned to other APs. As a result, the throughput for these users could be much smaller. As shown

in Table 4.2, the minimum achievable throughput is 2.9 Gbps, which is on user 9. Note that user 9 is among a group of five users (4, 5, 8, 9 and 10) that are assigned to AP 3. Among these five users, users 5, 8, 10 are given 2 time slots, while user 4 is closer to the AP than user 9. Therefore, user 9 has the smallest rate.

In this case study, we find that both per-time slot AP assignment and one-shot AP assignment have significant advantage over strongest-signal AP assignment in terms of our throughput objective, while there is only marginal improvement of per-time slot AP assignment over one-shot AP assignment.

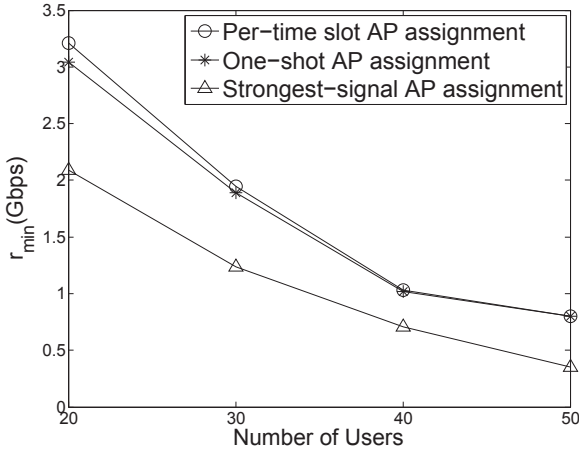
4.6.3 Complete Results

In this section, we offer extensive results for the three AP assignment and transmission scheduling schemes to substantiate the observations that we had in the case study. For each comparison study, we generate 50 random network instances and take the average from the results. To simulate blockage between an AP and a user, we randomly generate a binary number b_{ij} for each AP i and user j. If $b_{ij} = 1$, we will have a LOS path between AP i and user j; otherwise ($b_{ij} = 0$), there is no LOS between the two. In our comparison study, we vary the number of users, the number of APs, and the number of time slots in a frame.

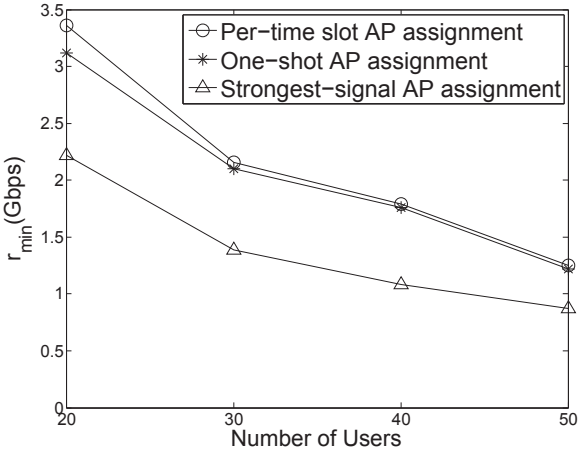
Varying Number of Users. We compare the objective values obtained by the three AP assignment and transmission scheduling schemes for different number of users in the area. We consider a network with 4 APs and increase the number of users in the network from 20 to 50. Figure 4.6(a)–(c) show the trend of objective values as the number of users increases from 20 to 50 when there are 16, 24 and 32 time slots in a time frame, respectively. Each point on the curve is averaged over results from 50 randomly generated network instances. As shown in the figures, the objective values obtained by all three schemes decreases as the number of users in the network increases, as expected. It is easy to observe that (i) both per-time slot and one shot AP assignment-based scheduling schemes significantly outperform the strongest-signal AP assignment based-scheduling scheme; and (ii) the benefits of per-time slot AP assignment scheme over one-shot AP assignment scheme is really marginal. Table 4.3 shows the ratio between the objective values of two schemes.

Table 4.2: Details of Scheduling under three AP assignment schemes for a network with 4 APs and 10 users.

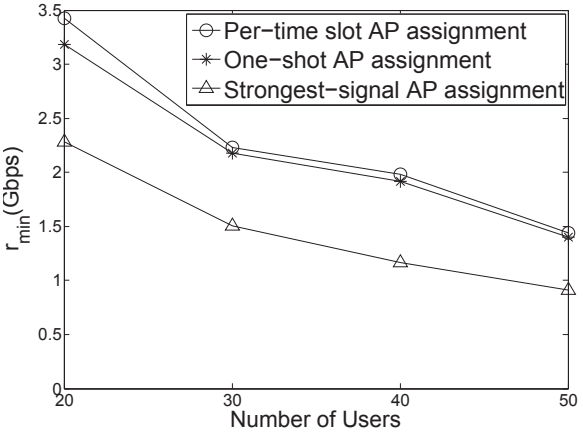
User	Per-time slot AP Assignment			One-shot AP Assignment			Strongest-signal AP Assignment		
	Timeslot	AP	Average rate (Gbps)	Timeslot	AP	Average rate (Gbps)	Timeslot	AP	Average rate (Gbps)
1	(3,5)	2	7.3	(5,6,7)	4	6.8	8	2	3.6
2	(1,2,3,4)	1	7.4	(1,3,4,7)	1	7.4	(1,2,3,4,5,6,7)	1	14.8
3	(7,8)	4	6.9	(1,2)	4	6.9	8	4	3.4
4	(1,4)	3	6.8	(1,2)	3	6.8	2	3	3.4
5	(4,6)	2	6.7	(1,2,3,5)	2	6.0	(5,6)	3	3.9
	(2,8)	3							
6	(1,2,3,5)	4	6.8	(3,4,5,6)	3	6.6	(1,2,3,4,5,6,7)	4	11.9
7	(5,6,7,8)	1	6.9	(2,5,6,8)	1	6.9	(1,2,3,4,5,6,7)	2	12.5
8	(1,2,7,8)	2	7.0	(4,6,7,8)	2	7.0	(3,4)	3	3.7
9	(5,6,7)	3	8.9	(7,8)	3	6.1	8	3	2.9
10	3	3	7.2	(3,4,8)	4	6.5	(1,7)	3	4.7
	(4,6)	4							



(a) 16 time slots.

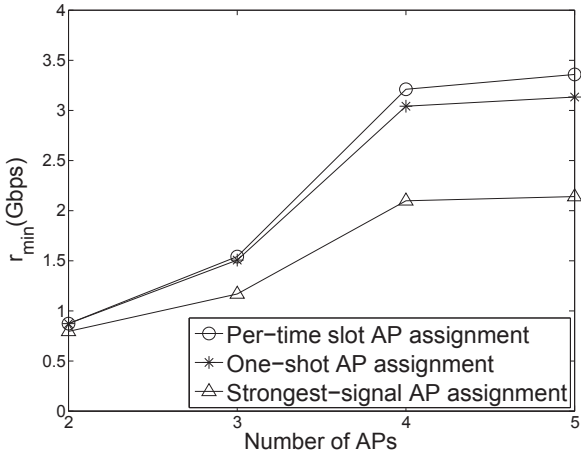


(b) 24 time slots.

