On Programmable Control and Optimization for Multi-Hop Wireless Networks

Brian A. Jalaian

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
Electrical Engineering

Y. Thomas Hou, Chair
R. Michael Buehrer
Venkat Dasari
Wenjing Lou
Jeffrey H. Reed
Hanif D. Sherali
Yi Shi

September 8, 2016
Blacksburg, Virginia

Keywords: wireless network, optimization, cross-layer design, multi-hop, programmable network

On Programmable Control and Optimization for Multi-Hop Wireless Networks

Brian A. Jalaian

(ABSTRACT)

Traditionally, achieving good performance for a multi-hop wireless network is known to be difficult. The main approach to control the operation of such a network relies on a distributed paradigm, assuming that a centralized approach is not feasible. Relying on a distributed paradigm could be justified at the time when the basic technical building blocks (e.g., node computational power, communication technology, positioning technology) were the bottlenecks. Recent advances and breakthroughs in these technical areas along with the emergence of programmable networks with softwarized control plane intelligence allow us to consider employing a centralized optimization paradigm to control and manage the operation of a multi-hop wireless network. The programmable control provides a platform on which the centralized global network optimization paradigm can be supported. The benefits of a centralized network optimization lie specially in that a network may be configured in such a way that offers optimal performance, which is hardly possible for a network relying on distributed operation.

The objectives of this dissertation are to fully understand the potential benefits of a centralized control plane for a multi-hop wireless network, to identify any new challenges under this new paradigm, and to devise innovative solutions for optimal performance via a centralized control plane. Given that the performance of a wireless network heavily depends on its physical layer capabilities, we will consider a number of advanced wireless technologies, including MIMO, full duplex, and interference cancellation at the physical layer. The focus is on building tractable computational models for these wireless technologies that can be used for modeling, analysis and optimization in the centralized control plane. Problem formulation and efficient solution procedures are developed for various centralized optimization problems across multiple layers. End-to-end throughput maximization is a key objective among these optimization problems on the centralized control plane and is used to demonstrate the superior advantage of this paradigm. We study several problems:

- **Integration of SIC and MIMO DoF IC.** We propose to integrate MIMO Degree-of-Freedom (DoF) interface cancellation (IC) and Successive Interference Cancellation (SIC) in MIMO multi-hop network under DoF protocol model. We show that DoF-based IC and SIC can be jointly integrated to combat the interference more effectively and improve the end-to-end throughput significantly. We develop the necessary mathematical models to realize the idea in a multi-hop wireless network.

- **Full-Duplex MIMO Wireless Networks Throughput.** We investigate the performance of MIMO full-duplex (FD) in a multi-hop network. We show that if IC is exploited, MIMO FD can achieve significant throughput gain over MIMO HD in a multi-hop network, which is contrary to the recent literature suggesting an unexpected marginal gain. Our proposed model handles the additional network interference by joint efficient link scheduling and interference cancellation.

- **PCP in Tactical Wireless Networking.** We propose the idea of the Programmable Control Plane (PCP) for the tactical wireless network under the protocol model. PCP decouples the control and data plane and allows the network control layer functionalities to be dynamically configured to adapt to specific wireless channel conditions, customized applications and/or certain tactical situations. The proposed PCP functionalities are cast into a centralized optimization problem, which can be updated as needed and provide a centralized intelligence to manage the operation of a wireless MIMO multi-hop network under the protocol model.

- **UPCP in Heterogeneous Wireless Networks.** We propose the idea of the Unified Programmable Control Plane (UPCP) for tactical heterogeneous wireless networks with interference management capabilities under the SINR model. The UPCP abstracts the complexity of the underlying network comprised of heterogeneous wireless technologies and provides a centralized intelligence over the network resources. We develop necessary mathematical model to realize the UPCP.

On Programmable Control and Optimization for Multi-Hop Wireless Networks

Brian A. Jalaian

(GENERAL AUDIENCE ABSTRACT)

In the past decades, wireless ad hoc communication networks have found a number of applications in both civilian and military environments. Such networks are comprised of a set of smart nodes, which are able to organize themselves into a multi-hop network (able to communicate from the source nodes to the destination nodes across multiple intermediary relay nodes) to provide various services such as unattended and real-time surveillance. Their capabilities of self-form and self-heal make them attractable for network deployment and maintenance, especially in the scenarios where infrastructure is hard to establish. Because of their ease of deployment and independence of infrastructure, wireless ad hoc network have motivated more and more research efforts to sustain their continued growth and well-being. Nevertheless, with rapidly increasing demand for data rate from various applications, we find ourselves still very much in the infancy of the development of such networks, which have the potential to offer orders-of-magnitude higher network-level throughput.

Traditionally, the main approach to control the operation of wireless ad hoc network relies on a distributed paradigm, assuming that a centralized approach is not feasible. Relying on a distributed paradigm could be justified at the time when were the bottlenecks. Recent advances and breakthroughs in basic technical areas the basic technical building blocks (e.g., node computational power, communication technology, positioning technology) along with the emergence of programmable networks with softwarized control plane intelligence allow us to consider employing a centralized optimization paradigm to control and manage the operation of a multi-hop wireless network. The objectives of this dissertation are to fully understand the potential benefits of a centralized optimization paradigm in multi-hop wireless network, to identify any new challenges under this new paradigm, and to devise innovative solutions for optimal performance.

Acknowledgments

This work took a long journey, and I have not made it alone.

First and foremost, I would like to thank my advisor Prof. Tom Hou for his guidance, help, support, and encouragement throughout my Ph.D. studies. His keen vision led me to see the outcome of this dissertation, his remarkable effort contributed substantially to the interesting findings in this dissertation, and his pursuit of true scholarship inspired me to challenge myself to the academic level which I would have never achieved otherwise. I will always remeber when I sat beside him and we discussed problems we faced in my research in details. Prof. Hou was beyond an adviser for my research and he was also a mentor. I enjoyed his advices regarding my future career and without his support I would not be able to see the light through the tunnel. His scholarship and dedication to research make him a role model for me as I continue along my career path.

I would also like to thank Dr. Venkat Dasari for his mentorship and guidance while I was continuing my PhD studies at the U.S. Army Research Lab (ARL). This dissertation could have not been completed without his support, guidance and dedication. Dr. Dasari introduced me to the entire new world of research for solving challenging problems in network science. He helped me to bridge the gap between theory and practice. I can always remember that he was always available even during the weekend and after work hours if I needed to discuss our research problems.

I would also like to thank Dr. Raju Namburu, chief of the Computational Sciences Division within ARL's Computational and Information Sciences Directorate for his support toward my PhD studies and research in ARL. I would like to also thank Mr. Dale Shires of ARL's Advanced

Architecture Branch for his support.

I also want to thank Dr. Yi Shi for his help and guidance. I appreciated the time that he spent reviewing my work and his valuable feedback.

I would also like to acknowledge and thank the rest of my dissertation committee for their time and efforts: Prof. Hanif Sherali, Prof. Wenjing Lou, Prof. Jeffery Reed, and Prof. Michael Buehrer. Their comments and questions have helped improve the overall quality of my dissertation.

I would like to thank my former colleague and a good friend Xu Yuan. I sat next to him in the office for several years and constantly bugged him with my research problems. Xu was always very helpful and never hesitated to review my work or share his opinion about my research. Xu was also a great friend and I enjoyed his company when we hung out for lunch or dinner times. I also like to thank my colleague and great friend Xiaoqi Qin. I appreciate our brainstorming sessions to discuss numerous research and coursework topics. I would like to acknowledge my current and former colleagues in my research group: Canming Jiang, Huacheng Zeng, and Liguang Xie. I not only enjoyed the time when we discussed research, but also enjoyed their company eating out at lunch or dinner time. I would like to also thank two of my best friends Robert L. Hall and Lee Barker. I enjoyed my time training Japanese karate and working out with them. They both always supported me and encouraged me along the way of my PhD journey. I would like to also thank Dr. Hooman Samani, my great friend for his support and encouragement during my PhD studies. I would like to thank Andrew Caballero and Debra Lee Rudy, two of my great friends for their support, encouragement, and constructive feedback on my defense presentation slides. I would like to thank my friends in and outside Virginia Tech for their friendship as well. I have thoroughly enjoyed my interactions with them in study and in life.

Finally, I wish to thank my parents, my sisters Yeganeh Jalaeian and Yasaman Rose Jalaian, and my uncle Ashkan Roshanzaer who have supported me through my PhD studies.

Sponser Acknowledgements

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.

I would like to express my gratitude to U.S. Army Research Laboratory for supporting this work. This work is supported in part by an appointment to the Student Research Participation Program at the U.S. Army Research Laboratory administered by the the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and U.S. Army Research Laboratory.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

Since its inception, the designs for a multi-hop wireless network adopted a fully distributed approach at all layers, similar to the Internet in its early days. On the control plane, each node was assumed to maintain only local and neighboring node information without having a global view of the network. Communication and networking relied on best-effort and hop-by-hop knowledge. Under such a distributed approach, connectivity was unreliable and throughput was low. Over the years, the poor performance under such a distributed design has become the most notable bottleneck for the success of multi-hop wireless networks.

Recent significant advances in computational, radio, localization, and networking technologies have motivated us to consider a centralized/distributed programmable control plane to handle globally-optimized management for multi-hop wireless networks. On the computation side, powerful computing resource is becoming readily available in the cloud and complex control and optimization problems may now be solved efficiently in real-time. On the radio side, we have seen the emergence of programmable or software-defined radio that allows waveforms and scheduling algorithms to be reconfigured on the fly (on a per-packet basis). Such radio is also becoming increasing

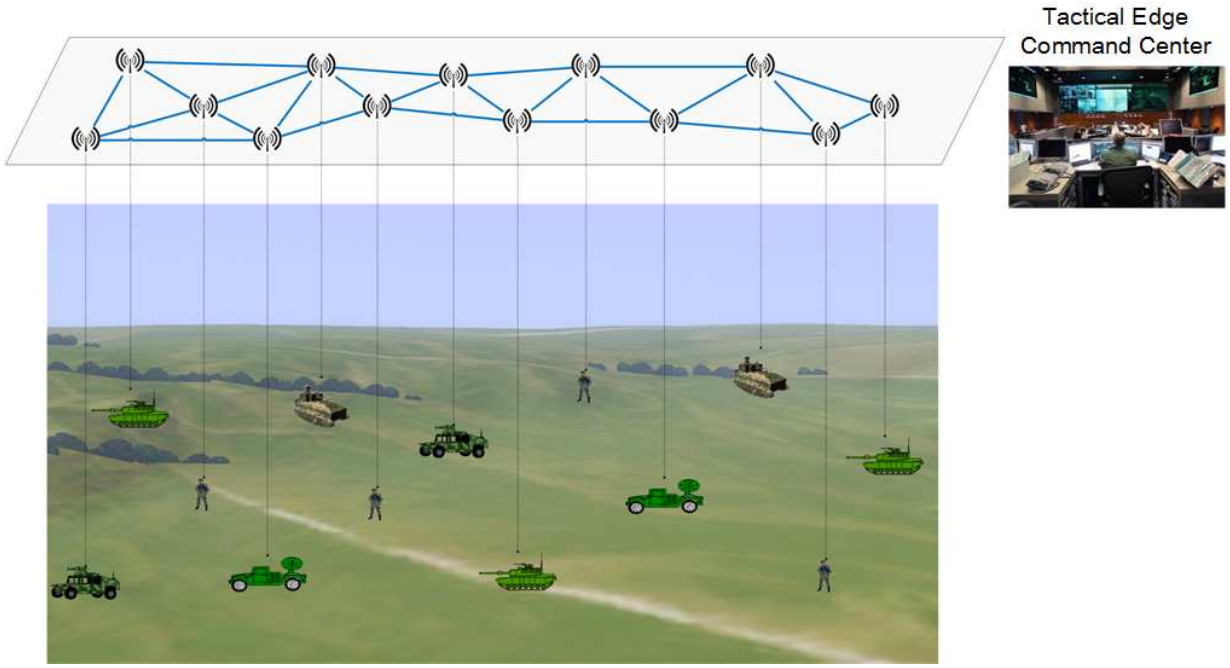


Figure 1.1: An illustration of a centralized control plane for the control and operation of a multi-hop wireless network.

intelligent (or cognitive) and can learn from its environment and adapt its operation behavior. On the localization side, the wide availability of GPS and other powerful positioning algorithms allow precise location information to be readily available in a very cost effective manner. This allows a network operator to obtain a precise topology map that directly matches the physical environment on the network (rather than merely a logical connectivity map). On the networking side, the decoupling of data and control planes is becoming widely accepted in the Internet community, as evident in the recent progress in SDN.

Capitalizing on these advances in high-performance computing, programmable radio and hardware, precision localization and positioning, and decoupling of data/control plane, we advocate a centralized control paradigm for multi-hop wireless networks (see Fig. 1.1). We believe this new paradigm will revolutionize the operational behavior and performance capabilities of multi-hop wireless networks. It will unleash the full potential of a multi-hop wireless network from bottom up. For the first time, the network operator can have a complete view of physical layer configu-

ration and capability of each node; the operator will be able to optimally configure (program) the physical layer (radio) on the fly, which is not typically possible without a programmable control plane. By performing global optimization via high-performance computing in the cloud, the multi-hop network can be configured to achieve a new performance envelope far beyond what was once achieved under the traditional distributed paradigm.

1.2 Dissertation Outline and Contributions

The objectives of this dissertation are to fully understand the potential benefits of a centralized control plane for a multi-hop wireless network, to identify any new challenges under this new paradigm, and to devise innovative solutions for optimal performance via a centralized control plane. Given that the performance of a wireless network heavily depends on its physical layer capabilities, we will consider a number of advanced wireless technologies, including MIMO, full duplex, interference cancellation and others at the physical layer. The focus will be on building tractable computational models for these wireless technologies that can be used for modeling, analysis and optimization in the centralized control plane. Problem formulation and efficient solution procedures will be developed for various centralized optimization problems across multiple layers. End-to-end throughput maximization will be a key objective among these optimization problems on the centralized control plane and will be used to demonstrate the superior advantage of this paradigm.

We summarize the main contributions of this research report as follows:

- In Chapter 2, we propose a union between MIMO DoF interface cancellation (IC) and successive interference cancellation (SIC) in MIMO multi-hop network under degree-of-freedom (DoF) protocol model. We show that DoF-based IC and SIC can help each other as follows: (i) SIC is exploited to decode multiple received signals to conserve DoF resources in IC, and (ii) DoF IC resolves the potential SINR barrier that SIC may encounter. We develop the necessary mathematical models to realize the two ideas in a multi-hop wireless

network. Together with scheduling and routing constraints, we develop a cross-layer optimization framework with joint DoF IC and SIC. By applying the framework on a throughput maximization problem, we find that SIC and DoF IC can indeed offer significant performance improvement by addressing each other's limitation.

- In Chapter 3, we study the performance of full duplex (FD) technology in MIMO multi-hop network under SINR model. The recent literature suggests that the network throughput gain of MIMO FD radios over MIMO half duplex (HD) is unexpectedly marginal due to the additional network interference, when both ends of the FD links are transmitting. In this Chapter, we show that if interference cancellation (IC) is employed, MIMO FD can achieve significant throughput gain over MIMO HD. By understanding the state-of-the-art MIMO full duplex transceiver architecture, we proposed a model for MIMO full duplex in a multi-hop network, where nodes can be scheduled to operate in HD or FD mode and perform MIMO spatial multiplexing or zero-forcing interference cancellation. We develop the necessary mathematical models to realize such a MIMO full duplex model. Together with scheduling and routing constraints, we develop a cross-layer optimization framework. By applying the framework on a throughput maximization problem, we study the throughput gain of MIMO FD over MIMO HD in a multi-hop wireless network. Our results suggest that MIMO FD can have a significant gain over MIMO HD given proper IC technique is employed in the network.

- In Chapter 4, we propose the idea of the programmable control plane for the tactical wireless network under protocol model. In the conventional wireless network, the control and data planes are coupled. This tight coupling limits the flexibility of the control plane. The programmable control plane decouples the control and data plane. Realizing a separated control plane for a wireless network requires a clear theoretical understanding for modeling the various functionalities of such control plane, which remains limited in the current literature. In our proposed programmable control plane, the network control layer functionalities can be dynamically configured to adapt to specific physical conditions, customized applications

and/or certain tactical situations. The proposed programmable control plane functionalities can be cast into a centralized optimization problem, which can be updated as needed. Specifically, we develop a cross-layer optimization framework, which characterizes the interaction between the physical, link, and network layers. By applying the framework to a throughput maximization problem, we show how the envisioned control plane programming framework can solve complex problems in tactical networks.

- In Chapter 5, we propose the idea of the unified programmable control plane for tactical heterogeneous wireless networks with interference management capabilities under SINR model. Tactical networks are wireless in nature to support mobility and rapid deployment in contested environments. These networks also consist of various heterogeneous networking technologies. Since the tactical networks are specialized purpose built networks and their main motivation is to solve special purpose functions, the interoperability between different technologies was not initially viewed as essential. As the deployment of wireless networks become ubiquitous both in private and government sectors, the lack of interoperability has become a disadvantage in creating a unified control plane. A unified control plane can abstract the complexity of heterogeneous wireless networks and can provide a centralized control over the network resources. In this Chapter, we develop the necessary mathematical model to realize the unified programmable control plane for heterogeneous wireless networks. We develop a cross-layer optimization framework, which characterizes the interaction between physical, link, and network layer for the unified programmable control plane in a heterogeneous wireless network. By applying the framework on a throughput maximization problem, we will show an application of the model to solve practical issues in a tactical network and gain some theoretical insight on the optimal behavior of the unified programmable control plane for a heterogeneous wireless network.

Chapter 2

Joint SIC and MIMO DoF for Interference Cancellation in Wireless Networks

2.1 Introduction

MIMO has been widely adopted by the communications industry and the research community due to its capabilities of spatial multiplexing (SM) gain, interference cancellation (IC), and diversity gain. Until recently, research on MIMO has been limited at the physical (PHY) layer or for single-hop communications due to the lack of tractable MIMO models. Recent advances in MIMO degree-of-freedom (DoF) models removed this stagnation and allowed MIMO research to penetrate the networking community [9, 12, 36, 40, 62, 71]. The concept of DoF was originally defined to represent the multiplexing gain of a MIMO channel in the information theory (IT) community. This DoF concept was then extended by the networking community to characterize a node's spatial freedom provided by its multiple antennas. Under a DoF model, only simple numerical computation is needed to account for a node's resource allocation for SM and IC. The basic idea of DoF-based MIMO models is as follows [71]: (i) The number of available DoFs at a node is equal to the number of its antennas. A node may use its DoFs for either SM or IC. (ii) For SM, both transmit and receive nodes need to consume DoFs. For each data stream, both the transmit

and receive nodes need to consume one DoF. (iii) For IC, either the transmit node or the receive node may consume DoFs. Clearly, under a DoF model, the number of available DoFs at a node is considered a precious recourse and must be utilized efficiently. In particular, when a node uses its DoFs for IC, its remaining DoFs for SM will be reduced. Therefore, there is a critical need to conserve DoFs for IC if one wishes to maximize the number of DoFs for SM.

Independent from MIMO, successive interference cancellation (SIC) is a powerful physical layer technique used in multi-user detection [40, 81]. It allows a receiver to take multiple interfering signals from different transmitters and decode each signal iteratively. In its simplest form, a receiver may decode the strongest signal from the aggregate received signals and considers all the other interfering signals as noise. If the strongest signal meets a SINR threshold, then it can be decoded successfully. Then the receiver subtracts it from the aggregate signals and repeats the process for the second strongest signal and so forth, until the desired signal is decoded successfully. Current research on exploiting SIC for single-antenna nodes in a wireless network can be found in [11, 29, 33, 35, 41, 52, 53, 54, 55, 56, 83, 91]. SIC has also been exploited in MIMO for point-to-point and multi-user communications (see related work in Section 2.6). Although attractive, SIC does have its limitations. Most notably, if the SINR threshold cannot be satisfied at any stage, the process cannot continue and the desired signal cannot be decoded.

Recognizing the strengths and weaknesses of MIMO DoF IC and SIC, we propose to have DoF IC and SIC help each other based on the following two ideas: (i) Since MIMO DoFs represent precious resources at the node, we shall exploit SIC for IC to conserve DoF resources; (ii) Since at a particular stage, SIC may fail to meet SINR threshold to successfully decode the signal, we may exploit MIMO DoF IC capability to selectively cancel a subset of interfering signals so that SIC can be successful at a receiver. In other words, we want to exploit SIC to conserve DoF resources while at the same time have DoF IC resolve the potential SINR barrier problem that SIC may encounter. The goal of this paper is to develop the mathematical models to realize these two ideas in a multi-hop wireless network.