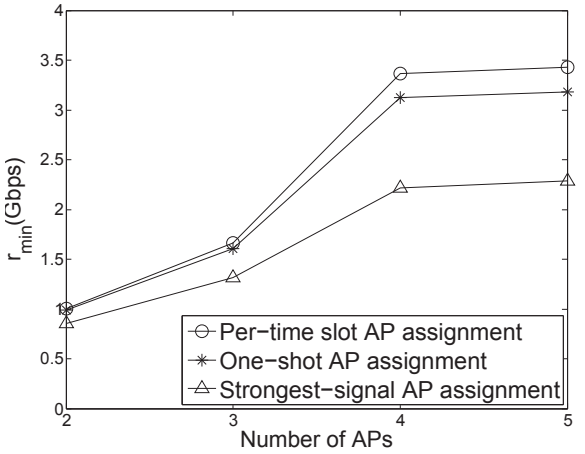


(c) 32 time slots.

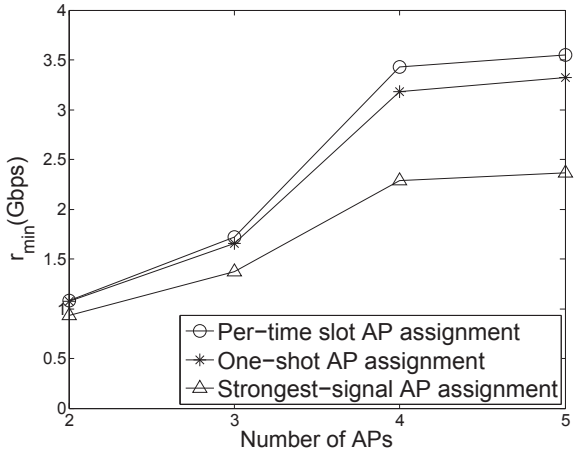
Figure 4.6: Maximum guaranteed user rate when the number of users increases from 20 to 50. Each curve is based on 50 randomly generated network instances, each with 4APs.



(a) 16 time slots.

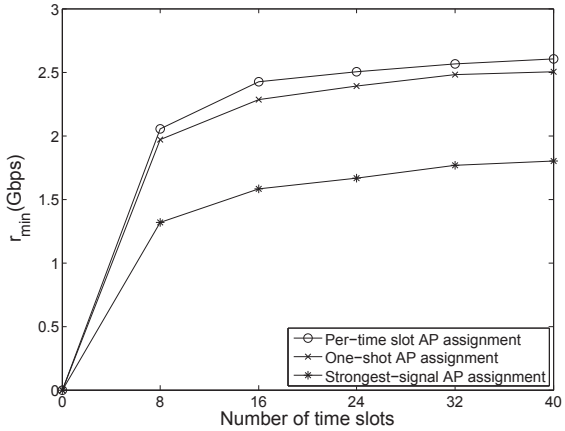


(b) 24 time slots.

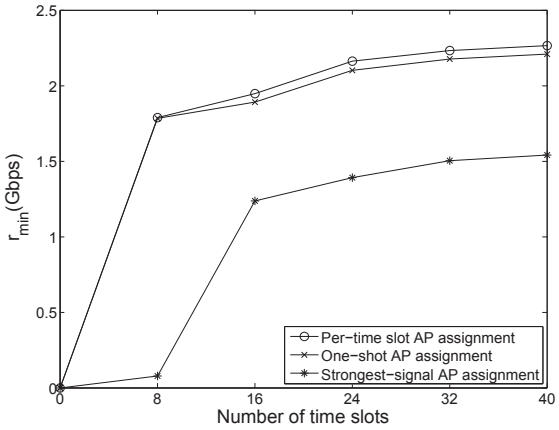


(c) 32 time slots.

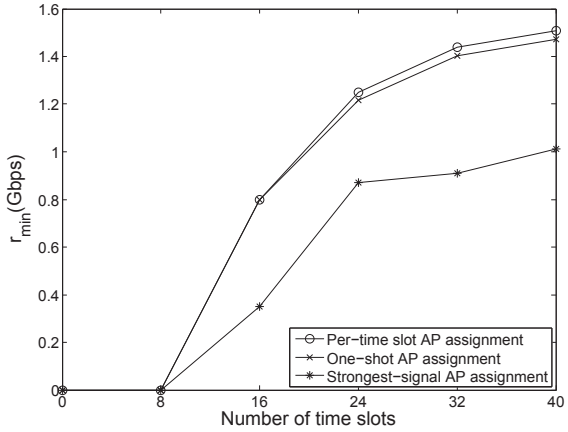
Figure 4.7: Maximum guaranteed user rate when the number of APs increases from 2 to 5. Each curve is based on 50 randomly generated network instances, each with 20 users.



(a) 20 users.



(b) 30 users.



(c) 50 users.

Figure 4.8: Maximum guaranteed user rate when the number of time slots increases from 0 to 40. Each curve is based on 50 randomly generated network instances, each with 4APs.

Table 4.3: Ratio between objective values obtained by per-time slot AP assignment and those by one-shot AP assignment for different number of users.

Number of Users	Ratio between Two Objectives		
	Time Slot = 16	Time Slot = 24	Time Slot = 32
20	1.06	1.08	1.08
30	1.03	1.03	1.04
40	1.01	1.02	1.03
50	1	1.02	1.02

Table 4.4: Ratio between objective values obtained by per-time slot AP assignment and those by one-shot AP assignment for different number of APs.

Number of APs	Ratio between Two Objectives		
	Time Slot = 16	Time Slot = 24	Time Slot = 32
2	1	1.02	1.01
3	1.03	1.03	1.04
4	1.06	1.08	1.08
5	1.07	1.08	1.08

Varying Number of APs. We now compare the objective values obtained by the three AP assignment and transmission scheduling schemes under different number of APs. We consider a network with 20 users and increase the number of APs in the network from 2 to 5. Figure 4.7(a)–(c) show the results when there are 16, 24 and 32 time slots in a time frame, respectively. Again, we have the same observations. In particular, Table 4.4 shows the ratio between the objective values for the per-time slot and one-shot AP assignment based transmission scheduling schemes. The difference between the two schemes is marginal.

Table 4.5: Ratio between objective values obtained by per-time slot AP assignment and those by one-shot AP assignment for different number of APs.

Number of Time Slots	Ratio between Two Objectives		
	Number of users = 20	Number of users = 30	Number of users = 50
8	1.03	1	1
16	1.06	1.03	1.02
24	1.08	1.03	1.02
32	1.08	1.04	1.02
40	1.09	1.04	1.03

Varying Number of Time slots. Finally, we compare the objective values under the three schemes for different time frame length. We consider a network with 4 APs and increase the number of time slots in a frame from 0 to 40. Figure 4.8(a) –(c) show the results when there are 20, 30, and 50 users, respectively. Note that for 50 users (Figure 4.8(c), 8 time slots are not sufficient to schedule transmission of all 50 users. Also, the conclusion from the three figures are consistent to our earlier observations. In particular, Table 4.5 shows the ratio between the objective values for the per-time slot AP assignment and one-shot AP assignment based transmission scheduling schemes. The difference between the two schemes is marginal.

4.7 An Online Algorithm for Practical Implementation

4.7.1 Motivation

The results in the last section show that per-time slot AP assignment only has marginal improvement in throughput over one-shot AP assignment. On the other hand, per-time slot AP assignment has substantially more control overhead than one-shot AP assignment. This makes it more attractive to employ one-shot AP assignment in practice. But still, there is more work needs to be done for practical implementation. The one-shot AP assignment based scheduling discussed in

Section 4.5.2 takes the instance of the current network as input. Then it formulates an optimization problem based on this instance and offers an optimal solution for AP assignment and transmission scheduling in each time slot. When a new user arrives the network, or an exisiting user departs the network, the instance of the network also changes. Then it is necessary to formulate a new optimization problem and find a new optimal solution for AP assignment and transmission schedule. That is, although the AP assignment does not need to change on a per-time slot basis , it may still need change per user arrival/departure. Again this could be disruptive and bring additional control overhead across multiple layers.

We are interested in an online algorithm that only performs one-time optimization of AP assignment for a new user and then binds this assignment permenently for this user throughout the life of its communication session (i.e., until it departs). This online algorithm will enjoy some potential benefits that come with one-shot AP assignment while avoiding the overhead in switching AP for each user arraival/departure event. For transmission scheduling, its optimal solution will change whenever a new user arrives and an existing user departs. But such change is not as disruptive as changing a user's AP, especially given that modern software-based radio is capable of changing its waveform and scheduling behavior on a per packet basis.

4.7.2 Algorithm Details

The design of our online algorithm for one-shot AP assignment and transmission scheduling is centered around solving two optimization problems – one for new user arrival and one for exisiting user departure. We decribed some details as follows.

Arrival of A New User

When a new user m enters the network, it can detect the subset of APs that it can communicate with (LOS) via the beacon frames sent periodically by these APs. Upon its response, the new user's LOS APs will add user m to their LOS user sets and report to the central controller. The central controller will add user m to the existing user set by updating $\mathcal{N}$ (i.e., $\mathcal{N} \leftarrow \mathcal{N} \cup \{m\}$), and solves the following new optimization problem (OPT-Arrival).

OPT-Arrival

max $r_{\min}$

s.t $r_{\min} \leq r_j$ ($j \in \mathcal{N}$) ;

AP selection constraint only for the new user m: $\sum_{i \in \mathcal{A}_m} y_{im} \leq 1$;

Transmission scheduling constraints for all users: (4.2), (4.10) ;

Interference constraints for all users: (4.4) ;

Data rate constraints for all users: (4.7), (4.8) .

In this formulation, α_{ij}^k, β_{km}^i, c_{ij} and y_{ij} ($j \in \mathcal{N} \backslash \{m\}$) are constants, while $x_{ij}(t)$, y_{im}, r_{ij}, r_j and $r_{\min}$ are optimization variables. Note that the AP assignment for existing users, y_{ij} where $j \in \mathcal{N} \backslash \{m\}$, are constants. Only AP assignment for new user m, y_{im}, is a variable. For transmission scheduling, $x_{ij}(t)$, are variables for all users, including new user m.

Upon the central controller solves the optimization problem, the new user is assigned to its optimal AP and each AP adjusts its transmission scheduling to the users following the new optimal solution.

Departure of An Existing User

When an existing user k leaves the network, the user sends a termination message to its AP, which relays the message to the central controller. Upon receiving this message, the central controller will remove user k from the user set $\mathcal{N}$, i.e., $\mathcal{N} = \mathcal{N} \backslash \{k\}$. Then it solves the following optimization problem (for optimal transmission scheduling) for the remaing users. Note that the AP assignment for the remaining users are not changed.

OPT-Departure

max $r_{\min}$

s.t $r_{\min} \leq r_j$ ($j \in \mathcal{N}$) ;

Transmission scheduling constraints for all users: (4.2), (4.10) ;

Interference constraints for all users: (4.4) ;

Data rate constraints for all users: (4.7), (4.8) .

In this formulation, y_{ij}, α_{ij}^{k}, β_{km}^{i} and c_{ij} are constants, and $x_{ij}(t)$, r_{ij}, r_j and $r_{\min}$ are variables. Upon solving the optimization problem, each AP adjusts its transmission scheduling to the remaining users following the new optimal solution..

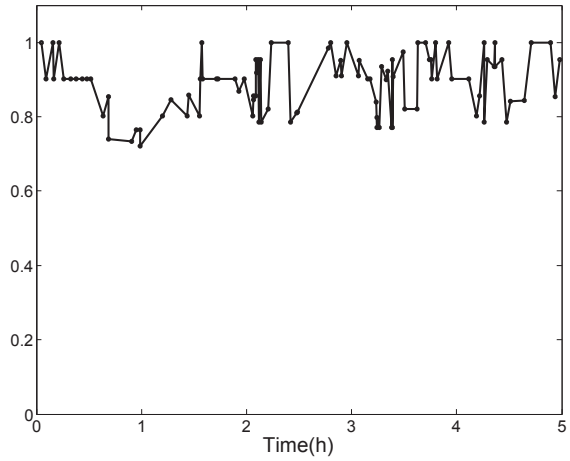
4.7.3 Performance Evaluation

To evaluate the performance of the online algorithm, we compare the objective values in real-time for different network instances (with arrivals and departures) under the proposed online algorithm and those under the offline algorithm (one-shot AP assignment). We also compare our proposed online algorithm with the strongest-signal AP assignment algorithm. The latter comparison is to ensure there is merit in optimizing AP assignment even for an online algorithm.

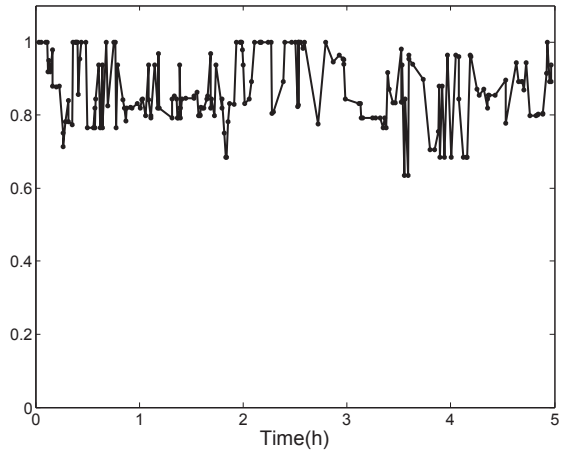
In the comparison study, we use 4 APs and assume there are 20 time slots in a frame. We consider different new user arrival rates (Poisson), ranging from 10, 20, and 30 per hour. The holding time for each user is exponential with an average of 1 hour.