The main contributions of this paper are as follows:

- This is the first paper that incorporates SIC under the MIMO DoF model with the goals of exploiting SIC for DoF resource conservation and employing DoF IC to selectively cancel a subset of interfering signals so that SIC can be successful to decode the desired signal. Note that our use of SIC in MIMO differs from SIC in multi-user MIMO communication (see Section 2.6) where there is no concern of exploiting MIMO's IC capability to remove the SINR threshold barrier that SIC may encounter.

- This is also the first paper that studies SIC under the MIMO DoF model for a multi-hop wireless network. In a multi-hop MIMO network, one needs to address the scheduling and flow routing problem. Coupling of both routing and scheduling with DoF allocation and SIC is clearly a nontrivial problem. Note that this problem differs from SIC in multi-user MIMO communication (see Section 2.6) where there is no concern of scheduling and routing.

- We propose a MIMO DoF IC scheme that incorporates SIC. Based on this scheme, we develop a model for DoF allocation at a transmitter and a receiver that incorporates SIC. We also develop a sequential SIC model when DoF IC is employed to cancel a subset of interfering signals that may hinder SIC from meeting its SINR decoding threshold. Both of these two models are developed for a multi-hop wireless network. Together with other scheduling and routing constraints, we develop an optimization framework for joint DoF IC and SIC in a multi-hop MIMO network.

- Through an application of the framework on a throughput maximization problem, we find that SIC and MIMO DoF IC can indeed work in harmony and help each other as we intended. When compared to a MIMO network where SIC is not used, there is a significant increase in throughput.

The remainder of this chapter is organized as follows. In Section 2.2, we review a MIMO DoF model that we shall employ in this study. We also review the basic concept of SIC for the single-antenna case. Subsequently, we describe the motivation and basic idea of this chapter, as well as the technical challenges that we will encounter. In Section 2.3, we extend the single-antenna SIC

model to MIMO with multiple data streams from each user. In Section 2.4, we propose our MIMO SIC scheme and develop SIC in the MIMO DoF model for a multi-hop network, which is the core of this paper. In Section 2.5, we present a case study for a throughput maximization problem along with a performance comparison to the same problem when SIC is not used. Section 2.6 reviews related work. Section 2.7 concludes this chapter.

2.2 SIC in MIMO: Background and Motivation

In this section, we first review a MIMO IC model that we will use in this study. Then we review the basic concept of SIC for single-antenna case. Based on this background, we motivate our idea of how SIC may be exploited to conserve DoF resources for MIMO IC.

2.2.1 MIMO DoF Model

In this section, we first review the MIMO IC model that we will employ in this study. Then we review the basic concept of SIC for a single-antenna case. Based on this background, we motivate our idea of how SIC may be exploited to conserve DoF resources for MIMO IC.

2.2.2 MIMO DoF Model

For notation, we denote vectors and matrices in bold-face lower and upper case letters, respectively. For matrix $\mathbf{G}$, $\mathbf{G}^T$ and $\mathbf{G}^\dagger$ denote transpose and Hermitian operations, respectively. $\|\mathbf{g}\|$ denotes the norm of vector $\mathbf{g}$. $\mathbf{g}^q$ denotes the q-th column of matrix $\mathbf{G}$. A diagonal matrix is denoted as $\mathrm{diag}\{\dots\}$. Table 2.1 shows the notation used in this paper.

Consider a multi-hop MIMO network consisting of a set of nodes $\mathcal{N}$ which has N elements. Each node is assumed to have M antennas. Suppose that there are L possible links in the network. Denote $\mathrm{Tx}(l)$ and $\mathrm{Rx}(l)$ as the transmit and receive nodes of link l, $1 \leq l \leq L$. We consider a time-

Table 2.1: Notation.

Symbol	Definition
M	Number of antennas at node each node
$\mathbf{H}_{ki}$	Channel fading matrix between transmitter k and receiver i
L_{ki}	Path-loss between node k and receiver node i
p_j	Transmit power of data streams from node j
$\mathbf{U}_k$	Precoding vector at transmitter node k
$\mathbf{x}_k$	Symbol vector at node k
$\mathbf{A}_k$	Diagonal transmit amplitude matrix at transmit node k
$\mathbf{n}$	Gaussian noise vector with power N_0
N_0	Noise power
$\mathbf{V}_{ji}$	Receive matrix to decode signal from node j at node i
$\mathbf{y}_{ji}$	Received signal vector at node i from node j
$\mathcal{I}_i$	Set of nodes within node i's interference range
L	Total number of links in the network
β	SIC SINR threshold
$\mathcal{L}_i^{\text{in}}$	Set of incoming links at node i
$\mathcal{L}_i^{\text{out}}$	Set of outgoing links at node i
N	Number of nodes in the network
$\mathcal{N}$	Set of nodes in the network
$\text{Rx}(l)$	Receiver of link l
$\text{Tx}(l)$	Transmitter of link l
$r(f)$	Rate of session f
$r_l(f)$	Rate for session f on link l
$w(f)$	Weight of session f
$x_i[t]$	Indicator variable to show if node i is a transmitter in time slot t
$y_i[t]$	Indicator variable to show if node i is a receiver in time slot t
$z_l(t)$	Number of data streams on link l in time slot t
$\pi_i[t]$	Node i's position in a node level ordering $\pi[t]$
$\theta_{ji}[t]$	A binary variable to indicate whether node i is placed after node j in $\pi[t]$
$\eta_{ji}[t]$	A binary variable to indicate interference from node j is canceled at i by SIC in time slot t
$\gamma_{ji}[t]$	A binary variable to indicate interference from node j is canceled at i by MIMO-IC in time slot t
$\lambda_{ji}[t]$	A binary variable to indicate intended transmit node j at least transmits one data stream to receive node i in time slot t

slotted scheduling, where a time frame consists of T time slots. Depending on link scheduling, a subset of links will be active in time slot t, $1 \leq t \leq T$.

The basic idea of DoF-based MIMO models is as follows [71]: (i) The number of available DoFs at a node is equal to the number of its antennas. (ii) For SM, both transmit and receive nodes need to consume DoFs. For each data stream, both the transmit and receive nodes need to consume one DoF. (iii) For IC, either the transmit node or the receive node may consume DoFs. If a transmit node A is to cancel its interference to a receive node B, then it needs to consume x DoFs, where x is the number of DoFs that is being sent to node B (via SM) from B's intended transmitter. Likewise, if a receive node B is to cancel the interference from a transmit node A, then it needs to consume y DoFs, where y is the number of DoFs that is being sent by transmit node A (via SM) to A's intended receiver. (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.

Half-Duplex Constraint. Although there has been significant advance on full duplex for single antenna node, there remain significant challenges to have a practical design for full duplex on a MIMO node. Therefore, we assume half duplex on a MIMO node in this paper. Denote $x_i[t]$ as a binary variable to indicate whether node $i \in \mathcal{N}$ is transmitting in time slot t, i.e., $x_i[t] = 1$ if node i is a transmitter in time slot t and 0 otherwise. Similarly, denote $y_i[t]$ as a binary variable to indicate whether node $i \in \mathcal{N}$ is receiving in time slot t, i.e., $y_i[t] = 1$ if node i is a receiver in time slot t and 0 otherwise. For half-duplex, we have the following constraint:

$$x_i[t] + y_i[t] \leq 1, \quad (1 \leq i \leq N, 1 \leq t \leq T) \,. \tag{2.2.1}$$

Node's SM Constraints. Denote $\mathcal{L}_i^{\text{in}}$ and $\mathcal{L}_i^{\text{out}}$ as the set of potential incoming and outgoing links at node i, respectively. Denote $z_{(l)}[t]$ as the number of data streams over link l. If node i is not a transmitter, then we have $\sum_{l \in \mathcal{L}_i^{\text{out}}} z_{(l)}[t] = 0$. Otherwise, the total number of outgoing streams should be positive and lesser than the number of antennas, i.e., $1 \leq \sum_{l \in \mathcal{L}_i^{\text{out}}} z_{(l)}[t] \leq M$. These two cases can be expressed in a compact form as follows:

$$x_i[t] \leq \sum_{l \in \mathcal{L}_i^{\text{out}}} z_{(l)}[t] \leq M x_i[t], \quad (1 \leq i \leq N, 1 \leq t \leq T) \,. \tag{2.2.2}$$

Similarly, depending on whether node i is an active receiver, we have the following constraint:

$$y_i[t] \leq \sum_{l \in \mathcal{L}_i^{\text{in}}} z_{(l)}[t] \leq M y_i[t], \quad (1 \leq i \leq N, 1 \leq t \leq T) \ . \tag{2.2.3}$$

Ordering Constraint. In a multi-hop MIMO network, to avoid duplication in IC while ensuring feasibility of DoF allocation, Shi *et al.* [71] introduced a novel IC scheme among the nodes based on a node ordering concept. Under this scheme, all nodes in the network are put into a logical list with the position of the node in the list representing its order. Specifically, denote $\pi[t]$ as an ordered list of nodes in the network in time slot t and denote $\pi_i[t]$ as the position of node $i \in \mathcal{N}$ in $\pi[t]$. Then we have:

$$1 \leq \pi_i[t] \leq N, \qquad (1 \leq i \leq N, 1 \leq t \leq T) \ . \tag{2.2.4}$$

To model the relative ordering between any two nodes i and j in $\pi[t]$, we define an indicator variable $\theta_{ji}[t]$ as follows:

$$\theta_{ji}[t] = \begin{cases} 1 & \text{if node } j \text{ is before node } i \text{ in } \pi[t], \\ 0 & \text{otherwise.} \end{cases}$$

Denote $\mathcal{I}_i$ as the set of nodes that are located within the interference range of transmitter i. Then the ordering relationship between any two nodes in the network can be represented by the following mathematical programming constraints [71]. For $(1 \leq i \leq N, j \in \mathcal{I}_i, 1 \leq t \leq T)$,

$$\pi_i[t] - N \cdot \theta_{ji}[t] + 1 \leq \pi_j[t] \leq \pi_i[t] - N \cdot \theta_{ji}[t] + N - 1 \ . \tag{2.2.5}$$

Based on $\pi[t]$, each node in this list has the following responsibility in IC:

- *Transmit node.* If this node is a transmit node, then it only needs to cancel its interference to those receive nodes that are before itself in the ordered node list. It does not need to consume DoFs to cancel its interference to those receive nodes that are after itself in the ordered node list. Interference from this transmit node to receive nodes after itself will be canceled by those receive nodes later. The number of DoFs consumed at this transmit node for IC is equal to the total number of desired data streams received by those receive nodes.

- *Receive node.* If this node is a receive node, then it only needs to cancel interference from those transmit nodes that are before itself in the ordered node list. It does not need to cancel interference from those transmit nodes that are after itself in the ordered node list. Interference from transmit nodes after this node will be canceled by those transmit nodes later. The number of DoFs consumed at this receive node for IC is equal to the total number of data streams transmitted by those transmit nodes.

The above IC rules can also be cast into mathematical programming constraints. Then the DoF constraint at a transmit node and a receive node for $(1 \leq i \leq N, 1 \leq t \leq T)$ can be written as follows [71]:

$$\sum_{l \in \mathcal{L}_i^{\text{out}}} z_{(l)}[t] + \sum_{j \in \mathcal{I}_i} \theta_{ji}[t] \sum_{k \in \mathcal{L}_j^{\text{in}}}^{\text{Tx}(k) \neq i} z_{(k)}[t] \leq M x_i[t] + (1 - x_i[t]) B_i, \tag{2.2.6}$$

$$\sum_{k \in \mathcal{L}_i^{\text{in}}} z_{(k)}[t] + \sum_{j \in \mathcal{I}_i} \theta_{ji}[t] \sum_{l \in \mathcal{L}_j^{\text{out}}}^{\text{Rx}(l) \neq i} z_{(l)}[t] \leq M \cdot y_i[t] + (1 - y_i[t]) B_i, \tag{2.2.7}$$

where B_i is a large constant and is no small than $\sum_{j \in \mathcal{I}_i} \theta_{ji}[t] \sum_{k \in \mathcal{L}_j^{\text{in}}}^{\text{Tx}(k) \neq i} z_{(k)}[t]$ and $\sum_{j \in \mathcal{I}_i} \theta_{ji}[t] \sum_{l \in \mathcal{L}_j^{\text{out}}}^{\text{Rx}(l) \neq i} z_{(l)}[t]$. For example, we can set $B_i = M \cdot |\mathcal{I}_i|$.

2.2.3 SIC under Single-antenna Node

SIC allows a receiver to take multiple interfering signals from different transmitters (see Fig. 2.1) and decode each one of them iteratively [81]. As shown in Fig. 2.2, for the composite received signal, the receiver attempts to decode the strongest signal and considers all other signals as interference. If the strongest signal is decoded successfully (upon meeting a certain SINR threshold), the receiver subtracts it from the original composite signal and then starts to decode the second strongest signal and so forth. The process continues until all signals are successfully decoded, or the SINR threshold cannot be satisfied at certain stage.

Without loss of generality, suppose that the power levels of the signals from the K transmitters

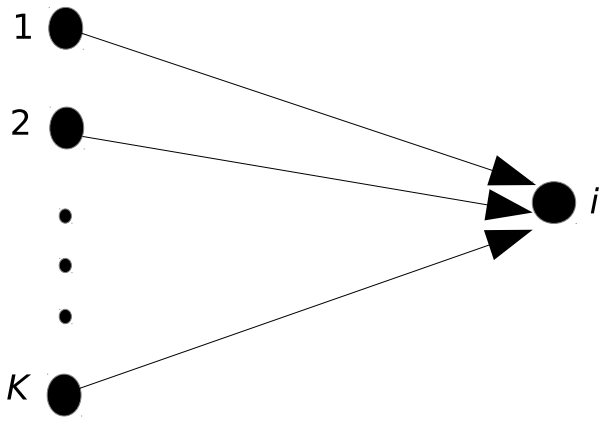


Figure 2.1: A receiver with K concurrent transmitters.

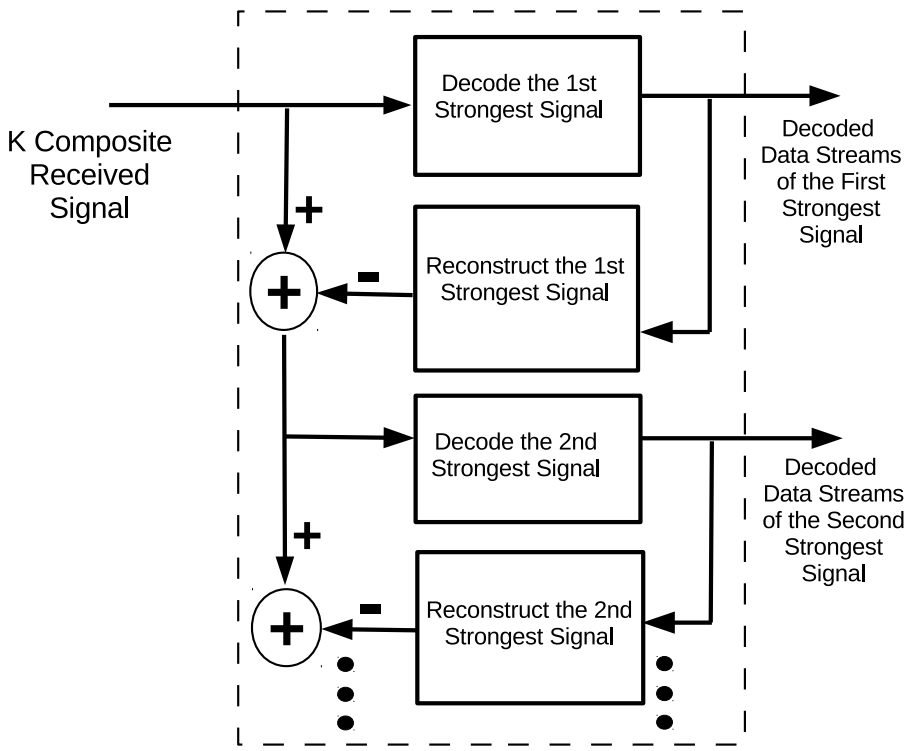


Figure 2.2: A schematic of SIC.

received at node i are in nondecreasing order as $P_{1i} \leq P_{2i} \leq \ldots \leq P_{Ki}$. Receiving node i tries to decode signal from node n in the order of $K, K-1, \ldots n$. Then, the signal with received power P_{ni} can be decoded successfully if and only if

$$\begin{aligned}
&\text{Step 1} && \frac{P_{Ki}}{\sum_{l=1}^{K-1} P_{ki} + \sigma^2} \geq \beta, \\
&\text{Step 2} && \frac{P_{(K-1)i}}{\sum_{l=1}^{K-2} P_{ki} + \sigma^2} \geq \beta, \\
&\vdots && \vdots \\
&\text{Step}(K-n+1) && \frac{P_{ni}}{\sum_{l=1}^{n-1} P_{ki} + \sigma^2} \geq \beta \ .
\end{aligned} \tag{2.2.8}$$

2.2.4 Motivation and Basic Idea

The above background on MIMO IC model and SIC model motivates us to propose the following ideas.

- **DoF IC to Remove Barrier Signal in SIC.** For SIC, at any stage when the SINR threshold is no longer satisfied at a receiver, SIC will fail to continue. This is the limitation of SIC. But with MIMO IC capability, we could use a MIMO DoF (either at this receiver or the transmitter of this signal) to cancel this particular interference without decoding it. After this impeding interfering signal is removed, SIC can resume its decoding of the remaining signals from other transmitters. As an example, in (2.2.8), suppose in Step 1, the SINR threshold β for P_{Ki} is not satisfied. With DoF IC, either the transmitter K or receiver i can use 1 DoF to cancel this interfering signal, thereby allowing SIC to continue to work on the remaining signals. At any step when the SINR threshold β is no longer satisfied for some transmitter, we can apply the same DoF IC technique and remove this barrier signal, until the desired signal n is decoded successfully.

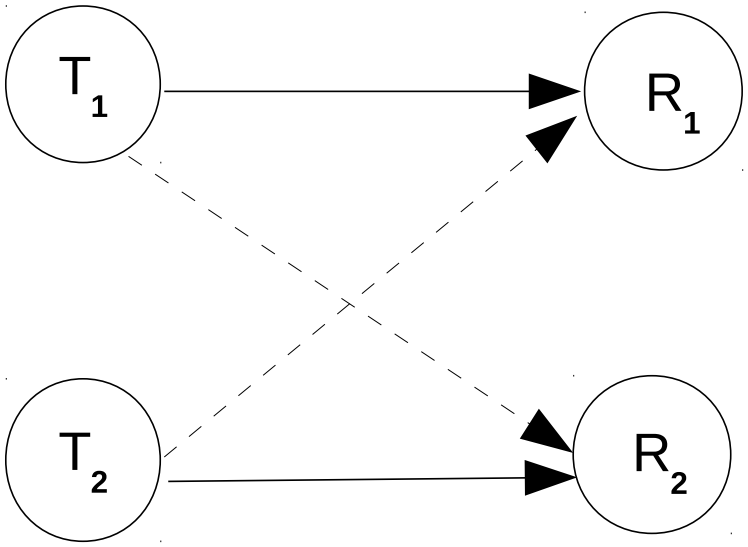


Figure 2.3: A simple example illustrating how SIC can help conserve DoFs for IC.

- **SIC to Conserve DoFs in IC.** Likewise, before we expend precious DoFs for IC at a receive node, we could exploit SIC to its fullest extent at this receive node (to decode as many concurrent signals as possible). This exploitation of SIC capability will help conserve precious DoFs at the node. As an example, consider Fig. 2.3, where (T_1, R_1) and (T_2, R_2) are two pairs of transmitting and receiving nodes in the network. Assume that all nodes share the same channel and are equipped with 4 antennas (with DoFs being 4 at each node). Suppose that both T_1 and T_2 wish to transmit 2 data streams to R_1 and R_2, respectively. For an ordered node list, say $\pi = (T_1, R_1, T_2, R_2)$, we need to expend 2, 2, 4, and 4 DoFs on T_1, R_1, T_2, R_2, respectively. Now, suppose that we employ SIC decoder at both receivers R_1 and R_2. Then it may be possible that the interference from T_2 be handled by SIC at R_1, allowing T_2 to save 2 DoFs for canceling its interference to R_1. Likewise, it may be possible that the interference from T_1 be handled by SIC at R_2, allowing R_2 to save 2 DoFs for canceling the interference from T_1. That is, when SIC is successful at R_1 and R_2, we only need to expend 2 DoFs on T_1, R_1, T_2, R_2, respectively.