Figure 4.9(a), (b), and (c) show the ratio between the objective values obtained by our proposed online algorithm and those by the offline algorithm based on one-shot AP assignment when user arrival rates are 10, 20, and 30 per hour, respectively. In Fig. 4.9(a), there are a total of 97 events during the 5-hour simulation time, among which there are 82 events with ratio over 80%, 59 events with ratio over 90%. The average ratio for all instances is 88.54%. In Fig. 4.9(b), there are a total of 209 events during the 5-hour simulation time, among which there are 149 events with ratio over 80%, 74 events with ratio over 90%. The average ratio for all instances is 85.87%. In Fig. 4.9(c), there are a total of 305 events during 5 hours, among which there are 217 events with ratio over 80%, 118 events with ratio over 90%. The average ratio for all instances is 84.71%. These results show that our proposed online algorithm is competitive compared to the offline algorithm.

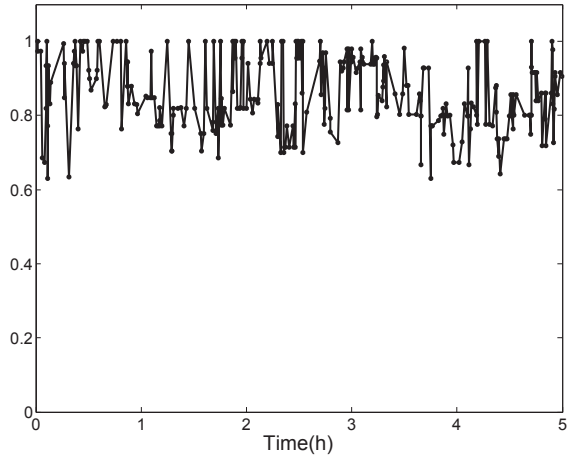
Figure 4.10(a), (b), and (c) show the ratio between the objective values obtained by our proposed online algorithm and those by the strongest-signal AP assignment algorithm when user arrival rates are 10, 20, and 30 per hour, respectively. As shown in the figures, most of time the proposed online algorithm has superior advantage over the strongest-signal AP assignment algorithm. In Fig. 4.10(a), there are a total of 105 events during the 5-hour simulation time. The average ratio



(a) user arrival rate is 10 per hour.

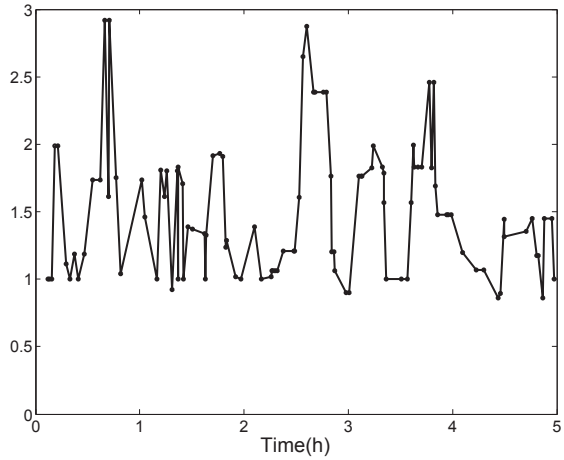


(b) user arrival rate is 20 per hour.

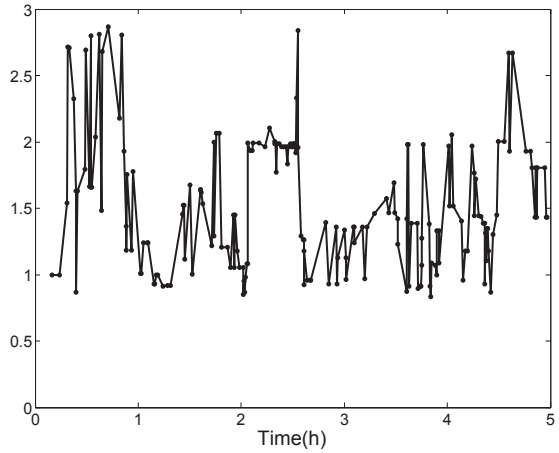


(c) user arrival rate is 30 per hour.

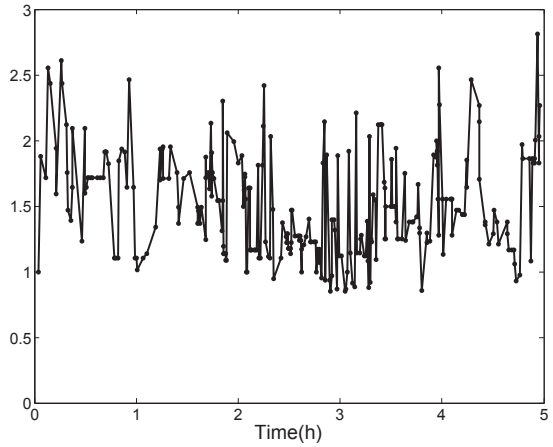
Figure 4.9: Ratio between the objective values obtained by the online algorithm and those by offline algorithm with one-shot AP assignment.



(a) user arrival rate is 10 per hour.



(b) user arrival rate is 20 per hour.



(c) user arrival rate is 30 per hour.

Figure 4.10: Ratio between the objective values obtained by the proposed online algorithm and those by the strongest-signal AP assignment algorithm.

between the two is 1.49. In Fig. 4.10(b), there are a total of 197 events during 5 hours. The average ratio between the two is 1.51. In Fig. 4.10(c), there are a total of 296 events during 5 hours, The average ratio between the two is 1.52. These results confirm that optimizing AP assignment, even for an online algorithm, is critical in maximizing throughput performance.

4.8 Chapter Summary

In this chapter, we investigated AP assignment and transmission scheduling in a multi-AP 60 GHz WLAN. Three AP assignment schemes were considered, namely per-time slot AP assignment, one-shot AP assignment, and strongest-signal AP assignment. Each AP assignment scheme is jointly optimized with transmission scheduling with the objective of maximizing minimum achievable user throughput. We find that both the per-time slot and one-shot AP assignment schemes significantly outperform the strongest-signal AP assignment scheme, which demonstrates that a judicious matching between APs and users is critical in optimizing network performance. On the other hand, there is little difference in the performance of the per-time slot AP assignment and one-shot AP assignment schemes. Due to its simplicity and lower overhead, we advocate to use one-shot AP assignment-based scheduling in practice. To address potential change in AP assignment under real-time user arrival/departure, we designed an online one-shot algorithm which does not alter AP assignment for existing users in the network. Our performance evaluation showed that the proposed online algorithm is competitive.

Chapter 5

Summary and Future Work

5.1 Dissertation Summary

In this dissertation, we investigated the impact of several advanced communication technologies on throughput performance in wireless networks. These technologies include MIMO, full duplex, and mmWave communication, all of which are going to play key roles in future wireless communications. The main contributions of this dissertation are summarized as follows:

- **Joint Flow Routing and DoF Allocation in Multi-hop MIMO Networks.** MIMO is a powerful PHY layer technology that exploits the capabilities of multiple antennas at a node. Although the benefits of MIMO at PHY layer for single-hop communications have been studied extensively, there has been limited work on its potential in multi-hop networks. In this work, we employed a DoF based model that captures SM and IC capabilities of MIMO. We studied a throughput maximization problem through joint consideration of flow routing and DoF allocation, The resulting formulation is in the form of a mixed-integer linear program (MILP). Our main contribution is an efficient polynomial time algorithm that offers a competitive solution.

- **Impact of Full Duplex Scheduling on End-to-end Throughput in Multi-hop Wireless Networks.** Full duplex aims to double the throughput at a wireless node by enabling simul-

taneous transmission and reception on the same channel. Although there has been extensive work on the design of full duplex transceivers, the benefit of full duplex for throughput performance at network level remains unclear. In this work, we investigated the performance gain of full duplex on end-to-end session throughput in a multi-hop wireless network. Through mathematical modeling and formulation, we formulated a throughput maximization problem. Using simulation results, we showed that session throughput under full duplex can exceed $2\times$ of that under half duplex in many cases. This work offered new understanding on the potential of full duplex capability on end-to-end throughput at network level.

- **On AP Assignment and Transmission Scheduling for Multi-AP 60 GHz WLAN.** mmWave communication in 60 GHz band is a new technology to realize multi-gigabit WLAN due to its massive bandwidth. Since 60 GHz signals are vulnerable to blockage, the deployment of multiple APs are proposed for a WLAN. In this work, we studied the important problem of AP assignment and transmission scheduling for a multi-AP 60 GHz WLAN. We proposed three AP assignment schemes, namely per-time slot AP assignment, one-shot AP assignment, and strongest-signal AP assignment. Under each scheme, we formulated a user throughput maximization problem by jointly considering AP assignment and transmission scheduling. Among the proposed schemes, we advocated to use one-shot AP assignment-based scheduling due to its competitive performance and simplicity. We also proposed an online one-shot algorithm for practical implementation. Through performance evaluation, we showed that our online algorithm is competitive when compared with the optimal offline algorithm.

5.2 Future Work

In this dissertation, we studied a set of problems on throughput performance in wireless networks that exploit advanced communication technologies. As expected, many problems remain to be explored. The following is a small set of problems that arise from our research and are worth further investigation.

- **Bridging MIMO DoF and data rate in multi-hop networks.** In Chapter 2, we employed a DoF-based model to study a throughput maximization problem in a multi-hop wireless network. The throughput was measured in terms of the number of DoFs. But ultimately, one would be interested in achievable data rate measured in bits per second. Therefore, it is important to map the number of DoFs to actual data rate by taking into consideration the channel condition.

- **Full duplex in multi-hop MIMO networks.** In Chapter 3, we investigated the impact of full duplex scheduling on network throughput performance in a multi-hop network. We assumed that each node is equipped with a single antenna. Recently, full duplex MIMO radio has been demonstrated and it will be interesting to extend our full duplex study from single-antenna to MIMO. The main challenge here is to model the IC capability of MIMO when each node can transmit and receive simultaneously.

- **User selection and scheduling in multi-user 60 GHz WLAN.** In Chapter 4, we studied AP assignment and transmission scheduling problem in a multi-AP 60 GHz WLAN. We assume that each AP can select at most one of tis LOS users for communication in each time slot. Recent studies showed the feasibility of multi-user beamforming in 60 GHz, i.e., a node can form multiple beams toward different directions for data transmission. In a time-slotted system, user selection (grouping) affects the user rate in each time slot and transmission scheduling affects the averaged user throughput over a frame. It would be interesting to consider these two problems jointly under multi-user beamforming.

- **Scheduling algorithm for multi-AP 60 GHz WLAN with user mobility.** User mobility poses serious challenges to multi-AP 60 GHz WLAN. First, since 60 GHz communication is directional, user mobility introduces significant overhead in beam steering for an AP to track its user. Therefore, good tracking algorithms are needed to give a good estimation of user position. Moreover, 60 GHz communication requires a LOS path between an AP and a user. As user moves, the LOS path between an AP and its user may be blocked. Therefore, carefully designed handover algorithms between APs are needed.

Bibliography

[1] E. Ahmed, A. Eltawil, and A. Sabharwal, "Self-interference cancellation with nonlinear distortion suppression for full-duplex systems," in *Proc. Asilomar Conference on Signals, Systems and Computers*, pp. 1199–1203, Pacific Grove, CA, USA, Nov. 2013.

[2] S. Alalusi and R. Brodersen, "A 60GHz phased array in CMOS," in *Proc. IEEE CICC*, pp. 393–396, San Jose, USA, Sept. 2006.

[3] S.M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[4] X. An, C. Sum, R. V. Prasad, J. Wang, Z. Lan, J. Wang, R. Hekmat, H. Harada, and I. Niemegeers, "Beam switching support to resolve link-blockage problem in 60GHz WPANs," in *Proc. IEEE Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 390–394, Tokyo, Japan, Sept. 2009.

[5] C. Anderson and T. Rappaport, "In-building wideband partition loss measurements at 2.5 and 60GHz," *IEEE Transactions on Wireless Communications.*, vol. 3, no. 3, pp. 922–928, May 2004.

[6] E. Aryafar, M.A. Khojastepour, K. Sundaresan, S. Rangarajan, and M. Chiang, "MIDU: Enabling MIMO full duplex," in *Proc. ACM MobiCom*, pp. 257-268, Istanbul, Turkey, Aug. 2012.

[7] G. Athanasiou, P. C. Weeraddana, C. Fischione, and L. Tassiulas, "Optimizing client association for load balancing and fairness in millimeter-wave wireless networks," *IEEE/ACM Transactions on Networking.*, vol. 23, no. 3, pp. 836–850, June 2015.

[8] T.K. Baranwal, D.S. Michalopoulos, and R. Schober, "Outage analysis of multihop full duplex relaying," *IEEE Communications Letters,* vol. 17, no. 1, pp. 63–66, Jan. 2013.

[9] R. Bhatia and L. Li, "Throughput optimization of wireless mesh networks with MIMO links," in *Proc. IEEE INFOCOM,* pp. 2326–2330, Anchorage, AK, May 2007.

[10] D. Bharadia, E. McMilin, and S. Katti, "Full duplex radios," in *Proc. ACM SIGCOMM,* pp. 375–386, New York, NY, USA, Aug. 2013.

[11] D. Bhatia and S. Katti, "Full duplex MIMO radios," in *Proc. USENIX NSDI,* pp. 359–372, Seattle, WA, April 2014.

[12] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H.V. Poor, *MIMO Wireless Communications,* Cambridge University Press, Jan. 2007.

[13] D.M. Blough, G. Resta, P. Santi, R. Srinivasan, and L.M. Cortes-Pena, "Optimal one-shot scheduling for MIMO networks," in *Proc. IEEE SECON,* pp. 377–385, Salt Lake City, Utah, June 2011.

[14] R.S. Blum, "MIMO capacity with interference," *IEEE Journal on Selected Areas in Communications,* vol. 21, no. 5, pp. 793–801, Jun. 2003.