The above two ideas and examples illustrate the benefits of using DoF IC and SIC jointly for interference management. The goal of this paper is to establish its theoretical foundation. To do so, it is important to understand the underlying technical challenges.

2.2.5 Technical Challenges

Although there has been active research on the MIMO DoF model (e.g. [9, 12, 36, 62, 71, 74, 94]) and SIC in wireless networks (e.g. [11, 35, 41, 52, 55, 81, 83]), they are mostly done independently, without exploiting the potential mutual benefits when used jointly to overcome each other's limitations. Although MIMO and SIC have been studied in the context of MIMO-MMSE-SIC in both the information theory (IT) and communications (COMM) communities (see, e.g., [10, 24, 25, 76]), SIC is not explicitly considered as a technique to conserve DoFs in IC; likewise, neither are DoFs explicitly used to remove large interference signal to meet SINR threshold. As a result, there is hardly any result in the literature addressing the ideas that we are proposing in this paper — bundling MIMO DoF and SIC to overcome each other's limitation in IC.

The main technical challenges that we need to address are as follows:

- The calculation of SINR at each stage of SIC requires received power from both intended and unintended transmitters. Such information is given explicitly in a single antenna SIC model. However, MIMO DoF model, which is a protocol model by nature, does not explicitly offer such receive power information. As a result, it is necessary to dig into the MIMO matrix model (inherently a physical model) and extract relevant parameters for power calculation. Such an intertwined approach in studying joint SIC and MIMO DoF models is not trivial.

- Once we know how to calculate SINR for SIC with MIMO, the next challenge is how to model SIC capability into MIMO's DoF constraints (at both transmitter and receiver). Coupling SIC with MIMO IC in mathematical programming is intrinsically complex and would call for effective reformulation techniques to ensure the tractability of the final formulation. Again, this is a challenging problem.

- Finally, instead of limiting ourselves to point-to-point or single-hop communications, we are interested in studying joint SIC and MIMO IC in the general context of a multi-hop wireless network. In a multi-hop environment, a MIMO-SIC scheme is also coupled with the upper layer scheduling and routing algorithms. These upper layer algorithms determine, in each time slot, the set of transmitters, the set of receivers, the set of links, and the number of data streams. The joint DoF-SIC scheme shall again be jointly designed with these upper layer scheduling and routing algorithms. As expected, such a mathematical formulation is intrinsically complex and usually results in a challenging problem.

2.3 SIC in MIMO

In Sections 2.3 and 2.4, we address the above problems. In this section, we extend SIC model in (2.2.8) for single-antenna system to MIMO. In the next section, we present a mathematical modeling of employing SIC in the MIMO DoF model.

2.3.1 Calculating SINR in MIMO

Consider the multiuser MIMO model in Fig. 2.4, where there are multiple transmit nodes and one receive node. We assume that nodes are symbol synchronous and each node $j \in \mathcal{I}_i$ may transmit up to M data streams. For a given symbol time, the data streams from transmit node j are denoted by a vector of symbols $\mathbf{x}_j = [x_j^1, x_j^2, \ldots, x_j^M]^T$. We assume if transmit node j has fewer data streams than M, then the remaining elements of $\mathbf{x}_j$ are filled with zeros.

The complex MIMO signal from transmitter j received at node i (after passing through a linear receiver) can be written as:

$$\mathbf{y}_{ji} = L_{ji}\mathbf{V}_{ji}^{\dagger}\mathbf{H}_{ji}^{\dagger}\mathbf{U}_j\mathbf{A}_j\mathbf{x}_j + \sum_{k\in\mathcal{I}_i, k\neq j} L_{ki}\mathbf{V}_{ji}^{\dagger}\mathbf{H}_{ki}^{\dagger}\mathbf{U}_k\mathbf{A}_k\mathbf{x}_k + \mathbf{V}_{ji}^{\dagger}\mathbf{n}_i \ , \tag{2.3.1}$$

where $\mathbf{H}_{ki} \in \mathbb{C}^{M\times M}$ is the channel matrix between transmit node k and receive node i and is

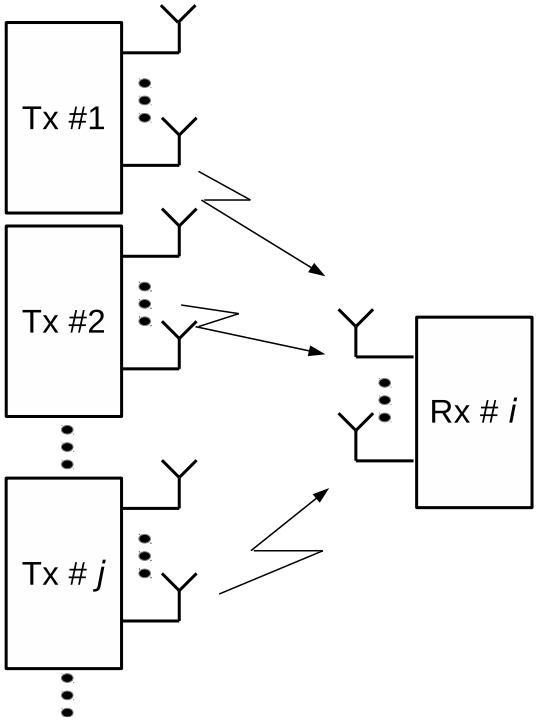


Figure 2.4: System configuration of a multiuser MIMO system

normalized to mean power 1, L_{ji} is the path-loss factor between j and i, $\mathbf{U}_k \in \mathbb{C}^{M\times M}$ is the unitary transmit precoding matrix at transmit node k,

$\mathbf{A}_j \in \mathbb{R}^{M\times M}$, $\mathbf{A}_j = \mathrm{diag}\{\sqrt{p_j}, \sqrt{p_j}, \ldots, \sqrt{p_j}\}$ is the real-valued diagonal transmit amplitude matrix. $\mathbf{V}_{ji} \in \mathbb{C}^{M\times M}$ is the unitary receive matrix at node i for decoding data streams from node j, and $\mathbf{n}_i \in \mathbb{C}^{M\times 1}$ is the white Gaussian noise vector with variance N_0 per element.

Depending on the transmit precoder and receiver matrices, SINR can be calculated [77, 86]. Assuming data streams are uncorrelated, SINR for the q-th element in $\mathbf{y}_{ji}$ is:

$$\mathrm{SINR}_{ji}^{q} = \frac{p_j \cdot L_{ji}^2 \|\mathbf{v}_{ji}^{q\,\dagger}\mathbf{H}_{ji}^{\dagger}\mathbf{u}_j^q\|^2}{\sum\limits_{k\in\mathcal{I}_i, k\neq j} p_k L_{ki}^2 \|\mathbf{v}_{ji}^{q\,\dagger}\mathbf{H}_{ki}^{\dagger}\mathbf{U}_k\|^2 + N_0 ||\mathbf{v}_{ji}^q||^2} .$$

Since $\mathbf{V}_{ji}$ is a unitary matrix, we have $\|\mathbf{v}_{ji}^q\|^2 = 1$. Therefore,

$$\mathrm{SINR}_{ji}^{q} = \frac{p_j \cdot L_{ji}^2 \|\mathbf{v}_{ji}^{q\,\dagger}\mathbf{H}_{ji}^{\dagger}\mathbf{u}_j^q\|^2}{\sum\limits_{k\in\mathcal{I}_i, k\neq j} p_k L_{ki}^2 \|\mathbf{v}_{ji}^{q\,\dagger}\mathbf{H}_{ki}^{\dagger}\mathbf{U}_k\|^2 + N_0} . \tag{2.3.2}$$

2.3.2 Sequential Constraints for MIMO-SIC

In (2.2.8), we showed a mathematical model for SIC in a single-antenna system. For MIMO, there are usually multiple data streams from a transmit node. Performing SIC at data stream level across different transmit nodes is not compatible to the MIMO DoF model, as the latter is intrinsically a node-based model. Instead, we choose to decode *aggregate* data streams on a per transmit node level in this paper. That is, we use the minimum SINR among all data streams from the same transmitter as the worst-case aggregate SINR for this transmit node. Note that there is no interference among data streams from node j to node i due to SM. Thus, if the worst-case (with the smallest SINR) data stream from node j is decodable, all other data streams from node j must also be decodable. Therefore, all data streams from transmit node j are decodable at receive node i if

$$\frac{p_j \cdot L_{ji}^2 \cdot \min_q \|\mathbf{v}_{ji}^{q\,\dagger}\mathbf{H}_{ji}^{\dagger}\mathbf{u}_j^q\|^2}{\sum\limits_{k\in\mathcal{I}_i,k\neq j} p_k L_{ki}^2 \cdot \max_q \|\mathbf{v}_{ji}^{q\,\dagger}\mathbf{H}_{ki}^{\dagger}\mathbf{U}_k\|^2 + N_0} \geq \beta \ . \tag{2.3.3}$$

Denote $C_{ji} = \min_q \|\mathbf{v}_{ji}^{q\,\dagger}\mathbf{H}_{ji}^{\dagger}\mathbf{u}_j^q\|^2$ and $D_{jki} = \max_q \|\mathbf{v}_{ji}^{q\,\dagger}\mathbf{H}_{ki}^{\dagger}\mathbf{U}_k\|^2$. To have SIC operate on the node level with aggregate data streams, without loss of generality, suppose that the minimum received power levels of the data streams from each of the K transmit nodes at node i are listed in non-decreasing order as $p_1 \cdot L_{1i}^2 \cdot C_{1i} \leq \ldots \leq p_n \cdot L_{ni}^2 \leq \ldots \leq p_K \cdot L_{Ki}^2 \cdot C_{Ki}$, where $p_n \cdot L_{ni}^2$ corresponds to the minimum received power from intended transmit node n while the others correspond to minimum received power of unintended transmit nodes. Based on SIC, receiver i will decode the signals in the order of $K, K-1, \ldots n$ (i.e., the strongest signal first, until the intended transmit node n, inclusive). That is, the set of data streams from intended transmitter

node n is decodable if

$$\begin{aligned}
&\text{Step 1} && \frac{p_K \cdot L_{Ki}^2 \cdot C_{Ki}}{\sum_{k=1}^{K-1} p_k L_{ki}^2 \cdot D_{Kki} + N_0} \geq \beta, \\
&\text{Step 2} && \frac{p_{(K-1)} \cdot L_{(K-1)i}^2 \cdot C_{(K-1)i}}{\sum_{k=1}^{K-2} p_k L_{ki}^2 \cdot D_{(K-1)ki} + N_0} \geq \beta, \\
&\quad \vdots && \quad \vdots \\
&\text{Step } (K-n+1) && \frac{p_n \cdot L_{ni}^2}{\sum_{k=1}^{n-1} p_k L_{ki}^2 \cdot D_{nki} + N_0} \geq \beta \ ,
\end{aligned} \tag{2.3.4}$$

where for intended transmit node n, $C_{ni} = 1$ due to SM requirements.

Note that although (2.3.4) shows the iterative SIC process, it is not written in a mathematical program. To address this issue, we adopt a similar approach as in [41] by defining the so-called *residual SINR* (or r-SINR). r-SINR is a compact expression to calculate SINR value for the transmit-receive pair under consideration after all the transmit nodes with stronger received signals have been successfully decoded. Specifically, for the aggregate data streams from transmit node j to receive node i in time slot t, we define r-SINR$_{ji}[t]$ as follows:

$$\text{r-SINR}_{ji}[t] = \frac{p_j \cdot L_{ji}^2 \cdot C_{ji}}{\sum\limits_{k \in \mathcal{I}_i, k \neq j}^{p_k L_{ki}^2 \cdot C_{ki} \leq p_j L_{ji}^2 \cdot C_{ji}} p_k \cdot L_{ki}^2 \cdot D_{jki} + N_0} \ , \tag{2.3.5}$$

where the summation in the denominator includes all transmit nodes k with weaker received signals than j.

2.4 SIC in MIMO DOF Model

A receive node may receive signals from multiple transmit nodes, including both signals from the intended transmit node and interference from any unintended transmit node. In the proposed scheme, receive node i divides the signals from these transmit nodes into five sets:

Set 1 Signals from unintended transmit nodes that are canceled by DoFs at the transmit nodes;

Set 2 Signals from unintended transmit nodes that are canceled by DoFs at receive node i;

Set 3 Signals from unintended transmit nodes that are decoded and subtracted from the composite received signals by SIC at receive node i before the intended transmit node (i.e., the received powers from these transmit nodes are greater than the powers from the intended transmit node n);

Set 4 Signals from intended transmit node n;

Set 5 Signals from unintended transmit nodes that are treated as noise during SIC at receive node i (i.e., the received powers from these transmit nodes are less than the powers from the intended transmit node n).

For the signals from an unintended transmitter to receive node i, the question of which sets (1, 2, 3, or 5) the signals belong to will be solved by an optimization problem. The goal of this section is to define and formulate the decision variables into the necessary constraints for DoF IC and SIC. Figure 2.5 shows how the signals from the intended transmitter (Set 4) are successfully decoded at receive node i. First, set 1 signals are canceled at the transmitter side. The remaining composite signals from sets 2, 3, 4 and 5 are received at Rx node i. As shown in the figure, Rx node has one reconfigurable receive matrix, which is updated iteratively during SIC. In each SIC iteration (except the last iteration), Rx node performs DoF IC and SIC as follows. In the first iteration, Rx node configures its receive matrix to perform DoF IC for signals in set 2. Then it decodes the strongest received signal (minimum received power among the data streams from a transmit node) in the union of sets 3, 4 and 5 by SM while treating the remaining signals in sets 3, 4 and 5 as noise. The decoded signals are reconstructed and subtracted from the composite signal before the next iteration. The process goes on from iteration to iteration. In the last iteration, the Rx node will decode the signals from the intended transmitter (set 4) while treating the signals in set 5 as noise. The output from the last iteration are the signals from the intended transmitter. In the rest of this

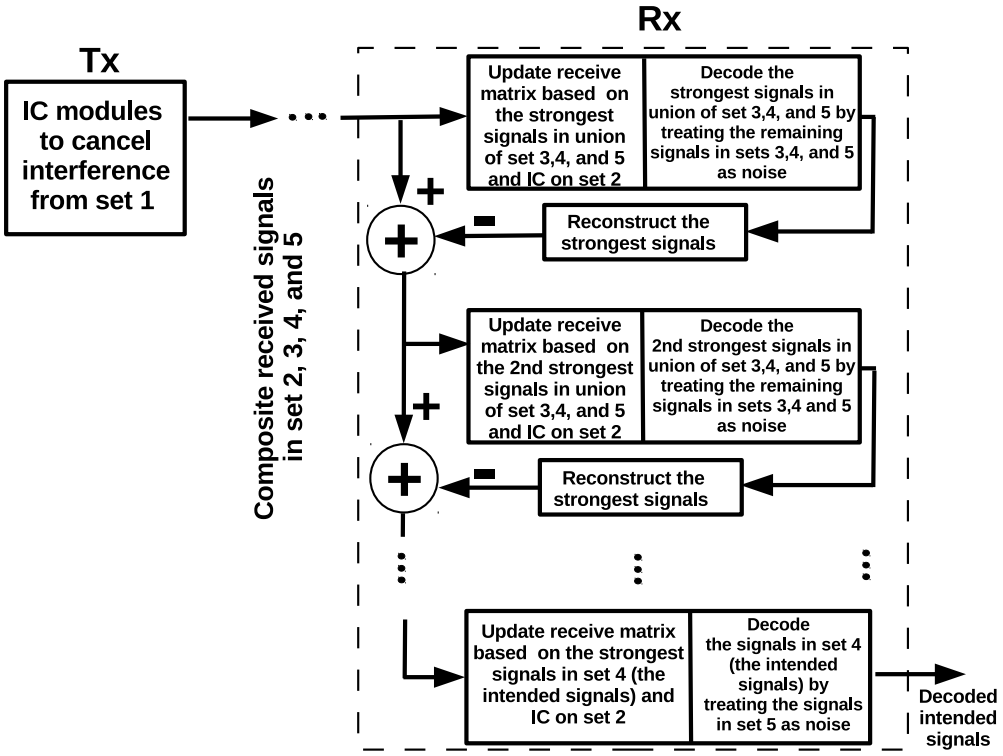


Figure 2.5: A schematic of the proposed DoF-SIC scheme.

section, we show mathematically how the DoF IC model and SIC model can be coupled together. First, we need to introduce some notations.

Indicator Variables for DoF IC and SIC. For the 5 sets of signals, we define three binary indicator variables $\gamma_{ji}[t]$ (for sets 1 and 2), $\eta_{ji}[t]$ (for sets 3 and 5) and $\lambda_{ji}[t]$ (for set 4) as follows:

- $\gamma_{ji}[t]$: a binary variable. $\gamma_{ji}[t] = 1$ if the interference from unintended transmit node j to receive node i is canceled by DoF (either at transmit node j or receive node i), and 0 otherwise. Note that when $\gamma_{ji}[t] = 1$, it does not tell which node does the IC with DoF (transmit node j or receive node i). For that, we need the value of $\theta_{ji}[t]$. Also note that, if $\gamma_{ji}[t] = 1$, then we have $x_j[t] = 1$ and $y_i[t] = 1$. This sufficient condition can be modeled by the following constraints:

$$x_j[t] \geq \gamma_{ji}[t],\ (1 \leq i \leq N, j \in \mathcal{I}_i, 1 \leq t \leq T)\ , \tag{2.4.1}$$

$$y_i[t] \geq \gamma_{ji}[t],\ (1 \leq i \leq N, j \in \mathcal{I}_i, 1 \leq t \leq T)\ . \tag{2.4.2}$$

- $\eta_{ji}[t]$: a binary variable. $\eta_{ji}[t] = 1$ if the interference from unintended transmit node j to receive node i is canceled by SIC (or being treated as noise), and 0 otherwise. Also note that, if $\eta_{ji}[t] = 1$, then we have $x_j[t] = 1$ and $y_i[t] = 1$. This sufficient condition can be modeled by the following constraints:

$$x_j[t] \geq \eta_{ji}[t],\ (1 \leq i \leq N, j \in \mathcal{I}_i, 1 \leq t \leq T)\ , \tag{2.4.3}$$

$$y_i[t] \geq \eta_{ji}[t],\ (1 \leq i \leq N, j \in \mathcal{I}_i, 1 \leq t \leq T)\ . \tag{2.4.4}$$

  Also note that, if $x_j[t] = y_i[t] = 1$, then we have $\eta_{ji}[t] + \gamma_{ji}[t] = 1$. This sufficient condition can be modeled by the following constraints:

$$x_j[t] + y_i[t] - 1 \leq \eta_{ji}[t] + \gamma_{ji}[t] \leq 1, \quad (1 \leq i \leq N, j \in \mathcal{I}_i, 1 \leq t \leq T)\ . \tag{2.4.5}$$

- $\lambda_{ji}[t]$: a binary variable. $\lambda_{ji}[t] = 1$ if intended transmit node j transmits *at least* one data stream successfully to receive node i via SM, and 0 otherwise. For SM, we have:

$$\lambda_{ji}[t] \leq z_l[t] \leq M \cdot \lambda_{ji}[t], \quad (1 \leq l \leq L, j = \mathrm{Tx}(l), i = \mathrm{Rx}(l), 1 \leq t \leq T)\ . \tag{2.4.6}$$

Coupling SIC with MIMO DoF Model. With the above definitions for $\gamma_{ji}[t]$, $\eta_{ji}[t]$ and $\lambda_{ji}[t]$, the DoF consumption constraints in (2.2.6) and (2.2.7) for a transmit node and a receive node can be extended by taking into account SIC as follows. When node i is a transmit node, then the DoF consumption at this node must satisfy for $(1 \leq i \leq N, 1 \leq t \leq T)$:

$$\sum_{l \in \mathcal{L}_i^{\mathrm{out}}} z_{(l)}[t] + \sum_{j \in \mathcal{I}_i} \theta_{ji}[t]\gamma_{ij}[t] \sum_{k \in \mathcal{L}_j^{\mathrm{in}}}^{\mathrm{Tx}(k) \neq i} z_{(k)}[t] \leq M x_i[t] + (1 - x_i[t]) B_i\ . \tag{2.4.7}$$

Note that in the above expression, the use of $\gamma_{ij}[t]$ limits the accounting of DoFs (used in IC) only to those interfering data streams that are canceled by transmit node i. Similarly, when node i is a receive node, then the DoF consumption at this node must satisfy for $(1 \leq i \leq N, 1 \leq t \leq T)$:

$$\sum_{k \in \mathcal{L}_i^{\mathrm{in}}} z_{(k)}[t] + \sum_{j \in \mathcal{I}_i} \theta_{ji}[t]\gamma_{ji}[t] \sum_{l \in \mathcal{L}_j^{\mathrm{out}}}^{\mathrm{Rx}(l) \neq i} z_{(l)}[t] \leq M y_i[t] + (1 - y_i[t]) B_i\ . \tag{2.4.8}$$

Sequential SIC Model with DoF IC. In Section 2.3, we developed MIMO SIC constraint (2.3.5) without DoF IC consideration. With DoF IC, the potential barrier signals that do not meet threshold β can now be removed and SIC can continue to decode. We incorporate DoF IC into the r-SINR definition in (2.3.5) through the $\eta_{ji}[t]$ and $\lambda_{ji}[t]$ variables, which allows us to account for only those interfering signals that are to be handled by SIC (i.e., not canceled by DoF IC). So r-SINR$_{ji}[t]$ can be re-defined as follows:

$$\text{r-SINR}_{ji}[t] = \frac{p_j \cdot L_{ji}^2 \cdot C_{ji}}{\substack{p_k L_{ki}^2 \cdot C_{ki} \le p_j L_{ji}^2 \cdot C_{ji} \\ \sum\limits_{k \in \mathcal{I}_i, k \neq j, \eta_{ki}[t]=1 \text{or} \lambda_{ki}[t]=1} p_k \cdot L_{ki}^2 \cdot D_{jki} + N_0}} . \tag{2.4.9}$$

When j is the intended transmit node, i.e., $j = n$, $C_{ni} = 1$ due to SM, r-SINR$_{ni}[t]$ is:

$$\text{r-SINR}_{ni}[t] = \frac{p_n \cdot L_{ni}^2}{\substack{p_k L_{ki}^2 \cdot C_{ki} \le p_n L_{ni}^2 \\ \sum\limits_{k \in \mathcal{I}_i, k \neq n, \eta_{ki}[t]=1} p_k \cdot L_{ki}^2 \cdot D_{nki} + N_0}} . \tag{2.4.10}$$

Note that if $\lambda_{ni}[t] = 1$ (i.e., we have at least one data stream from intended transmit node n to receive node i), then we must have:

- (i) The r-SINR$_{ji}$'s of all stronger received signals from other transmit nodes j with $\eta_{ji}[t] = 1$ are no less than the SINR threshold β; and
- (ii) The r-SINR$_{ni}[t]$ of the intended signals from node n to node i is no less than the SINR threshold β.

That is, if $\lambda_{ni}[t] = 1$, we have

$$\text{r-SINR}_{ji}[t] \ge \beta, \quad (1 \le i \le N, j \in \mathcal{I}_i, \eta_{ji}[t] = 1, p_j L_{ji}^2 C_{ji} > p_n L_{ni}^2, 1 \le t \le T) , \tag{2.4.11}$$

$$\text{r-SINR}_{ni}[t] \ge \beta, \quad (1 \le i \le N, n \in \mathcal{I}_i, 1 \le t \le T) . \tag{2.4.12}$$

Note that r-SINR$_{ji}[t]$ and r-SINR$_{ni}[t]$ in above constraints refers to definitions (2.4.9) and (2.4.10), respectively.

2.5 Case Study: A Throughput Maximization Problem

In Sections 2.2 to 2.4, we established key models for MIMO-SIC. In this section, we show how these models can be used to study networking problems. Let's consider a typical throughput maximization problem in a multi-hop MIMO network. Suppose there is a set of active sessions $\mathcal{F}$. Denote $r(f)$ as the rate of session $f \in \mathcal{F}$ and $r_{\min}$ as the minimum session rate, i.e., $r_{\min} = \min_{f\in\mathcal{F}} r(f)$. Our objective is to maximize the minimum session rate $r_{\min}$ among all sessions $\mathcal{F}$.

To formulate this problem, we need to have flow routing constraints and link capacity constraints, in addition to those constraints in Sections 2.2 to 2.4.

Flow Routing Constraints. Denote $r_l(f)$ as the amount of data rate on link l that is attributed to session $f \in \mathcal{F}$. Denote $s(f)$ and $d(f)$ as the source and destination nodes of session $f \in \mathcal{F}$, respectively. Then at source node, $s(f)$, $f \in \mathcal{F}$, we have the following flow balance:

$$\sum_{l\in\mathcal{L}_i^{\text{out}}} r_l(f) = r(f), \quad (i = s(f), f \in \mathcal{F}) \ . \tag{2.5.1}$$

At any intermediate relay node, we have

$$\sum_{l\in\mathcal{L}_i^{\text{in}}} r_l(f) = \sum_{l\in\mathcal{L}_i^{\text{out}}} r_l(f), (1 \le i \le N, i \ne s(f), i \ne d(f), f \in \mathcal{F}) \ . \tag{2.5.2}$$

At a destination node, we have

$$\sum_{l\in\mathcal{L}_i^{\text{in}}} r_l(f) = r(f), \quad (i = d(f), f \in \mathcal{F}) \ . \tag{2.5.3}$$

It can be easily verified that if (2.5.1) and (2.5.2) are satisfied, then (2.5.3) is also satisfied. Therefore, it is sufficient to have (2.5.1) and (2.5.2).

Link Capacity Constraints. For simplicity, we assume the granularity of the data rate is DoF per time slot. Since the aggregate data rate on link l cannot exceed the link's average rate, we have

$$\sum_{f\in\mathcal{F}} r_l(f) \le \frac{1}{T}\sum_{t=1}^{T} z_{(l)}[t], \quad (1 \le l \le L) \tag{2.5.4}$$

where the right-hand-side represents the average throughput on link l over a frame (T time slots).

Putting all the constraints together, we have the following formulation for the throughput maximization problem:

$$
\begin{aligned}
\textbf{TMP} \quad \max \quad & r_{\min} \\
\text{s.t.} \quad & r_{\min} \le r(f) \qquad (f \in \mathcal{F}); \\
& \text{Half duplex constraint: (2.2.1);} \\
& \text{Node's SM constraints: (2.2.2), (2.2.3);} \\
& \text{Node ordering constraints: (2.2.4), (2.2.5);} \\
& \text{DoF consumption with SIC:} \\
& \text{(2.4.1)–(2.4.5),(2.4.7)–(2.4.8);} \\
& \text{Sequential SIC with IC: (2.4.6),(2.4.11), (2.4.12) ;} \\
& \text{Flow balance constraints: (2.5.1), (2.5.2);} \\
& \text{Link capacity constraints: (2.5.4).}
\end{aligned}
$$

In the formulation, $M, N, T, B_i, p_j, L_{ji}^2, \beta,\ N_0,\ C_{ji}$ and D_{jki} are constants[1] and $x_i[t]$, $y_i[t]$, $z_{(l)}[t]$, $\pi_i[t]$, $\theta_{ji}[t]$, $\eta_{ji}[t]$, $\gamma_{ji}[t]$,$\lambda_{ji}[t]$, $r_l(f)$, $r(f)$ are variables. Through reformulation on (2.4.7), (2.4.8), (2.4.11) and (2.4.12) (see appendix), TMP can be reformulated into a mixed integer linear program (MILP). Although the theoretical worst-case complexity to a general MILP problem is exponential [28, 68], there exist highly efficient heuristics (e.g., sequential fixing algorithm [38, Chapter 10]) to solve it. Another approach is to apply an off-the-shelf solver (CPLEX [98]), which we find can handle up to a moderate-sized network successfully. Since the main goal of this paper is to explore DoF IC and SIC jointly, it is sufficient to demonstrate our results with moderate-sized networks. Therefore, we will use CPLEX to solve MILP.

[1]For the purpose of this paper, we set C_{ji} to its average value over a large number of realizations and D_{jki} to its worst case bound.

2.5.1 Reformulation

Reformulation of (2.4.7) and (2.4.8). First, we introduce binary variables $\kappa_{ji}[t]$ and $\beta_{ji}[t]$ to replace $\theta_{ji}[t] \cdot \gamma_{ij}[t]$ and $\theta_{ji}[t] \cdot \gamma_{ji}[t]$. That is, $\kappa_{ji}[t] = \theta_{ji}[t] \cdot \gamma_{ij}[t]$ and $\beta_{ji}[t] = \theta_{ji}[t] \cdot \gamma_{ji}[t]$. This change of variables will introduce the following new constraints for $\kappa_{ji}[t]$ and $\beta_{ji}[t]$:

$$\kappa_{ji}[t] \geq \theta_{ji}[t] + \gamma_{ij}[t] - 1, (1 \leq i \leq N, j \in \mathcal{I}_i, 1 \leq t \leq T), \tag{2.5.5}$$

$$\theta_{ji}[t] \geq \kappa_{ji}[t], (1 \leq i \leq N, j \in \mathcal{I}_i, 1 \leq t \leq T), \tag{2.5.6}$$

$$\gamma_{ij}[t] \geq \kappa_{ji}[t], (1 \leq i \leq N, j \in \mathcal{I}_i, 1 \leq t \leq T), \tag{2.5.7}$$

$$\beta_{ji}[t] \geq \theta_{ji}[t] + \gamma_{ji}[t] - 1, (1 \leq i \leq N, j \in \mathcal{I}_i, 1 \leq t \leq T), \tag{2.5.8}$$

$$\theta_{ji}[t] \geq \beta_{ji}[t], (1 \leq i \leq N, j \in \mathcal{I}_i, 1 \leq t \leq T), \tag{2.5.9}$$

$$\gamma_{ji}[t] \geq \beta_{ji}[t], (1 \leq i \leq N, j \in \mathcal{I}_i, 1 \leq t \leq T). \tag{2.5.10}$$

Now we can rewrite constraints (2.4.7) and (2.4.8) for $(1 \leq i \leq N, 1 \leq t \leq T)$ as

$$\sum_{l \in \mathcal{L}_i^{\text{out}}} z_{(l)}[t] + \sum_{j \in \mathcal{I}_i} \kappa_{ji}[t] \sum_{k \in \mathcal{L}_j^{\text{in}}}^{\text{Tx}(k) \neq i} z_{(k)}[t] \leq M \cdot x_i[t] + (1 - x_i[t]) B_i \ . \tag{2.5.11}$$

$$\sum_{k \in \mathcal{L}_i^{\text{in}}} z_{(k)}[t] + \sum_{j \in \mathcal{I}_i} \beta_{ji}[t] \sum_{l \in \mathcal{L}_j^{\text{out}}}^{\text{Rx}(l) \neq i} z_{(l)}[t] \leq M \cdot y_i[t] + (1 - y_i[t]) B_i \ . \tag{2.5.12}$$

Note that we still have nonlinear terms in (2.5.11) and (2.5.12), i.e., $\kappa_{ji}[t] \sum_{k \in \mathcal{L}_j^{\text{in}}}^{\text{Tx}(k) \neq i} z_{(k)}[t]$ and $\beta_{ji}[t] \sum_{l \in \mathcal{L}_j^{\text{out}}}^{\text{Rx}(l) \neq i} z_{(l)}[t]$. To reformulate these nonlinear terms, we again introduce new variables and adding new constraints. Specifically, we define new integer variable $\psi_{ji}[t] = \kappa_{ji}[t] \cdot \sum_{k \in \mathcal{L}_j^{\text{in}}}^{\text{Tx}(k) \neq i} z_{(l)}[t]$. Then (2.5.11) can be rewritten as

$$\sum_{i \in \mathcal{L}_i^{\text{out}}} z_{(l)}[t] + \sum_{j \in \mathcal{I}_i} \psi_{ji}[t] \leq M \cdot x_i[t] + (1 - x_i[t]) B_i \quad (1 \leq i \leq N, 1 \leq t \leq T) \ , \tag{2.5.13}$$

along with new constraints for $\psi_{ji}[t]$ for $(1 \leq i \leq N, j \in \mathcal{I}_i, 1 \leq t \leq T)$.

$$\psi_{ji}[t] \leq \sum_{k \in \mathcal{L}_j^{\text{in}}}^{\text{Tx}(k) \neq i} z_{(k)}[t] \ , \tag{2.5.14}$$

$$\psi_{ji}[t] \leq M \cdot \kappa_{ji}[t] \ , \tag{2.5.15}$$

$$\psi_{ji}[t] \geq M \cdot \kappa_{ji}[t] + \sum_{k \in \mathcal{L}_j^{\text{in}}}^{\text{Tx}(k) \neq i} z_{(k)}[t] - M \ . \tag{2.5.16}$$

Similarly, for (2.5.12), we define new variable $\epsilon_{ji}[t] = \beta_{ji}[t] \sum_{l \in \mathcal{L}_j^{\text{out}}}^{\text{Rx}(l) \neq i} z_{(l)}[t]$. Then (2.5.12) can be rewritten as:

$$\sum_{i \in \mathcal{L}_i^{\text{in}}} z_{(l)}[t] + \sum_{j \in \mathcal{I}_i} \epsilon_{ji}[t] \leq M \cdot y_i[t] + (1 - y_i[t]) B_i \quad (1 \leq i \leq N, 1 \leq t \leq T) \ , \tag{2.5.17}$$

along with new constraints for $\epsilon_{ji}[t]$ for $(1 \leq i \leq N, j \in \mathcal{I}_i, 1 \leq t \leq T)$,

$$\epsilon_{ji}[t] \leq \sum_{l \in \mathcal{L}_j^{\text{out}}}^{\text{Rx}(l) \neq i} z_{(l)}[t] \ , \tag{2.5.18}$$

$$\epsilon_{ji}[t] \leq M \cdot \beta_{ji}[t] \ , \tag{2.5.19}$$

$$\epsilon_{ji}[t] \geq M \cdot \beta_{ji}[t] + \sum_{l \in \mathcal{L}_j^{\text{out}}}^{\text{Rx}(l) \neq i} z_{(l)}[t] - M \ . \tag{2.5.20}$$

Reformulation of (2.4.11) and (2.4.12). The two sets of constraints in (2.4.11) and (2.4.12) are stated in the form of sufficient conditions rather than mathematical programming. To reformulate both, we first move $\eta_{ji}[t] = 1$ out of the range in (2.4.11) by treating it as part of the sufficient condition. That is, if ($\eta_{ji}[t] = 1$ *and* $\lambda_{ni}[t] = 1$) then r-SINR$_{ji}[t] \geq \beta$ for $(1 \leq i \leq N, j \in \mathcal{I}_i, p_j L_{ji}^2 \cdot C_{ji} > p_n L_{ni}^2, 1 \leq t \leq T)$. To combine $\eta_{ji}[t] = 1$ and $\lambda_{ni}[t] = 1$ into one condition, we introduce a binary variable $\delta_{(ji),(ni)}[t]$, where $\delta_{(ji),(ni)}[t] = 1$ *if and only if* ($\eta_{ji}[t] = 1$ *and* $\lambda_{ni}[t] = 1$) for $(1 \leq i \leq N, (n, j) \in \mathcal{I}_i, p_j L_{ji}^2 \cdot C_{ji} > p_n L_{ni}^2, 1 \leq t \leq T)$. This logical condition can be expressed in mathematical form as following:

$$\delta_{(ji),(ni)}[t] \geq \eta_{ji}[t] + \lambda_{ni}[t] - 1 \ , \tag{2.5.21}$$

$$\eta_{ji}[t] \geq \delta_{(ji),(ni)}[t] \ , \tag{2.5.22}$$

$$\lambda_{ni}[t] \geq \delta_{(ji),(ni)}[t] \ . \tag{2.5.23}$$

Now, we can re-write MIMO SIC sequential SINR constraints derived in (2.4.11) and (2.4.12) based on the above newly defined variables and substituting r-SINR definitions for intended and unintended transmissions in (2.4.10) and (2.4.9), respectively. For $(1 \leq i \leq N,\ (n,j) \in \mathcal{I}_i, p_j L_{ji}^2 \cdot C_{ji} > p_n L_{ni}^2, 1 \leq t \leq T)$,

$$\text{if } \delta_{(ji),(ni)}[t] = 1 \text{ then } \frac{p_j \cdot L_{ji}^2 \cdot C_{ji}}{\sum_{k \in \mathcal{I}_i, k \neq j, \eta_{ki}[t]=1 \text{ or } \lambda_{\mathrm{ki}}[\mathrm{t}]=1}^{p_k L_{ki}^2 \cdot C_{ki} \leq p_j L_{ji}^2 \cdot C_{ji}} p_k \cdot L_{ki}^2 \cdot D_{jki} + N_0} \geq \beta\ , \tag{2.5.24}$$

and for $(1 \leq i \leq N, n \in \mathcal{I}_i, 1 \leq t \leq T)$,

$$\text{if } \lambda_{ni}[t] = 1 \text{ then } \frac{p_n L_{ni}^2}{\sum_{k \in \mathcal{I}_i, k \neq n, \eta_{ki}=1}^{p_k L_{ki}^2 \cdot C_{ki} \leq p_n L_{ni}^2} p_k \cdot L_{ki}^2 \cdot D_{nki} + N_0} \geq \beta\ . \tag{2.5.25}$$

The logical constraints (2.5.24) and (2.5.25) can now be reformulated into mathematical form. For $(1 \leq i \leq N,\ (n,j) \in \mathcal{I}_i, p_j L_{ji}^2 \cdot C_{ji} > p_n L_{ni}^2, 1 \leq t \leq T)$,

$$\frac{\delta_{(ji),(ni)}[t] \cdot p_j \cdot L_{ji}^2 \cdot C_{ji} + (1 - \delta_{(ji),(ni)}[t]) \cdot G'}{\sum_{k \in \mathcal{I}_i, k \neq j}^{p_k L_{ki}^2 \cdot C_{ki} \leq p_j L_{ji}^2 \cdot C_{ji}} p_k \cdot L_{ki}^2 \cdot \eta_{ki}[t] \cdot D_{jki} + \lambda_{ni}[t] \cdot p_n \cdot L_{ni}^2 \cdot D_{jni} + N_0} \geq \beta\ , \tag{2.5.26}$$

and for $(1 \leq i \leq N, n \in \mathcal{I}_i, 1 \leq t \leq T)$,

$$\frac{\lambda_{ni}[t] \cdot p_j \cdot L_{ji}^2 + (1 - \lambda_{ni}[t]) \cdot G}{\sum_{k \in \mathcal{I}_i k \neq n}^{p_k L_{ki}^2 \cdot C_{ki} \leq p_n L_{ni}^2} p_k \cdot L_{ki}^2 \cdot \eta_{ki}[t] \cdot D_{nki} + N_0} \geq \beta\ . \tag{2.5.27}$$

where G' is an upper bound of $\beta \cdot (\sum_{k \in \mathcal{I}_i, k \neq j}^{p_k L_{ki}^2 \cdot C_{ki} \leq p_j L_{ji}^2 \cdot C_{ji}} \eta_{ki}[t] \cdot p_k \cdot L_{ki}^2 \cdot D_{jki} + \lambda_{ni}[t] \cdot p_n \cdot L_{ni}^2 \cdot D_{jni} + N_0)$ to ensure that the constraint holds whenever $\delta_{(ji),(ni)}[t] = 0$. Define $G' = \beta \cdot (\sum_{k \in \mathcal{I}_i, k \neq j} p_k \cdot L_{ki}^2 \cdot D_{jki} + N_0)$. Then $G' \geq \beta \cdot (\sum_{k \in \mathcal{I}_i, k \neq j}^{p_k L_{ki}^2 \cdot C_{ki} \leq p_j L_{ji}^2 \cdot C_{ji}} \eta_{ki}[t] \cdot p_k \cdot L_{ki}^2 \cdot D_{jki} + \lambda_{ni}[t] \cdot p_n \cdot L_{ni}^2 \cdot D_{jni} + N_0)$. Similarly, G is an upper bound of $\beta \cdot (\sum_{k \in \mathcal{I}_i, k \neq n}^{p_k L_{ki}^2 \cdot C_{ki} \leq p_n L_{ni}^2} \eta_{ki}[t] \cdot p_k \cdot L_{ki}^2 \cdot D_{nki} + N_0)$ to ensure that the constraint holds whenever $\lambda_{ni}[t] = 0$. Define $G = \beta \cdot (\sum_{k \in \mathcal{I}_i, k \neq n} p_k \cdot L_{ki}^2 \cdot D_{nki} + N_0)$. Then $G \geq \beta \cdot (\sum_{k \in \mathcal{I}_i, k \neq n}^{p_k L_{ki}^2 \cdot C_{ki} \leq p_n L_{ni}^2} \eta_{ki}[t] \cdot p_k \cdot L_{ki}^2 \cdot D_{nki} + N_0)$.