[15] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. on Information Theory,* vol. 49, no. 7, pp. 1691–1706, July 2003.

[16] S. Cartreux, L.J. Greenstein, and P.F. Dressen, "Simulation results for an interference-limited multiple-input multiple-output cellular system," *IEEE Communications Letters,* vol. 4, no. 11, pp. 334–336, Nov. 2000.

[17] K. Chandra, R. V. Prasad, B. Quang, and I. G. M. M. Niemegeers, "Cogcell: cognitive interplay between 60GHz picocells and 2.4/5 GHz hotspots in the 5G era," *IEEE Communications Magazine,* vol. 53, no. 7, pp. 118–125, July 2015.

[18] B. Chen and M.J. Gans, "MIMO communications in ad hoc networks," *IEEE Transactions on Signal Processing,* vol. 54, no. 7, pp. 2773–2783, July 2006.

[19] W. Cheng, X. Zhang, and H. Zhang, "RTS/FCTS mechanism based full-duplex MAC protocol for wireless networks," in *Proc. IEEE GLOBECOM,* pp. 5017–5022, Atlanta, GA, USA, Dec. 2013.

[20] J.I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving single channel, full duplex wireless communication," in *Proc. ACM MobiCom,* pp. 1–12, New York, NY, USA, Sept. 2010.

[21] A.C. Cirik, Y. Rong, and Y. Hua, "Achievable rates of full-duplex MIMO radios in fast fading channels with imperfect channel estimation," *IEEE Transactions on Signal Processing,* vol. 62, no. 15, pp. 3874–3886, Aug. 2014.

[22] M.F. Demirkol and M.A. Ingram, "Power-controlled capacity for interfering MIMO links," in *Proc. IEEE Vehicular Technology Conference,* pp. 187–191, Atlantic City, NJ, Oct. 2001.

[23] M.F. Demirkol and M.A. Ingram, "Stream control in network with interfering MIMO links," in *Proc. IEEE Wireless Communications and Networking Conference,* pp. 343–348, New Orleans, LA, Mar. 2003.

[24] M. Duarte and A. Sabharwal, "Full-duplex wireless communications using off-the-shelf radios: Feasibility and first results," in *Proc. Asilomar Conference on Signals, Systems and Computers,* pp. 1558–1562, Pacific Grove, CA, USA, Nov. 2010.

[25] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Transactions on Wireless Communications,* vol. 11, no. 12, pp. 4296–4307, Dec. 2012.

[26] M. Duarte, A. Sabharwal, V. Aggarwal, R. Jana, K.K. Ramakrishnan, C.W. Rice, and N.K. Shankaranarayanan, "Design and characterization of a full-duplex multiantenna system for WiFi networks," *IEEE Transactions on Vehicular Technology,* vol. 63, no. 3, pp. 1160–1177, Mar. 2014.

[27] E. Everett, M. Duarte, C. Dick, and A. Sabharwal, "Empowering full-duplex wireless communication by exploiting directional diversity," in *Proc. Asilomar Conference on Signals, Systems and Computers,* pp. 2002–2006, Pacific Grove, CA, USA, Nov. 2011.

[28] E. Everett, A. Sahai, and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," *IEEE Transactions on Wireless Communications,* vol. 13, no. 2, pp. 680–694, Oct. 2013.

[29] X. Fang, D. Yang, and G. Xue, "Distributed algorithms for multipath routing in full-duplex wireless networks," in *Proc. IEEE MASS,* pp. 102–111, Valencia, Spain, Oct. 2011.

[30] M. Fukumoto and M. Bandai, "MIMO full-duplex wireless: Node architecture and medium access control protocol," in *Proc. Seventh International Conference on Mobile Computing and Ubiquitous Networking (ICMU),* pp. 76–77, Singapore, Jan. 2014.

[31] Z. Genc, U. H. Rizvi, E. Onur, and I. Niemegeers, "Robust 60GHz indoor connectivity: is it possible with reflections?" in *Proc. IEEE Vehicular Technology Conference,* pp. 1–5, Taipei, Taiwan, May 2010.

[32] D. Gesbert, M. Shafi, D. Shiu, P.J. Smith, and A. Naguib, "From theory to practice: An overview of MIMO space-time coded wireless systems," *IEEE Journal on Selected Areas in Communications,* vol. 21, no. 3, pp. 281–302, Apr. 2003.

[33] S. Goyal, P. Liu, O. Gurbuz, E. Erkip, and S. Panwar, "A distributed MAC protocol for full duplex radio," in *Proc. Asilomar Conference on Signals, Systems and Computers,* pp. 788–792, Pacific Grove, CA, USA, Nov. 2013.

[34] A. Gudipati, D. Perry, L. E. Li, and S. Katti, "SoftRAN: software defined radio access network," in *Proc. ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking,* pp. 25–30, Hong Kong, China, Aug. 2013.

[35] F. Gutierrez, S. Agarwal, K. Parrish, and T. S. Rappaport, "On-chip integrated antenna structures in CMOS for 60GHz WPAN systems," *IEEE Journal on Selected Areas in Communications.,* vol. 27, no. 8, pp. 1367–1378, Oct. 2009.

[36] B. Hamdaoui and K.G. Shin, "Characterization and analysis of multi-hop wireless MIMO network throughput," in *Proc. ACM MobiHoc,* pp. 120–129, Montreal, Quebec, Canada, Sep. 2007.

[37] Z. He, S. Mao, and T. S. Rappaport, "Minimum time length link scheduling under blockage and interference in 60GHz networks," in *Proc. IEEE WCNC,* pp. 1–6, New Orleans, USA, March 2015.

[38] S. Hong, J. , S. Katti, "Picasso: Flexible RF and spectrum slicing," in *Proc. ACM SIGCOMM,* pp. 37–48, New York, NY, USA, Aug. 2012.

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

[40] S.A. Jafar and M. Fakhereddin, "Degrees of freedom for the MIMO interference channel," *IEEE Trans. on Information Theory,* vol. 53, no. 7, pp. 2637–2642, July 2007.

[41] M. Jain, J.I. Choi, T. Kim, D. Bharadia, S. Seth, K. Srinivasan, P. Levis, S. Katti, and P. Sinha, "Practical, real-time, full duplex wireless," in *Proc. ACM MobiCom,* pp. 301–312, New York, NY, USA, Aug. 2011.

[42] D.B. Johnson, D.A. Maltz, and J. Broch, "DSR: the dynamic source routing protocol for multi-hop wireless ad hoc networks," *Ad Hoc Networking,* Addison-Wesley, 2001, ISBN-13: 978-0321579072.

[43] M.A. Khojastepour, K. Sundaresan, S. Rangarajan, X. Zhang, and S. Barghi, "The case for antenna cancellation for scalable full-duplex wireless communications," in *Proc. the 10th ACM Workshop on Hot Topics in Networks,* pp. 1–6, New York, NY, USA, 2011.

[44] M.A. Khojastepour and S. Rangarajan, "Wideband digital cancellation for full-duplex communications," in *Proc. Asilomar Conference on Signals, Systems and Computers,* pp. 1300–1304, Pacific Grove, CA, USA, Nov. 2012.