In summary we replace (2.4.7), (2.4.8), (2.4.11), and (2.4.12) with (2.5.5)–(2.5.10), (2.5.13)–(2.5.16), (2.5.17)–(2.5.20), (2.5.26), and (2.5.27) in the original formulation **TMP**. The resulting optimization problem which we denote **R-TMP**, can be written as

R-TMP $\max \quad r_{\min}$

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

Half duplex constraint: (2.2.1);

Node activity constraints: (2.2.2), (2.2.3);

Node ordering constraints: (2.2.4), (2.2.5);

DoF consumption with SIC: (2.4.1)–(2.4.5), (2.5.5)-(2.5.10), (2.5.13)-(2.5.20);

Sequential SIC with IC: (2.4.6), (2.5.21)-(2.5.23), (2.5.26),(2.5.27);

Flow balance constraints: (2.5.1), (2.5.2);

Link capacity constraints: (2.5.4);

Variables: $x_i[t], y_i[t], z_{(l)}[t], \pi_i[t], \theta_{ji}[t], \eta_{ji}[t], \gamma_{ji}[t], \lambda_{ji}[t],$
$\psi_{ji}[t], \kappa_{ji}[t], \epsilon_{ji}[t], \beta_{ji}[t], \delta_{(ji),(ni)}[t], r_l(f), r(f)$;

Constants: $M, N, T, B_i, p_j, L_{ji}^2, \beta, N_0, G, G', C_{ji}, D_{jki}$.

R-TMP is a mixed-integer linear problem (MILP). Therefore, we can apply a solver such as CPLEX [98] to obtain a solution efficiently.

2.5.2 A 25-node Example

The goal of this effort is twofold. First, we want to show how a solution to the TMP formulation looks like for an example network. By examine the details of our solution for an example network, one could gain some quantitative understanding of the interaction between DoF IC and SIC. Second, we want to perform a comparison study between our joint DoF IC and SIC framework and that without SIC.

We consider a randomly generated multi-hop wireless network with 25 nodes that are distributed in a 100×100 area. For generality, we normalize all units for distance, data rate, bandwidth, and power with appropriate dimensions. At the network layer, minimum-hop routing is employed. There are 4 active sessions in the network with each session's source node and destination node given in Table 2.2. Each node is equipped with $M = 4$ antennas. The transmit power for

Table 2.2: Source node and destination node in the 25-node network.

Session f	Source Node $s(f)$	Dest. Node $d(f)$
1	0	20
2	9	17
3	21	2
4	15	14

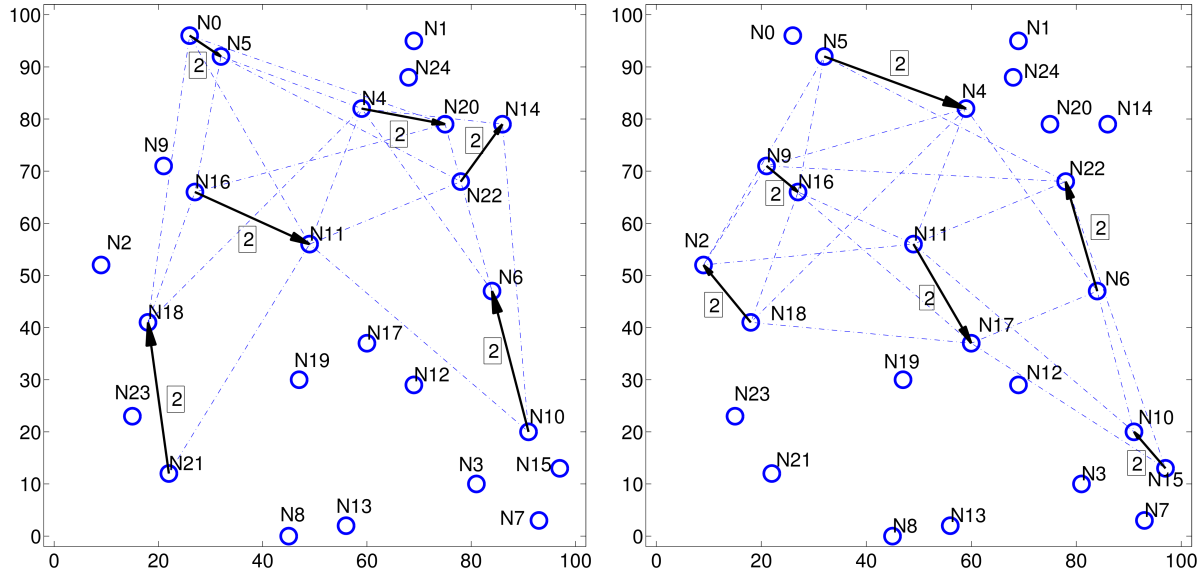
each data stream at a node is set to 1. The path-loss factor L_{ji}^2 between nodes i and j is $L_{ji}^2 = d_{ji}^{-\alpha}$, where d_{ji} is the distance between the two nodes and $\alpha = 3$ is the path-loss index. The power of ambient noise is $N_0 = 10^{-6}$. The average value of C_{ji} is 0.3460. The worst case upper bound value for D_{jki} is 7.3753. For the 25-node network, we apply CPLEX solver for the TMP formulation. In [83], it was recommended that SIC be used with direct sequence spread spectrum (DSSS). We follow this approach and assume DSSS's spreading gain is 3. We assume $\beta = 1.2$ and $T = 2$ time-slots.

Figure 2.6 shows the set of active links and the number of data streams per link in each time slot in the solution. Table 2.3 shows the details of SIC and DoF IC in each time slot. The first column identifies the time slot (1 or 2). The second column shows the active receivers in each time slot. The third column shows the transmitters (both desired and interfering) with respect to the receiver. The fourth column shows the number of DoFs for SM on the intended link (with 0 indicating interference). The fifth column shows whether the interference from the transmitter is handled by SIC (through SINR calculation). The last column shows whether the interference is canceled by DoF IC on the transmitter side or receiver side.

To show the benefits of the joint DoF IC and SIC scheme, we compare our solution to that without SIC. The achievable objective value is 1 under the proposed joint DoF IC and SIC scheme and it is 0.5 when SIC is not used. The increase in throughput is therefore 100%.

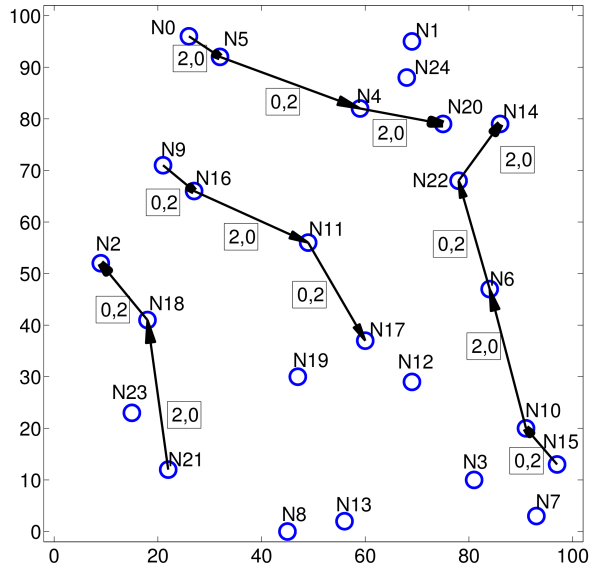
Table 2.3: Details of SIC and DoF IC on each link in each time slot.

Time Slot 1					Time Slot 2				
Rx	Tx	SM	SIC	IC	Rx	Tx	SM	SIC	IC
N_5	N_0	2		0	N_2	N_5	0	✓	0
	N_4	0	✓	0		N_9	0	✓	0
	N_{16}	0	✓	0		N_{11}	0	✓	0
	N_{22}	0	✓	0		N_{18}	2		0
N_6	N_4	0	✓	0	N_4	N_5	2		0
	N_{10}	2		0		N_6	0		2 at Rx
	N_{22}	0		2 at Rx		N_9	0		2 at Tx
N_{11}	N_0	0	✓	0		N_{11}	0		2 at Tx
	N_4	0		2 at Rx		N_{18}	0	✓	0
	N_{10}	0		2 at Tx	N_{10}	N_6	0	✓	0
	N_{16}	2		0		N_{11}	0	✓	0
	N_{21}	0		2 at Tx		N_{15}	2		0
	N_{22}	0		2 at Tx	N_{16}	N_5	0	✓	0
N_{14}	N_4	0	✓	0		N_9	2		0
	N_{10}	0	✓	1		N_{11}	0		2 at Tx
	N_{22}	2		0		N_{18}	0	✓	0
N_{18}	N_0	0	✓	0	N_{17}	N_6	0		2 at Tx
	N_4	0		2 at Tx		N_9	0	✓	0
	N_{16}	0		2 at Rx		N_{11}	2		0
	N_{21}	2		0		N_{15}	0	✓	0
N_{20}	N_0	0	✓	0		N_{18}	0		2 at Rx
	N_4	2		0	N_{22}	N_5	0		2 at Tx
	N_{16}	0	✓	0		N_6	2		0
	N_{22}	0		2 at Tx		N_9	0		2 at Rx
						N_{11}	0	✓	0



(a) Time slot 1

(b) Time slot 2



(c) Combined results

Figure 2.6: a) and (b) show scheduled links, DoFs allocation on each link, and interference pattern in time slots 1 and 2, respectively. A solid arrow line represents a directed transmission link (with the number of data streams on the link shown in a box) and a dashed arrow line represents an interference. (c) shows the combined results for both time slots (with the number of data streams for each time slot on the link shown in a box).

2.5.3 Complete Results

The previous section gives results for a 25-node example network. In this section, we provide three additional sets of results. First, we perform the same study for 25-node network, but for 50 randomly generated network instances, with each session's source and destination nodes being randomly selected among the nodes. Table 2.4 shows the results for both the proposed joint scheme and DoF IC only scheme. We find that the average percentage increase over the 50 instances is 87.75%.

Next, we perform the study on 50-node network randomly deployed in 150×150 area. There are 4 sessions in each network instance, with each session's source and destination nodes being randomly selected among the nodes. Table 2.5 and 2.6 lists the objective values under the two schemes when the number of time slots is 2 and 4, respectively. In the case when there are only 2 time slots, the average percentage increase in objective value under the joint scheme is 86.66%, while the average percentage increase is 65% when there are 4 time slots. This decrease is intuitive as more time slots will offer more room for scheduling, thus alleviating the DoF resource shortage issue in DoF IC only scheme.

2.6 Related Work

Related work on single-antenna SIC and MIMO DoF IC has been described in Section 2.1. In this section, we focus our review on related work that employs SIC in MIMO.

For point-to-point MIMO communication, there has been extensive research on SIC based receivers to decode received data streams. The first MIMO SIC for point-to-point MIMO communication was D-BLAST by Foschini [24]. The use of SIC helps boost the performance of a MIMO receiver by decoding and subtracting each data stream successively. The boost in power gain comes from increased SINR at each stage. It was shown in [24] that a receiver based on minimum mean-square-error (MMSE) outperforms zero forcing in terms of mitigating both in-

Table 2.4: Objective values under joint scheme and DoF IC only scheme for 25-node network over 50 instances.

IDX	DoF IC and SIC	DoF IC only	Incr %	IDX	DoF IC and SIC	DoF IC only	Incr %
1	1	0.5	100	26	1	0.5	100
2	1	0.5	100	27	1	0.5	100
3	1	0.5	100	28	2	1	100
4	1	0.5	100	29	1	0.5	100
5	1	0.5	100	30	1	0.5	100
6	1	0.5	100	31	1	0.5	100
7	1	0.5	100	32	1	0.5	100
8	1	0.5	100	33	1	0.5	100
9	1	0.5	100	34	1	0.5	100
10	1	0.5	100	35	1	0.5	100
11	0.5	0.5	0	36	1	0.5	100
12	0.5	0.5	0	37	1	0.5	100
13	0.5	0.5	0	38	1	0.5	100
14	1	0.5	100	39	1	0.5	100
15	2	1	100	40	1	0.5	100
16	1	0.5	100	41	2	1	100
17	0.5	0.5	0	42	1	0.5	100
18	0.5	0.5	0	43	1	0.5	100
19	2	0.5	300	44	1	0.5	100
20	1	0.5	100	45	1	0.5	100
21	1	0.5	100	46	1	0.5	100
22	1	0.5	100	47	0.5	0.5	0
23	1	0.5	100	48	0.5	0.5	0
24	1	0.5	100	49	0.5	0.5	0
25	1	0.5	100	50	1	0.5	100

Table 2.5: Objective values under joint scheme and DoF IC only scheme for 50-node network over 50 instances. Number of time slot is 2.

IDX	DoF IC and SIC	DoF IC only	Incr %	IDX	DoF IC and SIC	DoF IC only	Incr %
1	1	0.5	100	26	1	0.5	100
2	1	0.5	100	27	2	1.5	33.00
3	1	0.5	100	28	1	0.5	100
4	1	0.5	100	29	2	1	100
5	1	0.5	100	30	4	2	100
6	0.5	0.5	0	31	2	1	100
7	1	0.5	100	32	1	0.5	100
8	1	0.5	100	33	1	0.5	100
9	1	0.5	100	34	1	0.5	100
10	1	0.5	100	35	1	1	0
11	1	0.5	100	36	4	2	100
12	1	0.5	100	37	4	2	100
13	1	0.5	100	38	1	0.5	100
14	1	0.5	100	39	1	0.5	100
15	0.5	0.5	0	40	0.5	0.5	0
16	0.5	0.5	0	41	0.5	0.5	0
17	1	0.5	100	42	1	0.5	100
18	1	0.5	100	43	1	0.5	100
19	1	0.5	100	44	1	0.5	100
20	1	0.5	100	45	1	0.5	100
21	2	1	100	46	1	0.5	100
22	1	0.5	100	47	2	1	100
23	1	0.5	100	48	1	0.5	100
24	2	1	100	49	1	0.5	100
25	1	0.5	100	50	1	0.5	100

Table 2.6: Objective values under joint scheme and DoF IC only scheme for 50-node network over 50 instances. Number of time slot is 4.

IDX	DoF IC and SIC	DoF IC only	Incr %	IDX	DoF IC and SIC	DoF IC only	Incr %
1	0.75	0.5	50	26	0.75	0.5	50
2	0.5	0.25	100	27	0.75	0.5	50
3	1	0.75	33.33	28	0.5	0.25	100
4	0.75	0.5	50	29	0.75	0.5	50
5	1	0.5	100	30	0.75	0.5	50
6	0.5	0.25	100	31	0.5	0.25	100
7	0.5	0.25	100	32	1	0.5	100
8	0.5	0.25	100	33	0.5	0.25	100
9	0.5	0.25	100	34	1	0.5	100
10	0.5	0.25	100	35	1	1	0
11	0.75	0.5	50	36	0.75	0.5	50
12	1	0.5	100	37	0.5	0.5	0
13	1	0.75	33.33	38	0.75	0.5	50
14	0.75	0.5	50	39	1	0.75	33.33
15	0.5	0.25	100	40	1	0.5	100
16	0.25	0.25	0	41	1	0.5	100
17	1	0.5	100	42	0.75	0.5	50
18	0.75	0.5	50	43	0.75	0.5	50
19	0.5	0.25	100	44	1	0.75	33.33
20	1	0.5	100	45	0.75	0.75	0
21	1	0.75	33.33	46	1	0.5	100
22	0.75	0.5	50	47	0.75	0.5	50
23	1	0.5	100	48	0.75	0.5	50
24	0.75	0.5	50	49	1	0.75	33.33
25	0.5	0.5	0	50	0.5	0.25	100

terference and noise. The optimality of MMSE in conjunction with SIC was shown in [79] by Varanasi *et al.* A simplified version of D-BLAST, called V-BLAST, was proposed by Wolniansky *et al.* in [85]. A number of performance studies of V-BLAST and MIMO-MMSE SIC in the context of point-to-point MIMO communication can be found in [16, 50, 60, 92, 97].

In multi-user MIMO with SIC, although the incoming data streams may come from different MIMO transmitters, the receiver design is still similar to that for the point-to-point MIMO SIC receiver. That is, V-BLAST architecture remains prevalent in the design of multi-user MIMO-SIC receiver (e.g., [69]). Although the MMSE based receiver is prevalent in single-user and multi-user MIMO SIC, it is not tractable when we study MIMO SIC in a multi-hop network environment. This is because when scheduling algorithm is unknown (part of the optimization problem), the number of variables and constraints become prohibitively large when MMSE is employed in a multi-hop network with MIMO SIC. Due to this reason, we do not employ MMSE in our MIMO SIC receiver and instead employ the simple DoF model, which is based on zeroforcing. In this sense, DoF based MIMO SIC design may only achieve sub-optimal. But it offers an excellent starting point to understand the potential of SIC in a multi-hop MIMO network.

In [29], Gelal *et al.* studied how to maximally exploit SIC in multi-user MIMO networks through selection of a subset of links that can be concurrently active in each receiver's neighborhood. The proposed algorithms attempt to divide the network into a minimum number of sub-topologies where the set of links in each sub-topology can be active simultaneously. In [29], MIMO's capability is limited to selection diversity on the receiver side, while SM and IC capabilities are not considered. Further, it is not clear how the proposed the algorithm can be extended to address end-to-end data flow routing via multiple hops in the network.

2.7 Conclusions

DoF is an important concept to characterize a node's resource for SM and IC in a MIMO network. SIC is a powerful technique for IC and has the potential to conserve DoF resources once employed

in a MIMO network. However, SIC's effectiveness is limited by its stringent SINR threshold criteria. This paper investigated how to conserve DoF resources and meeting SIC's SINR threshold criteria by jointly exploiting the strengths of each technique. We developed the necessary mathematical models to characterize (i) how precious DoF resources can be conserved through the use of SIC and (ii) how the stringent SINR threshold criteria can be met through the use of DoF-based IC. Our modeling work was done in a general multi-hop MIMO network, which by default included scheduling and routing for user traffic sessions. Based on our cross-layer mathematical models, we studied a throughput maximization problem and confirmed that SIC and DoF IC can indeed achieved the two benefits that proposed in this paper.

Chapter 3

Throughput of Full-Duplex MIMO-Empowered Multi-hop Wireless Networks

3.1 Introduction

Half duplex has been the fundamental limitation in a transceiver design since the beginning of wireless communications. Due to half duplex, a transceiver can only perform one task at a time: transmit or receive. Conceptually, half duplex cuts down the potential throughput by half at a node. Recent breakthrough in full duplex transceiver design opens up new possibility in wireless communications. In [15], Choi et al. proposed the first practical design of SISO full-duplex system by using a combination of antenna cancellation, RF interference cancellation and digital interference cancellation with two transmit antennas and one receive antenna. This design can achieve 60dB self interference cancellation and is suitable for low-power narrow-band 802.15.4 systems. In [39], Jain et al. proposed a new design of full duplex by employing the Balun transformer with two antennas. This design can achieve 85dB self interference cancellation and can be used for low power,

narrow-band protocols such as Zigbee. In [3], Aryafar et al. proposed an approach to achieve full duplex by using two transmit antennas and two receive antennas to cancel interference in the spatial domain. This design can achieve 75dB self-interference cancellation. The current state-of-the art SISO full duplex design is the one in [7] by Bharadia *et al.* This design only requires a single antenna and is closest to what one would wish from a full duplex node. It can achieve 110dB of self-interference cancellation and thus meets the requirements of 802.11 standards, which are most widely deployed in commercial wireless devices. The state of the art MIMO full duplex design is recently proposed by Bhardia *et. al.* in [8]. In this fully working MIMO full-duplex, all antennas within a node can transmit and receive simultaneously. The self interference and the cross talk between antennas are properly canceled to the noise floor in [8].

On the other hand, MIMO has been widely adopted by the communications industry and the research community due to its capabilities of spatial multiplexing (SM) gain, interference cancellation (IC), and diversity gain. Until recently, research on MIMO has been limited at the physical (PHY) layer or for single-hop communications due to the lack of tractable MIMO models. Recent advances in MIMO degree-of-freedom (DoF) models removed this stagnation and allowed MIMO research to penetrate the networking community [9, 12, 36, 62, 49]. The concept of DoF was originally defined to represent the multiplexing gain of a MIMO channel in the information theory (IT) community. This DoF concept was then extended by the networking community to characterize a node's spatial freedom provided by its multiple antennas. Under a DoF model, only simple numerical computation is needed to account for a node's resource allocation for SM and IC. The basic idea of DoF-based MIMO models is as follows [49]: (i) The number of available DoFs at a node is equal to the number of its antennas. A node may use its DoFs for either SM or IC. (ii) For SM, both transmit and receive nodes need to consume DoFs. For each data stream, both the transmit and receive nodes need to consume one DoF. (iii) For IC, either the transmit node or the receive node may consume DoFs. Clearly, under a DoF model, the number of available DoFs at a node is considered a precious recourse and must be utilized efficiently. In particular, when a node uses its DoFs for IC, its remaining DoFs for SM will be reduced. Therefore, there is a critical need to conserve DoFs for IC if one wishes to maximize the number of DoFs for SM. It should be noted

the granularity of the data rate in MIMO DoF model is DoF per time slot for simplicity. However, it is not clear how to compute actual links data rates based on received SINR of each data stream.

In [59], Nguyen *et al.* shows that the network throughput gain of FD radios over HD onces in unexpectedly marginal. This is mainly due to additional network interference introduced in the network when both ends of the FD links are transmitting. In this paper, we claim that the true potential gain of MIMO FD in multi-hop network can only be unleashed when it is jointly integrated with IC to combat additional interference introduced by FD links. By understanding the state-of-the-art MIMO full duplex transceiver architecture, we proposed a model for MIMO full duplex in multi-hop network, where nodes can operate in FD mode and its spatial DoF resources can be used for SM and IC. We develop the necessary mathematical models to realize such a MIMO full duplex model. The main contributions of this paper are summarized as follows.

- We will show MIMO FD in wireless multi-hop networks when integrated with IC can achieve a significant gain over MIMO HD.
- We develop a simple receiver side IC model for a multi-hop full duplex MIMO network to ensure that all the strong interference in the network can be canceled using MIMOs IC capabilities.
- Based on the proposed model, we develop a cross-layer optimization problem with the objective to maximize the minimum session throughput. Due to its inherent property, the formulated optimization problem inevitably involves nonlinear constraints, making it notoriously difficult to solve. To make the problem more tractable, we linearize the nonlinear constraints by two steps: formulation and near optimal approximation. We show that our reformulation does not loss the optimality of the original problem and our approximation can achieve near optimality to the original problem. In particular, our approximation approach requires a minimum number of new linear constraint and, more notably, does not introduce any new variables.

The remainder of this chapter is organized as follows. In Section 3.2, we will develop our

Table 3.1: Notation

Symbol	Definition
N	Number of nodes in the network
$\mathcal{N}$	Set of nodes in the network (one TX and one RX chain per each node in the network)
$\mathcal{I}_i$	Set of nodes within node i's strong interference range
M	Number of antennas at node each node
G_{ij}	path loss from the transmit node i to the receive node j
$P_{\max}$	Total transmit power of an active transmitter
Q	Total number of transmit power levels
$p_l^q[t]$	Transmit power for the kth data stream at the transmit node of link l in time slot t
$c_l^q[t]$	rate of qth data stream of link l in time slot t
$c_l[t]$	aggregate data rate at link l over all data streams in time slot t
W	bandwidth of the channel
P_n	Noise power
L	Total number of links in the network
$\mathcal{L}$	Set of potential link in the network
$\mathcal{L}_i^{\text{in}}$	Set of incoming links at node i
$\mathcal{L}_i^{\text{out}}$	Set of outgoing links at node i
$A_l^q[t]$	power scaling coefficient of the interference and the noise for the kth data stream of link l
α	self-interference cancellation coefficient
$\text{Rx}(l)$	Receiver of link l
$\text{Tx}(l)$	Transmitter of link l
$r(f)$	Rate of session f
$r_l(f)$	Rate for session f on link l
$x_i[t]$	Indicator variable to show if node i is a transmitter in time slot t
$y_i[t]$	Indicator variable to show if node i is a receiver in time slot t
$z_l[t]$	Number of data streams on link l in time slot t
$m_i[t]$	Indicator variable to show whether node i is operating in FD mode in time slot t

cross-layer MIMO full duplex model. In Section 3.3, we first present a problem formulation for a throughput maximization problem based on the mathematical models developed in the previous sections. Next, we present the reformulation and a near optimal approximation necessary for solving the original problem. The near-optimality proof for the proposed near optimal approximation is also presented in Section 3.3. In Section 3.4, we explore the throughput performance of full duplex in a multi-hop wireless network through numerical results. Related work in full duplex communication is reviewed in Section 3.5. Section 3.6 concludes this chapter.

3.2 MIMO Full Duplex Cross-Layer Model

In this section, we will extend the DoF IC model in [49] into MIMO full duplex. Table 3.1 shows the notation used in this paper.

Consider a multi-hop MIMO network consisting of a set of nodes $\mathcal{N}$ which has N elements. Each node is assumed to have M antennas. Suppose that there are L possible links in the network. Denote $\mathrm{Tx}(l)$ and $\mathrm{Rx}(l)$ as the transmit and receive nodes of link l, $1 \leq l \leq L$. We consider a time-slotted scheduling, where a time frame consists of T time slots. Depending on link scheduling, a subset of links will be active in time slot t, $1 \leq t \leq T$.

SM Constraints. Denote $x_i[t]$ as a binary variable to indicate whether node $i \in \mathcal{N}$ is transmitting in time slot t, i.e., $x_i[t] = 1$ if node i is a transmitter in time slot t and 0 otherwise. Similarly, denote $y_i[t]$ as a binary variable to indicate whether node $i \in \mathcal{N}$ is receiving in time slot t, i.e., $y_i[t] = 1$ if node i is a receiver in time slot t and 0 otherwise. Denote $z_l[t]$ as the number of data streams over link l. If node i is not a transmitter, then we have $\sum_{l \in \mathcal{L}_i^{\mathrm{out}}} z_l[t] = 0$. Otherwise, the total number of outgoing streams should be positive and lesser than the number of antennas, i.e., $1 \leq \sum_{l \in \mathcal{L}_i^{\mathrm{out}}} z_l[t] \leq M$. These two cases can be expressed in a compact form as follows:

$$x_i[t] \leq \sum_{l \in \mathcal{L}_i^{\mathrm{out}}} z_l[t] \leq M x_i[t], \quad (1 \leq i \leq N, 1 \leq t \leq T) \,. \tag{3.2.1}$$

Similarly, depending on whether node i is an active receiver, we have the following constraint:

$$y_i[t] \leq \sum_{l \in \mathcal{L}_i^{\mathrm{in}}} z_l[t] \leq M y_i[t], \quad (1 \leq i \leq N, 1 \leq t \leq T) \,. \tag{3.2.2}$$

IC Constraints. We first introduce a concept of "strong" interference range D. For a receive node j, it may be interfered by all the unintended transmit nodes in the network. We distinguish an interference as either a "strong" interference or a "weak" interference through strong interference range D. Specifically, if the distance from a transmit node to its unintended receive node is less than or equal to D, we consider this interference as strong interference; otherwise, we consider it as week interference. In our model, only strong interference will be considered for IC, while weak

interference will not be considered for IC. Instead, weak interference will be treated as noise at the receive node when calculating its achievable data rate. We assume $\mathcal{I}_i$ is the set of node strongly interfered by node i. Denote $\mathcal{I}_i$ as the set of nodes that are located within the strong interference range of transmitter i.

The current state-of-the art MIMO full duplex radio design is the one in [8] by Bharadia *et al.* This design achieves MIMO full duplex with a cancellation design solely based on digital estimation and cancellation algorithms that eliminate almost all interference. The proposed design in [8] does not consume DoF for self-interference cancellation. Both self talk and cross talks between antennas of the MIMO node are canceled merely by underlying digital and analogue circuits and all DoF resources are available for exploitation in MIMO techniques. The proposed MIMO full duplex system proposed in [8] employs two separate RF chains for RX/TX. Therefore, it is possible to design independent precoding/receive matrices operating simultaneously to support SM/IC in each chain independently.

In this work, we will adapt a very simple IC strategy. Strong intereference is canceled by zero-forcing and consuming DoF at the receiver side. Denote $\mathcal{L}_i^{\text{in}}$ and $\mathcal{L}_i^{\text{out}}$ as the set of potential incoming and outgoing links at node i, respectively. Denote $z_l[t]$ as the number of data streams over link l. The above IC constraints can be written in mathematical form as following:

$$\sum_{k\in\mathcal{L}_i^{\text{in}}} z_k[t] + \sum_{j\in\mathcal{I}_i} \sum_{l\in\mathcal{L}_j^{\text{out}}}^{\text{Rx}(l)\neq i} z_l[t] \leq M\cdot y_i[t] + (1-y_i[t])B_i, (1\leq i\leq N, j\in\mathcal{I}_i, 1\leq t\leq T)\ , \quad (3.2.3)$$

where B_i is a large constant and is no smaller than $\sum\limits_{j\in\mathcal{I}_i}\sum_{l\in\mathcal{L}_j^{\text{out}}}^{\text{Rx}(l)\neq i} z_l[t]$. For example, we can set $B_i = M\cdot|\mathcal{I}_i|$.

Operation Mode Constraints Denote $m_i[t]$ as a binary variable to indicate whether node $i\in\mathcal{N}$ is operating in FD mode in time slot t, i.e., $m_i[t]=1$ if node i is simultaneously transmitting and receiving in time slot t and 0 otherwise. The logical relationship between $x_i[t]$, $y_i[t]$, and $m_i[t]$ can

be expressed by the following constraints:

$$m_i[t] \geq x_i[t] + y_i[t] - 1, \qquad (1 \leq i \leq N, 1 \leq t \leq T)\ , \tag{3.2.4}$$

$$m_i[t] \leq x_i[t], \qquad (1 \leq i \leq N, 1 \leq t \leq T)\ , \tag{3.2.5}$$

$$m_i[t] \leq y_i[t] \qquad (1 \leq i \leq N, 1 \leq t \leq T)\ . \tag{3.2.6}$$

Data Stream Based Power Allocation Constraints. For power control, we assume that the transmission power allocated to qth outgoing data stream at transmit node of link l can be tuned to a finite number of levels between 0 and $P_{\max}$. We further assume that the total transmit power of a node is P_s. To model this discrete data stream power control, we introduce an integer parameter Q that represents the total number of power levels to which a transmitter can be adjusted, i.e., transmission power can be $0, \frac{1}{Q}P_{\max}, \frac{2}{Q}P_{\max}, \ldots, P_{\max}$. Denote $p_l^q \in \{0, 1, \ldots, Q\}$ the integer levels for transmission power allocated to qth outgoing data stream at transmit node of link l. If node i is an active transmitter (i.e., $x_i(t) = 1$), then its total transmit power over all data streams is $P_{\max}$. Otherwise (i.e., $x_i(t) = 0$), it transmit power for each data stream is zero. Then we have

$$\sum_{l \in \mathcal{L}_i^{\text{out}}} \sum_{q=1}^{z_l[t]} p_l^q[t] = Q \cdot x_i[t], \qquad (1 \leq i \leq N, 1 \leq t \leq T)\ . \tag{3.2.7}$$

Data Stream Capacity Constraints. Denote $c_l^q[t]$ as the achievable rate of qth data stream of link l in time slot t. Denote $\gamma_l^q[t]$ as the effective SINR at the receive node of link l for receiving the kth data streams of link l.

$$c_l^q[t] = W \cdot \log_2(1 + \gamma_l^q[t]), (1 \leq l \leq L, 1 \leq q \leq z_l[t], 1 \leq t \leq T)\ . \tag{3.2.8}$$

where W is the system bandwidth.

We know calculate the effective SINR $\gamma_l^q[t]$. The DoF IC cancels the strong interfering signals and weak interference signals is treated as noise. Therefore, the effective SINR at the receive node of link l for receiving the qth data streams of link l for $(1 \leq l \leq L, 1 \leq q \leq z_l[t], 1 \leq t \leq T)$ can be written as

$$\gamma_l^q[t] = \frac{G_{\mathrm{Tx}(l)\mathrm{Rx}(l)} \frac{p_l^q[t]}{Q} P_{\max}}{A_l^q[t]\big[\sum\limits_{i \in \mathcal{N} \setminus \mathcal{I}_{\mathrm{Rx}(l)}}^{i \neq \mathrm{Tx}(l)} G_{i\mathrm{Rx}(l)} P_{\max} x_i[t] + \alpha P_{\max} m_{\mathrm{Rx}(l)}[t] + P_n\big]} \ , \tag{3.2.9}$$

where G_{ij} is the path loss from the transmit node i to the receive node j; P_n is the noise power at the receiver; $A_l^q[t]$ is a power scaling coefficient of the interference and the noise for the qth data stream of link l, which is determined by the channel matrices. The value of α characterizes the performance of self-interference cancellation, which is consist of the residue of self-talk and cross-talk – the smaller the value of α, the cleaner the cancellation of self interference.

Denote $c_l[t]$ as the aggregate data rate at link l over its $z_l[t]$ data streams in time slot t. Then we have

$$c_l[t] = \sum_{q=1}^{z_l[t]} c_l^q[t], \qquad (1 \leq l \leq L, 1 \leq t \leq T) \ . \tag{3.2.10}$$

Link Capacity Constraints. Denote $r_l(f)$ as the amount of data rate on link l that is attributed to session $f \in \mathcal{F}$. Since the aggregate data rate on link l cannot exceed the link's average rate, we have

$$\sum_{f \in \mathcal{F}} r_l(f) \leq \frac{1}{T} \sum_{t=1}^{T} c_l[t], \quad (1 \leq l \leq L) \ , \tag{3.2.11}$$

where the right-hand-side represents the average throughput on link l over a frame (T time slots).

Flow Routing Constraints. Suppose there is a set of active sessions $\mathcal{F}$. Denote $r(f)$ as the rate of session $f \in \mathcal{F}$ and $r_{\min}$ as the minimum session rate, i.e., $r_{\min} = \min_{f \in \mathcal{F}} r(f)$. Denote $s(f)$ and $d(f)$ as the source and destination nodes of session $f \in \mathcal{F}$, respectively. Then at source node, $s(f)$, $f \in \mathcal{F}$, we have the following flow balance:

$$\sum_{l \in \mathcal{L}_i^{\mathrm{out}}} r_l(f) = r(f), \quad (i = s(f), f \in \mathcal{F}) \ . \tag{3.2.12}$$

At any intermediate relay node, we have

$$\sum_{l\in\mathcal{L}_i^{\text{in}}} r_l(f) = \sum_{l\in\mathcal{L}_i^{\text{out}}} r_l(f), (1 \le i \le N, i \ne s(f), i \ne d(f), f \in \mathcal{F}) \ . \tag{3.2.13}$$

At a destination node, we have

$$\sum_{l\in\mathcal{L}_i^{\text{in}}} r_l(f) = r(f), \quad (i = d(f), f \in \mathcal{F}) \ . \tag{3.2.14}$$

It can be easily verified that if (3.2.12) and (3.2.13) are satisfied, then (3.2.14) is also satisfied. Therefore, it is sufficient to have (3.2.12) and (3.2.13).

3.3 Problem Formulation and Solution

3.3.1 Problem Formulation

In Sections 3.2, we established key models to extend DoF IC to MIMO full duplex and convert DoF to bit rate. In this section, we show how these models can be used to study networking problems. Let's consider a typical throughput maximization problem in a multi-hop MIMO network.

Putting all the constraints together, we have the following formulation for the throughput maximization problem:

$$\begin{aligned}
\mathbf{BR-TMP}_1 \quad \max \quad & r_{\min} \\
\text{s.t.} \quad & r_{\min} \le r(f) \qquad (f \in \mathcal{F}); \\
& \text{SM Constraints: (3.2.1),(3.2.2);} \\
& \text{IC Constraints: (3.2.3);} \\
& \text{Operation mode constraints: (3.2.4)–(3.2.6);} \\
& \text{Data Stream Power Allocation Constraints: (3.2.7);} \\
& \text{Data Stream Capacity Constraints: (3.2.8)–(3.2.10);} \\
& \text{Link Capacity Constraints: (3.2.11);} \\
& \text{Flow balance constraints: (3.2.12), (3.2.13).}
\end{aligned}$$

In $\mathbf{BR-TMP}_1$, constraints (3.2.7)–(3.2.10) are non-linear. Therefore, $\mathbf{BR-TMP}_1$ is a mixed-integer nonlinear program (MINLP), which in general is NP-hard [28]. MINLP problems are known to be difficult due to the combinatorial nature of mixed-integer programs and the difficulty in solving nonlinear programs. Note that there exist some techniques to address *general* MINLP problems (e.g., outer-approximation methods [23], branch-and-bound [34], extended cutting plane methods [84], and generalized Benders' decomposition [30]). However, these techniques do not exploit our problem-specific structures and properties, and hence can only handle small-sized problems.

In the remaining of this section, we will show through reformulation on (3.2.7)–(3.2.10) as well as linearization of the logarithmic function in (3.2.8), $\mathbf{TMP}_1$ can be reformulated into a mixed integer linear program (MILP). Although the theoretical worst-case complexity to a general MILP problem is exponential [28, 68], there exist highly efficient heuristics (e.g., sequential fixing algorithm [38, Chapter 10]) to solve it. Another approach is to apply an off-the-shelf solver (CPLEX [98]), which we find can handle up to a moderate-sized network successfully. Since the main goal of this paper is to explore DoF IC and FD jointly, it is sufficient to demonstrate our results with moderate-sized networks. Therefore, we will use CPLEX to solve MILP. The solution to this optimization problem gives us the optimal decision on when and how to perform scheduling, SM, IC, and/or operate in HD/FD in different time-slots.

3.3.2 Problem Reformulation

In $\mathbf{BR-TMP}_1$, the nonlinear constraints are (3.2.7)–(3.2.10). We first reformulate (3.2.7)–(3.2.10) using Reformulation Linearization Technique (RLT) [70] and then linearize the logarithmic function in (3.2.8) via near optimal approximation. We will show that the resulting optimization problem does not involve any nonlinear constraints and provides a near optimal solution to the original problem (i.e., $\mathbf{BR-TMP}_1$).

Reformulation of (3.2.7) and (3.2.10). (3.2.7) and (3.2.10) are non-linear constraints since

$z_l[t]$ is not a constant rather an optimization variable. In order to linearize these two constraints, we introduce $M - z_l[t]$ dummy outgoing data streams for node $\mathrm{Tx}(l)$ (transmitter of link l) and add constraints to force the data rate of the dummy data streams to zero. Denote $\lambda_l^q[t]$ as a binary variable to indicate whether or not outgoing data stream q from the transmitter of link l (node $\mathrm{Tx}(l)$) is dummy at time slot t. Specifically, $\lambda_l^q[t] = 0$ if data stream k is dummy and 1 if not. Therefore, we have

$$\sum_{q=1}^{M} \lambda_l^q[t] = z_l[t], \quad (1 \le l \le L, 1 \le t \le T) \; . \tag{3.3.1}$$

Consider the outgoing data streams at $\mathrm{Tx}(l)$ (transmitter of link l). For data stream q, If it is dummy (i.e., $\lambda_l^q[t] = 0$), then the transmit power allocated for this data stream should be 0. Otherwise (i.e., $\lambda_l^q[t] = 1$), the transmit power allocated for this data stream is limited by the total transmit power $P_{\max}$. Therefore, we have

$$0 \le p_l^q[t] \le Q \cdot \lambda_l^q[t], \quad (1 \le l \le L, 1 \le q \le M, 1 \le t \le T) \; . \tag{3.3.2}$$

Similarly, we use the following constraints to force the achievable data rate of a dummy data stream to zero.

$$0 \le c_l^q[t] \le A \cdot \lambda_l^q[t], \quad (1 \le l \le L, 1 \le q \le M, 1 \le t \le T) \; . \tag{3.3.3}$$

where A is a large enough constant.