[45] J. Kim, O. Mashayekhi, H. Qu, M. Kazandjieva, and P. Levis, "Janus: A novel mac protocol for full duplex radio," Technical Report, Stanford Univeristy, 2013. Available at http://web.stanford.edu/ skatti/pubs/mobicom10-fd.pdf

[46] S. Kim and W.E. Stark, “On the performance of full duplex wireless networks,” in *Proc. 47th Annual Conference on Information Science and Systems,* pp. 1–6, Baltimore, MD, USA, March 2013.

[47] D. Kim, H. Lee, and D. Hong, “A survey of in-band full-duplex transmission: From the perspective of PHY and MAC layers,” *IEEE Communications Surveys & Tutorials,* vol. PP, no. 99, pp. 1, Feb. 2015.

[48] J. Lee and T.Q.S. Quek, “Hybrid full-/half-duplex system analysis in heterogeneous wireless networks,” *IEEE Transactions on Wireless Communications,* Vol. 13, no. 5, pp. 2883–2895, May 2015.

[49] D. Liu and R. Sirdeshmukh, “A patch array antenna for 60GHz package applications,” in *Proc. IEEE Antennas and Propagation Society International Symposium*, pp. 1–4, San Diego, USA, July 2008.

[50] B. Mahboobi and M. Ardebilipour, “Joint power allocation and routing in full-duplex relay network: An outage probability approach,” *IEEE Communications Letters,* vol. 17, no. 8, pp. 1497–1500, Aug. 2013.

[51] R.J. Marhefka and J. D. Kraus, *Antennas for All Applications,* third edition, Chapter 2, McGraw–Hill, 2002.

[52] N. Mckeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, “OpenFlow: enabling innovation in campus networks,” *ACM SIGCOMM Computer Communication Review,* vol. 38, no. 2, pp. 69–74, Apr. 2008.

[53] S. Murthy and J. J. Garcia-Luna-Aceves, “Congestion-oriented shortest multi-path routing,” in *Proc. IEEE INFOCOM*, pp. 1038–1036, San Francisco, CA, May 1996.

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

[55] T. Newman, X. Chen, D. Datla, H. Volos, C. Dietrich, T. Bose, and J. H. Reed, “CORNET: cognitive radio mesh and dynamic spectrum allocation demonstration,” in *Proc. IEEE DySPAN Symposium*, Oct. 2008.

[56] T. Nitsche, A. B. Flores, E. W. Knightly and J. Widmer, “Steering with eyes closed: mm-wave beam steering without in-band measurement,” in *Proc. IEEE INFOCOM*, pp. 2416–2424, Hong Kong, China, May 2015.

[57] Y. Niu, Y. Li, D. Jin, L. Su, and D. Wu, “Blockage robust and efficient scheduling for directional mmWave WPANs,” *IEEE Transactions on Vehicular Technology.*, vol. 64, no. 2, pp. 728–742, Feb. 2015.

[58] J-S. Park, A. Nandan, M. Gerla, and H. Lee, “SPACE-MAC: Enabling spatial reuse using MIMO channel-aware MAC,” in *Proc. IEEE ICC*, pp. 3642–3646, Seoul, Korea, May 16–20, 2005.

[59] P. Papadimitratos, Z. J. Haas, and E. G. Sirer, “Path set selection in mobile ad hoc networks,” in *Proc. ACM MobiHoc*, pp. 1–11, Lausanne, Switzerland, Jun. 2002.

[60] A.J. Paulraj, D.A. Gore, R.U. Nabar, and H. Bolcskei, “An overview of MIMO communications – A key to gigabit wireless,” *Proceedings of the IEEE*, vol. 92, no. 2, pp. 198–218, Feb. 2004.

[61] C. Perkins, E. Belding-Royer, and S. Das, *Ad hoc on-demand distance vector (AODV) routing*, RFC 3561, IETF, July 2003.

[62] J. Qiao, X. Shen, J. W. Mark, Z. Shi, N. Mohammadizadeh, “MAC-layer integration of multiple radio bands in indoor millimeter wave networks,” in *Proc. IEEE WCNC*, pp. 889–894, Shanghai, China, April 2013.

[63] X. Qin, X. Yuan, Y. Shi, Y.T. Hou, W. Lou and S.F. Midkiff, “On throughput maximization for a multi-hop MIMO network,” in *Proc. IEEE MASS*, pp. 37–45, Hangzhou, China, Oct. 2013.

[64] B. Radunovic, D. Gunawardena, P. Key, A. Proutiere, N. Singh, V. Balan, and G. Dejean, "Rethinking indoor wireless mesh design: Low power, low frequency, full-duplex," in *Proc. IEEE WIMESH*, pp. 1–6, Boston, MA, USA, Jun. 2010.

[65] D. Ramirez and B. Aazhang, "Optimal routing and power allocation for wireless networks with imperfect full-duplex nodes," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 1536–1276, Sept. 2013.

[66] T. S. Rappaport, J. N. Murdock, F. Gutierrez, "State of the art in 60GHz integrated circuits and systems for wireless communications," *Proceedings of the IEEE.*, vol. 99, no. 8, pp. 1390–1436, Aug. 2011.

[67] T. Riihonen, S. Werner, and R. Wichman, "Residual self-interference in full-duplex MIMO relays after null-space projection and cancellation," in *Proc. Asilomar Conference on Signals, Systems and Computers*, pp. 653–657, Pacific Grove, CA, USA, Nov. 2010.

[68] A. Sahai, G. Patel, and A. Sabwarwal, "Pushing the limits of full duplex: Design and real-time implementation," Technical Report, Rice Univeristy, 2011. Available at http://arxiv.org/abs/1107.1276.

[69] K. Sato and T. Manabe, "Estimation of propagation-path visibility for indoor wireless LAN systems under shadowing condition by human bodies," in *Proc. IEEE Vehicular Technology Conference*, pp. 2109–2113, Ottawa, Canada, May 1998.

[70] M.S. Bazaraa, J.J. Jarvis, H.D. Sherali, *Linear Programming and Network Flows*, fourth edition, Chapter 8, John Wiley & Sons, 2010.

[71] Y. Shi, J. Liu, C. Jiang, C. Gao, and Y.T. Hou, "A DoF-based link layer model for multi-hop MIMO networks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 7, pp. 1395–1408, July 2014.

[72] S. Singh, R. Mudumbai, and U. Madhow, "Interference analysis for highly directional 60GHz mesh networks: the case for rethinking medium access control," *IEEE/ACM Transactions on Networking.*, vol. 19, no. 5, pp. 1513–1527, Mar. 2011.

[73] S. Singh, F. Ziliotto, U. Madhow, E. M. Belding, and M. Rodwell, "Blockage and directivity in 60GHz wireless personal area networks: from cross-layer model to multi-hop MAC design," *IEEE Journal on Selected Areas in Communications.*, vol. 27, no. 8, pp. 1400–1413, Oct. 2009.