The new constraints (3.3.1)–(3.3.3) ensure that a dummy data stream has zero transmit power and zero data rate. Therefore, it is equivalent to reformulate (3.2.7) to

$$\sum_{l \in \mathcal{L}_i^{\mathrm{out}}} \sum_{q=1}^{M} p_l^q[t] = Q \cdot x_i[t], \qquad (1 \le i \le N, 1 \le t \le T) \; . \tag{3.3.4}$$

and reformulate (3.2.10) to

$$c_l[t] = \sum_{q=1}^{M} c_l^q[t], \qquad (1 \le l \le L, 1 \le t \le T) \; . \tag{3.3.5}$$

Reformulation of (3.2.9). Based on the new constraints (3.3.1)–(3.3.3), we can rewrite constraint (3.2.9) for $(1 \leq l \leq L, 1 \leq q \leq M, 1 \leq t \leq T)$ as

$$\sum_{i \in \mathcal{N} \setminus \mathcal{I}_{\mathrm{Rx}(l)}}^{i \neq \mathrm{Tx}(l)} G_{i\mathrm{Rx}(l)} P_{\max} x_i[t] \gamma_l^q[t] + \alpha P_{\max} m_{\mathrm{Rx}(l)}[t] \gamma_l^q[t] + P_n \gamma_l^q[t] = \frac{G_{\mathrm{Tx}(l)\mathrm{Rx}(l)}}{A_l^q[t]} \frac{p_l^q[t]}{Q} P_{\max} \ .$$

This constraint is nonlinear due to nonlinear terms $x_i[t]\gamma_l^q[t]$ and $m_{\mathrm{Rx}(l)}[t]\gamma_l^q[t]$. We can use reformulation linearizion technique (RLT [70]) to linearize $x_i[t]\gamma_l^q[t]$ and $m_{\mathrm{Rx}(l)}[t]\gamma_l^q[t]$. Note since these terms are the product of a integer and continuous variable, the linearizion guarantees optimality. We define new variables $\omega_{il}^q[t] = x_i[t]\gamma_l^q[t]$, and $\nu_l^q[t] = m_{\mathrm{Rx}(l)}[t]\gamma_l^q[t]$. Then this constraint can be written for $(1 \leq l \leq L, 1 \leq q \leq M, 1 \leq t \leq T)$ as

$$\sum_{i \in \mathcal{N} \setminus \mathcal{I}_{\mathrm{Rx}(l)}}^{i \neq \mathrm{Tx}(l)} G_{i\mathrm{Rx}(l)} P_{\max} \omega_{il}^q[t] + \alpha P_{\max} \nu_l^q[t] + P_n \gamma_l^q[t] = \frac{G_{\mathrm{Tx}(l)\mathrm{Rx}(l)}}{A_l^q[t]} \frac{p_l^q[t]}{Q} P_{\max} \ . \tag{3.3.6}$$

To ensure that $\omega_{il}^q[t] = x_i[t]\gamma_l^q[t]$ always holds, we add the following two linear constraints in the formulation for $(1 \leq l \leq L, i \in \mathcal{N} \setminus \mathcal{I}_{\mathrm{Rx}(l)}, 1 \leq q \leq M, 1 \leq t \leq T)$

$$0 \leq \omega_{il}^q[t] \leq \frac{G_{\mathrm{Tx}(l)\mathrm{Rx}(l)} P_{\max}}{A_l^q[t] P_n} x_i[t] \ , \tag{3.3.7}$$

$$\frac{G_{\mathrm{Tx}(l)\mathrm{Rx}(l)} P_{\max}}{A_l^q[t] P_n} x_i[t] + \gamma_l^q[t] - \frac{G_{\mathrm{Tx}(l)\mathrm{Rx}(l)} P_{\max}}{A_l^q[t] P_n} \leq \omega_{il}^q[t] \leq \gamma_l^q[t] \ . \tag{3.3.8}$$

To ensure that $\nu_l^q[t] = m_{\mathrm{Rx}(l)}[t]\gamma_l^q[t]$ always holds, we add the following two linear constraints in the formulation for $(1 \leq l \leq L, 1 \leq q \leq M, 1 \leq t \leq T)$

$$0 \leq \nu_l^q[t] \leq \frac{G_{\mathrm{Tx}(l)\mathrm{Rx}(l)} P_{\max}}{A_l^q[t] P_n} m_{\mathrm{Rx}(l)}[t] \ , \tag{3.3.9}$$

$$\frac{G_{\mathrm{Tx}(l)\mathrm{Rx}(l)} P_{\max}}{A_l^q[t] P_n} m_{\mathrm{Rx}(l)}[t] + \gamma_l^q[t] - \frac{G_{\mathrm{Tx}(l)\mathrm{Rx}(l)} P_{\max}}{A_l^q[t] P_n} \leq \nu_l^q[t] \leq \gamma_l^q[t] \ . \tag{3.3.10}$$

Reformulation of (3.2.8). Based on (3.3.1)–(3.3.10), we know that the dummy data stream has zero transmit power, zero SINR, and zero data rate. Therefore, it is equivalent to translate (3.2.8) to the following constraint

$$c_l^q[t] = W \cdot \log_2(1 + \gamma_l^q[t]), (1 \leq l \leq L, 1 \leq q \leq M, 1 \leq t \leq T) \ . \tag{3.3.11}$$

In summary, we replace (3.2.7), and (3.2.8)–(3.2.10) with (3.3.1)–(3.3.11). The resulting optimization problem which we denote $\mathbf{BR-TMP}_2$, can be written as

$$
\begin{aligned}
\mathbf{BR-TMP}_2 \quad \max \quad & r_{\min} \\
\text{s.t.} \quad & r_{\min} \leq r(f) \qquad (f \in \mathcal{F}); \\
& \text{SM Constraints: (3.2.1),(3.2.2);} \\
& \text{IC Constraints: (3.2.3);} \\
& \text{Operation mode constraints: (3.2.4)–(3.2.6);} \\
& \text{Dummy Data Stream Constraints: (3.3.1)–(3.3.3);} \\
& \text{Data Stream Power Allocation Constraints: (3.3.4);} \\
& \text{Data Stream Capacity Constraints: (3.3.5)–(3.3.11);} \\
& \text{Link Capacity Constraints: (3.2.11);} \\
& \text{Flow balance constraints: (3.2.12), (3.2.13);} \\
& \text{Variables: } x_i[t], y_i[t], z_l[t], \pi_i[t], m_i[t], p_l^q[t], \gamma_l^q[t], c_l^q[t], \lambda_l^q[t], \\
& \omega_{il}^q[t], \nu_l^q[t], c_l[t], r_l(f), r(f).
\end{aligned}
$$

where $x_i[t]$, $y_i[t]$, $\pi_i[t]$, $m_i[t]$, and $\lambda_l^q[t]$ are binary variables; $z_l[t]$ and $p_l^q[t]$ are integer variables; $\gamma_l^q[t]$, $c_l^q[t]$, $\omega_{il}^q[t]$, $\nu_l^q[t]$, $c_l[t]$, $r_l(f)$, and $r(f)$ are continuous variables, and all the other parameters are constant. Denote $r_{\min}^*(\mathbf{BR-TMP}_1)$ as the optimal objective value of $\mathbf{BR-TMP}_1$ and $r_{\min}^*(\mathbf{BR-TMP}_2)$ as the optimal objective value of $\mathbf{BR-TMP}_2$. Then we have the following lemma:

Lemma 1. $\mathbf{BR-TMP}_1$ *and* $\mathbf{BR-TMP}_2$ *have the same objective value,i.e.,*

$$r_{\min}^*(\mathbf{BR-TMP}_1) = r_{\min}^*(\mathbf{BR-TMP}_2).$$

The proof is straightforward and we omit its discussion.

3.3.3 Near Optimal Approximation

In this paper, we exploit the unique mathematical structure of our MINLP problem and develop a novel near-optimal solution procedure, with a performance guarantee. Note that in Problem

$\mathbf{BR-TMP}_2$, the only nonlinear constraints are the link capacity constraints (3.3.11), which involve the log function. To address this problem, we adopt a piece-wise linear approximation technique proposed by Jiang et al. [42] to transform the nonlinear constraints to linear constraints. The main idea is as follows. We first use a set of linear segments to approximate the log term in (3.3.11) while ensuring that the linear approximation error does not exceed a specific threshold ϵ. Subsequently, the nonlinear constraints in $\mathbf{BR-TMP}_2$ are replaced by a set of linear constraints. Denote the linearized optimization problem as $\mathbf{BR-TMP}_3$, which is a mixed-integer linear problem (MILP). Since MILP problems are relatively easier to solve than MINLP problems, we can efficiently apply a solver such as CPLEX [98] to obtain a solution efficiently.

We will show that solving $\mathbf{BR-TMP}_3$ gives us a near-optimal solution to the original problem $\mathbf{BR-TMP}_1$. Denote ψ as the desired performance gap for the near-optimal solution, i.e., the difference in the objective values between the optimal solution and the near-optimal solution to OPT. We analyze the relationship between the performance gap ψ and the linear approximation error ϵ. Specifically, for a desired performance gap ψ, we compute the maximum allowed linear approximation error ϵ, and accordingly, we derive the linear approximation constraints and construct $\mathbf{BR-TMP}_3$. The solution to $\mathbf{BR-TMP}_3$ then provides a near-optimal solution with the performance guarantee ψ. We summarize the above steps in Fig. 3.1, and We provide details for the steps in the remainder of this section.

Piece-wise Linear Approximation. We will employ the scheme proposed in [42] by Jiang et al. for piece-wise linear approximation. Denote $\gamma_l^q[t]^{\max} = \frac{G_{\mathrm{Tx}(l)\mathrm{Rx}(l)}P_{\max}}{A_l^q[t]P_n}$ as the maximum effective SINR for the data stream q of link l in time slot t. This scheme introduces a set of consecutive linear segments to approximate $\ln(1+\gamma_l^q[t])$ for $\gamma_l^q[t] \in [0, \gamma_l^q[t]^{\max}]$(see Fig. 3.2). We can rewrite the nonlinear constraints in (3.3.11) as follows:

$$c_l^q[t] = \frac{W}{\ln 2} \cdot \ln(1+\gamma_l^q[t]), (1 \le l \le L, 1 \le q \le M, 1 \le t \le T) \ . \tag{3.3.12}$$

The nonlinear term in (3.3.12) is $\ln(1+\gamma_l^q[t])$. The range of $\gamma_l^q[t]$ is $[0, \gamma_l^q[t]^{\max}]$. Denote ϵ as the maximum allowed error for this linear approximation and let $K_l^q[t]$ be the number of linear segments needed to meet this error requirement. Denote $\gamma_l^q[t]^{(k)}$, $k = 0, 1, \ldots, K_l^q[t]$ as the $\gamma_l^q[t]$-

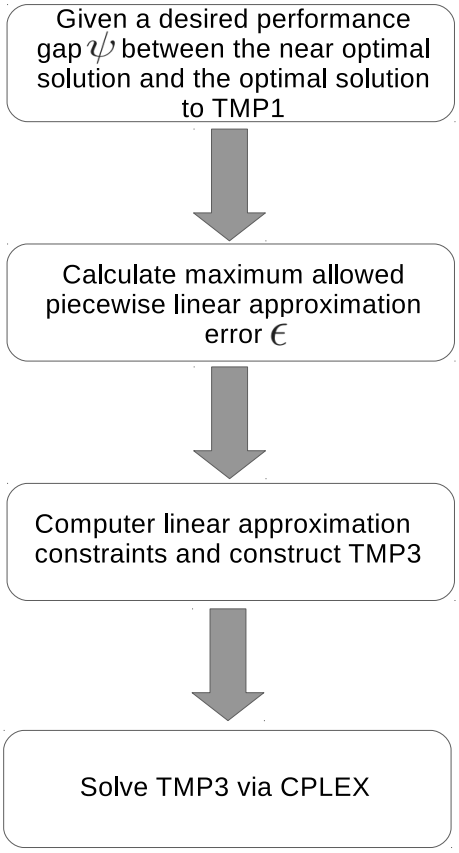


Figure 3.1: A flow chart for obtaining a near optimal solution to TMP1.

axis values of the endpoints for these $K_l^q[t]$ segments, with $\gamma_l^q[t]^{(0)} \equiv 0$ and $\gamma_l^q[t]^{(K_l^q[t])} \equiv \gamma_l^q[t]^{\max}$. Denote $a_l^q[t]^{(k)}$ as the slope of the k-th linear segments, i.e.,

$$a_l^q[t]^{(k)} = \frac{\ln(1+\gamma_l^q[t]^{(k)}) - \ln(1+\gamma_l^q[t]^{(k-1)})}{\gamma_l^q[t]^{(k)} - \gamma_l^q[t]^{(k-1)}} \,. \tag{3.3.13}$$

Denote $g_l^q[t]^{(k)}(\gamma_l^q[t])$ as the k-th linear approximation segment (see Fig. 3.3), which can be represented as follows. For $\gamma_l^q[t]^{(k-1)} \le \gamma_l^q[t] \le \gamma_l^q[t]^{(k)}$,

$$g_l^q[t]^{(k)}(\gamma_l^q[t]) = a_l^q[t]^{(k)} \cdot (\gamma_l^q[t] - \gamma_l^q[t]^{(k-1)}) + \ln(1+\gamma_l^q[t]^{(k-1)}). \tag{3.3.14}$$

The values of $\gamma_l^q[t]^{(0)}, \ldots \gamma_l^q[t]^{(K_l^q[t])}$ can by computed sequentially (for a give ϵ) using the algo-

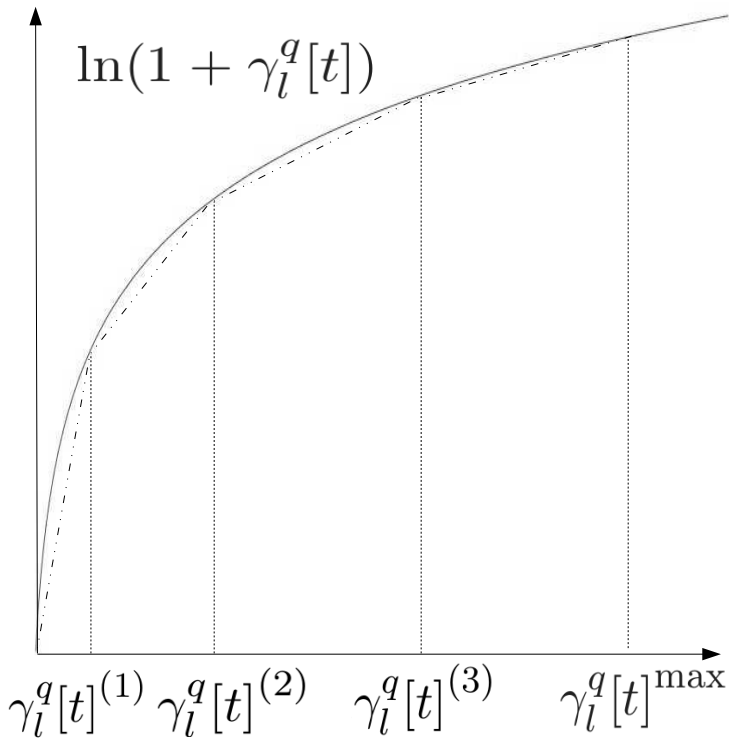


Figure 3.2: An illustration of piece-wise linear approximation with four line segments.

rithm proposed by Jiang et al. in [42]. The algorithm iteratively computes $a_l^q[t]^{(k)}$ that satisfies $-\ln(a_l^q[t]^{(k)}) + a_l^q[t]^{(k)}(1+\gamma_l^q[t]^{(k-1)}) - 1 - \ln(1+\gamma_l^q[t]^{(k-1)}) = \epsilon$ and with $a_l^q[t]^{(k)}$, computes $\gamma_l^q[t]^{(k)}$ that satisfies (3.3.13). The proposed algorithm in [42] further guarantees that the maximum approximation error of each linear segment is at most ϵ and *minimizes* the number of linear segments to approximate $\ln(1+\gamma_l^q[t])$, for a given approximation error bound ϵ for each linear segment. After piece-wise linear approximation of $\ln(1+\gamma_l^q[t])$, Constraint (3.3.11) can be replaced by the following set of constraints:

$$c_l^q[t] \leq \frac{W}{\ln 2} g_l^q[t]^{(k)}(\gamma_l^q[t]), (k = 0, 1, \ldots, K_l^q[t], 1 \leq l \leq L, 1 \leq q \leq M, 1 \leq t \leq T) \ ,$$

where $g_l^q[t]^{(k)}(\gamma_l^q[t])$ is given by (3.3.14). Substituting (3.3.14) into the above equation, for $(k = 0, 1, \ldots, K_l^q[t], 1 \leq l \leq L, 1 \leq q \leq M, 1 \leq t \leq T)$ we have

$$c_l^q[t] \leq \frac{W}{\ln 2}[a_l^q[t]^{(k)} \cdot (\gamma_l^q[t] - \gamma_l^q[t]^{(k-1)}) + \ln(1+\gamma_l^q[t]^{(k-1)})] \ . \tag{3.3.15}$$

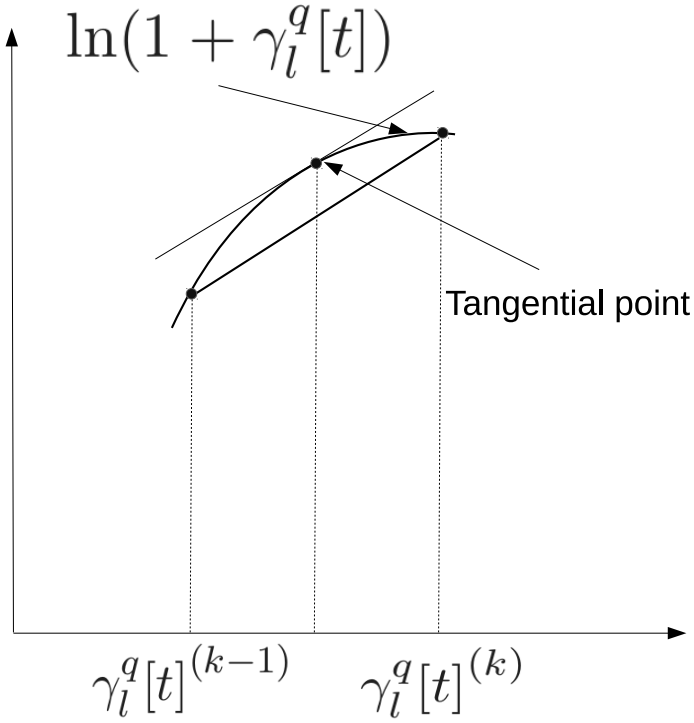


Figure 3.3: An illustration of maximum approximation error for the k-th linear segment.

By replacing the nonlinear constraints in (3.3.11) with the set of linear constraints in (3.3.15), we obtain the following revised formulation for $\mathbf{BR-TMP}_2$, which we denote as $\mathbf{BR-TMP}_3$

$\mathbf{BR-TMP}_3$ max $r_{\min}$

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

SM Constraints: (3.2.1),(3.2.2);

IC Constraints: (3.2.3);

Operation mode constraints: (3.2.4)–(3.2.6);

Dummy Data Stream Constraints: (3.3.1)–(3.3.3);

Data Stream Power Allocation Constraints: (3.3.4);

Link Capacity Constraints: (3.3.5)–(3.3.10),(3.3.15);

Link Rate Constraints: (3.2.11);

Flow balance constraints: (3.2.12), (3.2.13);

Variables: $x_i[t], y_i[t], z_l[t], \pi_i[t], m_i[t], p_l^q[t], \gamma_l^q[t], c_l^q[t], \lambda_l^q[t]$, $\omega_{il}^q[t], \nu_l^q[t], c_l[t], r_l(f), r(f)$.

Theorem 1. *The gap between the optimal objective values of* $\mathbf{BR-TMP}_3$ *and* $\mathbf{BR-TMP}_1$ *is no more than* $\frac{W \cdot M \cdot L \cdot \epsilon}{\ln 2}$.

The proof for the thorem is provided in appendix 3.3.4.

Based on Theorem 1, the following algorithm prescribes a near-optimal solution to $\mathbf{BR-TMP}_1$ with a performance guarantee.

Algorithm 3.3.1 An algorithm for obtaining near optimal solution.

Input: Given a desired performance gap psi for the solution to $\mathbf{BR-TMP}_1$.

1. Compute epsilon based on
$$\psi = \frac{W \cdot M \cdot L \cdot \epsilon}{\ln 2} \,. \tag{3.3.16}$$
2. Compute $a_l^q[t]^{(k)}$ and $\gamma_l^q[t]^{(k)}$ by the proposed Algorithm in [42].
3. Construct $\mathbf{BR-TMP}_3$ based on $a_l^q[t]^{(k)}$ and $\gamma_l^q[t]^{(k)}$.
4. Solve $\mathbf{BR-TMP}_3$ optimally using an MILP package (e.g.,CPLEX).

Upon the completion of algorithm 3.3.1, we have a near-optimal solution to $\mathbf{BR-TMP}_1$ with a guaranteed performance bound (no more than ψ from the optimal objective value).

3.3.4 Near-Optimality Proof

In this section, we present the proof to Theorem 1. In $\mathbf{BR-TMP}_3$, we linearize constraint (3.2.8) in $\mathbf{BR-TMP}_1$ by a piece-wise linear approximation scheme proposed in [42]. In this proof, we provide a bound for the gap between the optimal objective values of $\mathbf{BR-TMP}_1$ and $\mathbf{BR-TMP}_3$.