[74] P. F. M. Smulders, "Statistical characterization of 60GHz indoor radio channels," *IEEE Transactions on Antennas and Propagation.*, vol. 57, no. 10, pp. 2820–2829, Aug. 2009.

[75] I. K. Son, S. Mao, M. X. Gong, and Y. Li, "On frame-based scheduling for directional mmWave WPANs," in *Proc. IEEE INFOCOM*, pp. 2149–2157, Orlando, USA, March 2012.

[76] C. Sum, Z. Lan, R. Funada, J. Wang, T. Baykas, M. A. Rahman, and H. Harada, "Virtual time-slot allocation scheme for throughput enhancement in a millimeter-wave multi-Gbps WPAN system," *IEEE Journal on Selected Areas in Communications.*, vol. 27, no. 8, pp. 1379–1389, Oct. 2009.

[77] K. Sundaresan, R. Sivakumar, M. Ingram, and T-Y. Chang, "Medium access control in ad hoc networks with MIMO links: Optimization considerations and algorithms," *IEEE Transactions on Mobile Computing*, vol. 3, no. 4, pp. 350–365, Oct. 2004.

[78] S. Sur, V. Venkateswaran, X. Zhang, and P. Ramanathan, "60GHz indoor networking through flexible beams: A link-level profiling," in *Proc. ACM SIGMETRICS*, pp. 71–84, New York, USA, June 2015.

[79] S. Sur, X. Zhang, P. Ramanathan, and R. Chandra, "Beamspy: Enabling robust 60GHz links under blockage," in *Proc. USENIX NSDI*, pp. 193–206, Santa Clara, USA, March 2016.

[80] K. Tamaki, H. Ari Raptino, Y. Sugiyama, M. Bandai, S. Saruwatari, and T. Watanabe, "Full duplex media access control for wireless multi-hop networks," in *Proc. IEEE 77th Vehicular Technology Conference*, pp. 1–5, Dresden, Germany, June 2013.

[81] Z. Tong and M. Haenggi, "Throughput analysis for full-duplex wireless networks with imperfect self-interference cancellation," *IEEE Transactions on Communications*, vol. PP, no. 99, pp. 1, Aug. 2015.

[82] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication,* Cambridge University Press, 2005.

[83] S. Wang, V. Venkateswaran and X. Zhang, “Exploring full-duplex gains in multi-cell wireless networks: A spatial stochastic framework,” in *Proc. IEEE INFOCOM,* pp. 855–863, Hong Kong, China, May 2015.

[84] X. Xie and X. Zhang, “Semi-synchronous channel access for full-duplex wireless networks,” in *Proc. IEEE International Conference on Network Protocols,* pp. 209–214, Raleigh, NC, USA, Oct. 2014.

[85] X. Xie and X. Zhang, “Does full duplex double the capacity of wireless networks,” in *Proc. IEEE INFOCOM,* pp. 253–261, Toronto, Canada, April 2014.

[86] H. Xu, V. Kukshya, and T. S. Rappaport, “Spatial and temporal characteristics of 60 GHz indoor channel,” *IEEE Journal on Selected Areas in Communications.*, vol. 20, no. 3, pp. 620–630, Apr. 2002.

[87] Y. Yang and N.B. Shroff, “Scheduling in wireless networks with full-duplex cut-through transmission,” in *Proc. IEEE INFOCOM,* pp. 2164–2172, Hong Kong, China, May 2015.

[88] C. Yiu and S. Singh, “Empirical capacity of mmWave WLANs,” *IEEE Journal on Selected Areas in Communications.*, vol. 27, no. 8, pp. 1479–1487, Oct. 2009.

[89] H. Zeng, Y. Shi, Y.T. Hou, and W. Lou, “An efficient DoF scheduling algorithm for multi-hop MIMO networks,” in *Proc. IEEE INFOCOM,* pp. 1564–1554, Turin, Italy, April 2013.

[90] H. Zeng, Y. Shi, Y.T. Hou, R. Zhu, and W. Lou, “A novel MIMO DoF model for multi-hop networks, ” *IEEE Network*, Vol. 28, no. 5, pp. 81–85, Oct. 2014.

[91] X. Zhang, S. Zhou, X. Wang, Z. Niu, X. Lin, D. Zhu and M. Lei, “Improving network throughput in 60GHz WLANs via multi-AP diversity,” in *Proc. IEEE ICC,* pp. 4803–4807, Ottawa, Canada, June 2012.

[92] L. Zheng and D. Tse, "Communication on the Grassmann Manifold: A Geometric Approach to the Noncoherent Multiple-Antenna Channel," *IEEE Trans. on Information Theory,* vol. 48, no. 2, pp. 359–383, Feb. 2002.

[93] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. on Information Theory,* vol. 49, no. 5, pp. 1073–1096, May 2003.

[94] W. Zhou, K. Srinivasan, and P. Sinha, "RCTC: Rapid concurrent transmission coordination in full duplex wireless networks," in *Proc. IEEE International Conference on Network Protocols,* pp. 1–10, Goettingen, Germany, Oct. 2013.

[95] IBM ILOG CPLEX: http://www.ibm.com/software/integration/optimization/cplex/

[96] CTIA: Mobile Data Demand: Growth Forecasts Met, available at `http://www.ctia.org/docs/default-source/default-document-library/062115mobile-data-demands-white-paper-new.pdf`.

[97] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 20152020 White Paper, available at `http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.pdf`.

[98] Ericsson Mobility Report: On the Pulse of the Networked Society, available at `https://www.ericsson.com/res/docs/2016/ericsson-mobility-report-2016.pdf`.

[99] ETSI TR 125 942 (V3.3.0): Universal Mobile Telecommunications System (UMTS); RF system scenarios (3GPP TR 25.942 version 3.3.0).

[100] ITU-R TR M.2135-1: Guidelines for evaluation of radio interface technologies for IMT-Advanced.

[101] "IEEE standard for information technologytelecommunications and information exchange between systemslocal and metropolitan area networksspecific requirements Part 11: Wireless

lan medium access control (MAC) and physical layer (PHY) specifications amendment 5: Enhancements for higher throughput," IEEE Std 802.11n–2009, Oct. 2009

[102] "Air interface for fixed broadband wireless access systems," IEEE Std 802.16–2004, Oct. 2004.

[103] UTRA-UTRAN Long Term Evolution (LTE) and 3GPP System Architecture Evolution (SAE), available at `ftp://ftp.3gpp.org/Inbox/2008_web_files/LTA_Paper.pdf/`

[104] *Draft Standard for Information TechnologyTelecommunications and Information Exchange Between SystemsLocal and Metropolitan Area NetworksSpecific RequirementsPart 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) SpecificationsAmendment 4: Enhancements for Very High Throughput in the 60 Ghz Band*, IEEE P802.11ad/D9.0, Oct. 2012.

[105] IEEE 802.15 WPAN Millimeter Wave Alternative PHY Task Group 3c (TG3c), available at: `http://www.ieee802.org/15/pub/TG3c.html`.