To better illustrate the idea, we first introduce the following notations. In $\mathbf{BR-TMP}_1$, we assume that there exists a scheduling and power assignment solution $\boldsymbol{\varphi} = (\mathbf{x}[t], \mathbf{y}[t], \mathbf{m}[t], \mathbf{z}[t], \mathbf{p}[t])$ that satisfies constraints (3.2.1)–(3.2.7), where $\boldsymbol{\pi}[t]$, $\mathbf{x}[t]$, $\mathbf{y}[t]$, $\mathbf{m}[t]$ and $\mathbf{z}[t]$ represent the vectors $[x_1[t],\ x_2[t], ...,\ x_N[t]]$, $[y_1[t],\ y_2[t], ...,\ y_N[t]]$, $[m_1[t],\ m_2[t], ...,\ m_N[t]]$, and $[z_1[t],\ z_2[t], ...,\ z_L[t]]$; $\mathbf{p}[t]$ represents matrix $[p_l^q[t]]_{L\times Q}$. Then we define $\bar{\boldsymbol{\phi}} = (\boldsymbol{\varphi}, \bar{r}_l(f), \bar{r}(f))$ as a feasible solution to $\mathbf{BR-TMP}_1$, where $(\bar{r}_l(f), \bar{r}(f))$ is the optimal solution to the following linear program $\mathbf{BR-TMP}_1(\boldsymbol{\varphi})$.

$$
\begin{aligned}
\mathbf{BR-TMP}_1(\boldsymbol{\varphi}) \quad \max \quad & r_{\min} \\
\text{s.t.} \quad & r_{\min} \le r(f) && (f \in \mathcal{F}) \\
& \sum_{l\in\mathcal{L}_i^{\text{out}}} r_l(f) = r(f) && (i = s(f), f \in \mathcal{F}) \\
& \sum_{l\in\mathcal{L}_i^{\text{in}}} r_l(f) = \sum_{l\in\mathcal{L}_i^{\text{out}}} r_l(f) && (1\le i\le N, i\ne s(f), i\ne d(f), f \in \mathcal{F}) \\
& \sum_{f\in\mathcal{F}} r_l(f) \le \bar{c}_l && (1 \le l \le L) \\
& r_{\min}, r_l(f), r(f) \ge 0 && (l \in \mathcal{L}, f \in \mathcal{F})
\end{aligned}
$$

where $\bar{c}_l = \frac{1}{T}\sum_{t=1}^{T} \bar{c}_l[t]$. Note that once we fix the scheduling and power variables in $\mathbf{BR-TMP}_1$ to values in $\boldsymbol{\varphi}$, we can get the value of $\bar{c}_l$ by constraints (3.2.8)–(3.2.10). Therefore, given $\boldsymbol{\varphi}$, $\mathbf{BR-TMP}_1(\boldsymbol{\varphi})$ is an LP.

Based on the feasible solution $\bar{\boldsymbol{\phi}} = (\boldsymbol{\varphi}, \bar{r}_l(f), \bar{r}(f))$ to $\mathbf{BR-TMP}_1$, we define a feasible solution $\hat{\boldsymbol{\phi}} = (\boldsymbol{\varphi}, \hat{r}_l(f), \hat{r}(f))$ to $\mathbf{BR-TMP}_3$, where $(\hat{r}_l(f), \hat{r}(f))$ is the optimal solution to $\mathbf{BR-TMP}_3$ when the scheduling and power variables are fixed to values in $\boldsymbol{\varphi}$. That is, $(\hat{r}_l(f), \hat{r}(f))$ is the optimal solution to the following LP.

$$
\begin{aligned}
\mathbf{BR-TMP}_3(\boldsymbol{\varphi}) \quad \max \quad & r_{\min} \\
\text{s.t.} \quad & r_{\min} \le r(f) && (f \in \mathcal{F}) \\
& \sum_{l \in \mathcal{L}_i^{\text{out}}} r_l(f) = r(f) && (i = s(f), f \in \mathcal{F}) \\
& \sum_{l \in \mathcal{L}_i^{\text{in}}} r_l(f) = \sum_{l \in \mathcal{L}_i^{\text{out}}} r_l(f) && (1 \le i \le N, i \ne s(f), i \ne d(f), f \in \mathcal{F}) \\
& \sum_{f \in \mathcal{F}} r_l(f) \le \hat{c}_l && (1 \le l \le L) \\
& r_{\min}, r_l(f), r(f) \ge 0 && (l \in \mathcal{L}, f \in \mathcal{F})
\end{aligned}
$$

where $\hat{c}_l = \frac{1}{T} \sum_{t=1}^{T} \hat{c}_l[t]$. According to constraint (3.3.5), we know that $\hat{c}_l[t] = \sum_{q=1}^{M} \hat{c}_l^q[t]$, where $\hat{c}_l^q[t]$ is the linear approximation of link l's achievable rate of qth data stream in time slot t. Recall that in $\mathbf{BR-TMP}_3$, we use constraint (3.3.15) to replace constraint (3.2.8) in $\mathbf{BR-TMP}_1$. When scheduling variables $(\mathbf{x}[t], \mathbf{m}[t])$ and link l's power assignment $p_l^q[t]$ are fixed at the values given in $\boldsymbol{\varphi}$, we can obtain the value of $\gamma_l^q[t]$ by constraint (3.2.9). Therefore, we can determine which line segment is used in the linear approximation of $\ln(1 + \gamma_l^q[t])$. Suppose that the k-th segment is used, then $\hat{c}_l^q[t]$ can be obtained as $\hat{c}_l^q[t] = \frac{W}{\ln 2} \cdot g_l^q[t]^{(k)}(\gamma_l^q[t])$. As a result, with given $\boldsymbol{\varphi}$, $\mathbf{BR-TMP}_3(\boldsymbol{\varphi})$ is an LP.

Now, we quantify the gap between the optimal solutions to $\mathbf{BR-TMP}_1$ and $\mathbf{BR-TMP}_3$ by the following two steps. First, we show that for any feasible scheduling and power assignment solution $\boldsymbol{\varphi}$, the gap between the feasible objective values corresponding to $\bar{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\phi}}$ is $\frac{W \cdot L \cdot M \cdot \epsilon}{\ln 2}$. Then, we show that the gap between the optimal objective values of $\mathbf{BR-TMP}_1$ and $\mathbf{BR-TMP}_3$ is also bounded by $\frac{W \cdot L \cdot M \cdot \epsilon}{\ln 2}$. This gap is characterized in the dual domain of problem $\mathbf{BR-TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{BR-TMP}_3(\boldsymbol{\varphi})$ as follows.

Step 1: For a given $\boldsymbol{\varphi}$, we denote $\bar{r}$ as the feasible objective value corresponding to solu-

tion $\bar{\phi}$ to $\mathbf{BR}-\mathbf{TMP}_1$, and $\hat{r}$ as the feasible objective value corresponding to solution $\hat{\phi}$ to $\mathbf{BR}-\mathbf{TMP}_3$. In this step, we will show that $\bar{r}-\hat{r} \leq \frac{W \cdot L \cdot M \cdot \epsilon}{\ln 2}$.

Denote the dual problem of $\mathbf{BR}-\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{BR}-\mathbf{TMP}_3(\boldsymbol{\varphi})$ as $\mathbf{D}_1(\boldsymbol{\varphi})$ and $\mathbf{D}_3(\boldsymbol{\varphi})$, respectively. Since the only difference between the formulation of $\mathbf{BR}-\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{BR}-\mathbf{TMP}_3(\boldsymbol{\varphi})$ is the constant term $\bar{c}_l$ and $\hat{c}_l$, it is easy to see that $\mathbf{D}_1(\boldsymbol{\varphi})$ and $\mathbf{D}_3(\boldsymbol{\varphi})$ have the same constraints, but different objective functions.

Denote the dual variables corresponding to the first set of constraints in $\mathbf{BR}-\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{BR}-\mathbf{TMP}_3(\boldsymbol{\varphi})$ as $w(f), f \in \mathcal{F}$. Denote the dual variables corresponding to the second set of constraints in $\mathbf{BR}-\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{BR}-\mathbf{TMP}_3(\boldsymbol{\varphi})$ as $v(f), f \in \mathcal{F}$. Denote the dual variables corresponding to the third set of constraints in $\mathbf{BR}-\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{BR}-\mathbf{TMP}_3(\boldsymbol{\varphi})$ as $y_i(f), f \in \mathcal{F}, i \in \mathcal{N}, i \neq s(f), d(f)$. Denote the dual variables corresponding to the fourth set of constraints in $\mathbf{BR}-\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{BR}-\mathbf{TMP}_3(\boldsymbol{\varphi})$ as $h_l, l \in \mathcal{L}$. Then $\mathbf{D}_1(\boldsymbol{\varphi})$ can be written as follows:

$$
\begin{array}{lll}
\mathbf{D}_1(\boldsymbol{\varphi}) & \min & \sum_{l \in \mathcal{L}} \bar{c}_l h_l \\
& \text{s.t.} & w(f) \geq 1 \qquad\qquad (f \in \mathcal{F}) \\
& & -w(f)-v(f) \geq 0 \qquad (f \in \mathcal{F}) \\
& & v(f)+y_j(f)+h_l \geq 0 \quad (f \in \mathcal{F}, l \equiv (s(f), j) \in \mathcal{L}, j \neq d(f)) \\
& & y_j(f)-y_i(f)+h_l \geq 0 \\
& & (f \in \mathcal{F}, l \equiv (i, j) \in \mathcal{L}, i \neq s(f), i \neq d(f), j \neq s(f), j \neq d(f)) \\
& & -y_i(f)+h_l \geq 0 \qquad (f \in \mathcal{F}, l \equiv (i, d(f)) \in \mathcal{L}, i \neq s(f)) \\
& & v(f)+h_l \geq 0 \qquad (f \in \mathcal{F}, l \equiv (s(f), d(f)) \in \mathcal{L}) \\
& & w(f), h_l \geq 0, v(f) \text{ and } y_i(f) \text{ unrestricted } (f \in \mathcal{F}, i \in N, i \neq s(f), d(f), l \in \mathcal{L})
\end{array}
$$

The dual problem $\mathbf{D}_3(\boldsymbol{\varphi})$ can be written as follows:

$$
\begin{array}{lll}
\mathbf{D}_3(\boldsymbol{\varphi}) & \min & \sum_{l \in \mathcal{L}} \hat{c}_l h_l \\
& \text{s.t.} & \text{Same constraints as in } \mathbf{D}_1(\boldsymbol{\varphi})
\end{array}
$$

Since $\mathbf{D}_1(\boldsymbol{\varphi})$ and $\mathbf{D}_3(\boldsymbol{\varphi})$ share the same constraints, they have the same feasible region. If h_l^* is (part of) an optimal solution to $\mathbf{D}_3(\boldsymbol{\varphi})$, then since $\bar{r}$ and $\hat{r}$ is the optimal objective value of $\mathbf{BR-TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{BR-TMP}_3(\boldsymbol{\varphi})$ respectively, we have

$$\bar{r} - \hat{r} \leq \sum_{l \in \mathcal{L}} \bar{c}_l h_l^* - \sum_{l \in \mathcal{L}} \hat{c}_l h_l^* = \sum_{l \in \mathcal{L}} (\bar{c}_l - \hat{c}_l) h_l^* \ .$$

Now we quantify the gap between $\bar{c}_l$ and $\hat{c}_l$. Since the maximum error of our linear approximation is ϵ, we know that

$$\bar{c}_l^q[t] - \hat{c}_l^q[t] \leq \frac{W}{\ln 2}\epsilon \ .$$

Then based on constraint (3.3.5), we have

$$\bar{c}_l - \hat{c}_l = \frac{1}{T}\sum_{t=1}^{T}\sum_{q=1}^{M}(\bar{c}_l^q[t] - \hat{c}_l^q[t]) \leq \frac{W \cdot M}{\ln 2}\epsilon \ .$$

Therefore,

$$\bar{r} - \hat{r} \leq \frac{W \cdot M}{\ln 2}\epsilon \sum_{l \in \mathcal{L}} h_l^* \ . \tag{3.3.17}$$

According to the definition of marginal rate of change of dual variables in [6], h_l^* is upper bounded by the largest possible change of the optimal objective value $\hat{r}$ of $\mathbf{BR-TMP}_3(\boldsymbol{\varphi})$ with respect to the right-hand side $\hat{c}_l$. Since the objective function is the achievable session rate, and a small marginal $\triangle$-change (say, increase) in the capacity of a link can at most increase $\triangle$ units of session rate, we have

$$h_l^* \leq 1 \ \ (l \in L) \ . \tag{3.3.18}$$

Combining (3.3.17) and (3.3.18), we have

$$\bar{r} - \hat{r} \leq \frac{W \cdot M \cdot L \cdot \epsilon}{\ln 2} \ .$$

This completes Step 1 of the proof.

Step 2: Denote $\boldsymbol{\phi}_1^*$ and r_1^* as the optimal solution and its corresponding optimal objective value of $\mathbf{BR-TMP}_1$. Denote $\boldsymbol{\phi}_3^*$ and r_3^* as the optimal solution and its corresponding optimal objective value of $\mathbf{BR-TMP}_3$. In this step, we show that $r_1^* - r_3^* \leq \frac{W \cdot M \cdot L \cdot \epsilon}{\ln 2}$.

Since the optimal solution $\boldsymbol{\phi}_1^*$ to $\mathbf{BR}-\mathbf{TMP}_1$ is a particular case of feasible solution $\bar{\boldsymbol{\phi}}$, we know that there exists a corresponding feasible solution to $\mathbf{BR}-\mathbf{TMP}_3$, denoted as $\tilde{\boldsymbol{\phi}}_3$, with a corresponding objective value $\tilde{r}_3$. From Step 1, we know that

$$r_1^* - \tilde{r}_3 \leq \frac{W \cdot M \cdot L \cdot \epsilon}{\ln 2} \ .$$

Since r_3^* is the optimal objective value of $\mathbf{BR}-\mathbf{TMP}_3$, we know that $r_3^* \geq \tilde{r}_3$. Therefore,

$$r_1^* - r_3^* \leq \frac{W \cdot M \cdot L \cdot \epsilon}{\ln 2} \ .$$

This completes the proof. □

3.4 Numerical Results

In this section, we present some numerical results to study the performance of proposed multi-hop MIMO FD described in Section 3.2–3.3. The goal of this effort is twofold. First, we want to show how a solution to the $\mathbf{BR}-\mathbf{TMP}_1$ formulation looks like for an example network. By studying the details of our solution for an example network, we will develop some quantitative understanding on how the joint integration of FD and MIMO DoF IC can significantly improve the network throughput. Second, we want to perform a comparison study between MIMO FD, MIMO HD, MIMO FD without IC, and MIMO HD without IC. Our numerical result will show that MIMO FD does not flourish its full potential without IC due to increase of interference power by enabling FD links.

3.4.1 Simulation Settings

We consider a randomly generated multi-hop wireless network with 30 nodes that are distributed in a 100×100 area. For generality, we normalize all units for distance, data rate, bandwidth, and power with appropriate dimensions. At the network layer, minimum-hop routing is employed. The topology of the network is shown in Fig. 3.4. There are 2 active sessions in the network with each

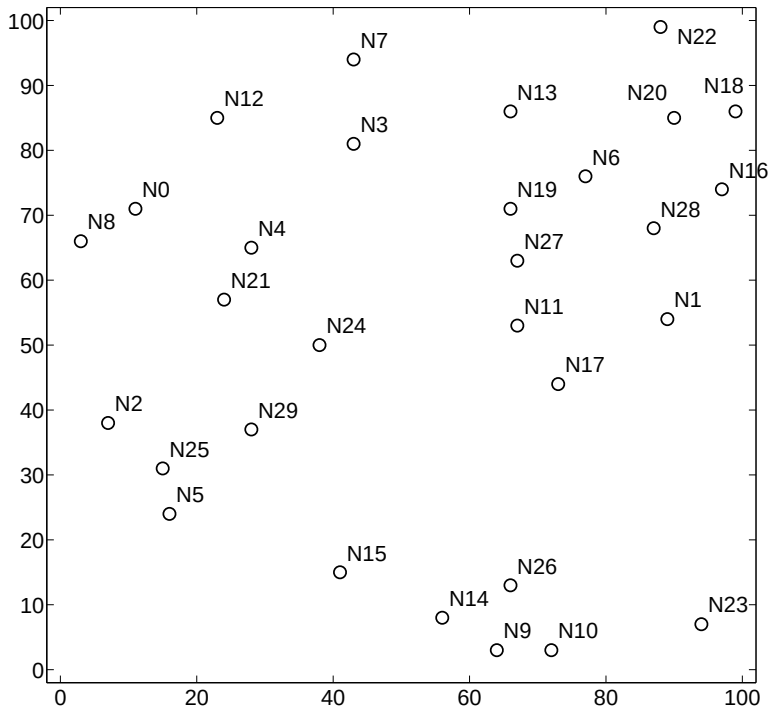


Figure 3.4: A 30-node network in a 100×100 area.

Table 3.2: Source node and destination node in the 30-node network.

Session f	Source Node $s(f)$	Dest. Node $d(f)$
1	7	22
2	11	2

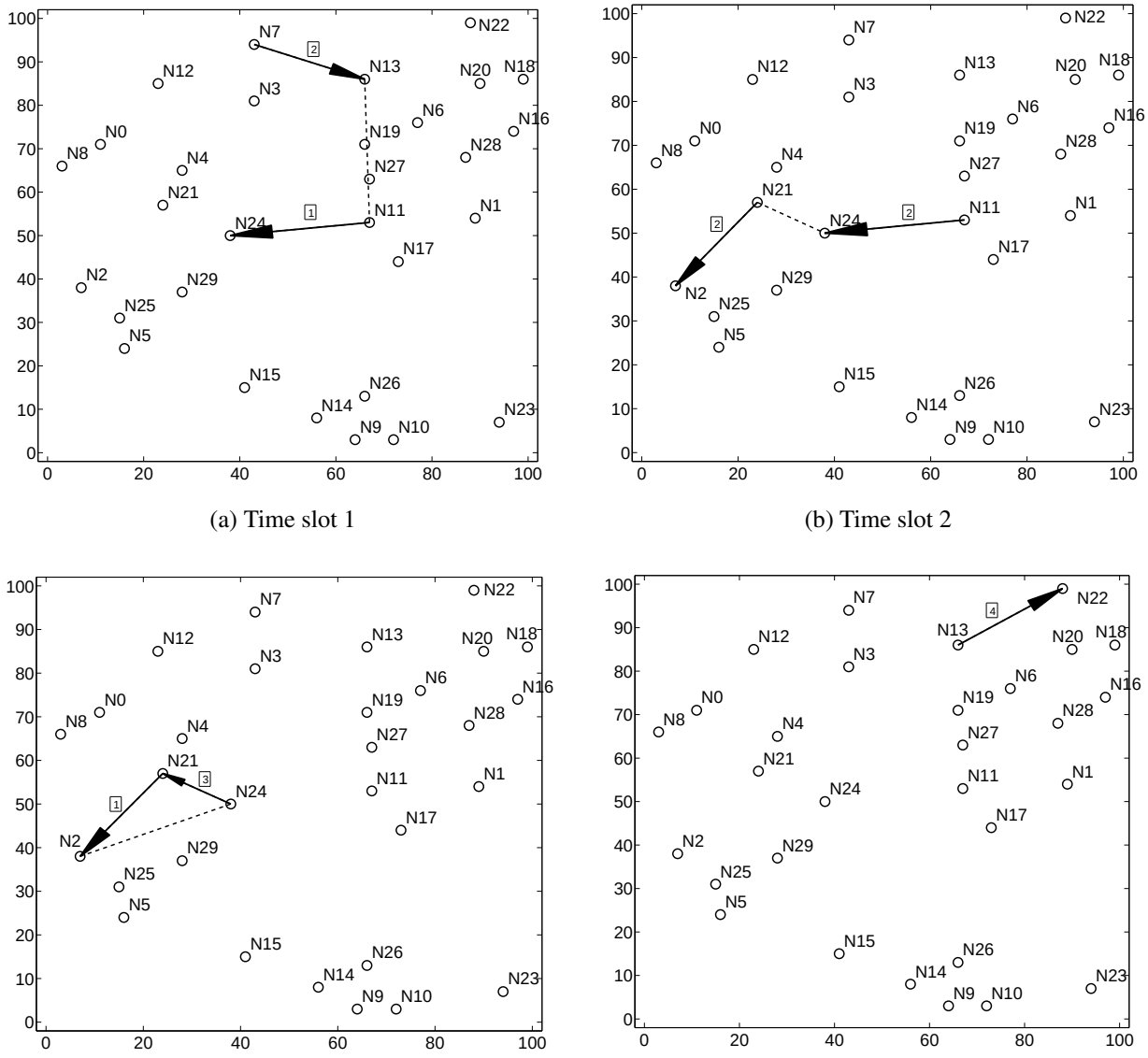
session's source node and destination node given in Table 3.2. Each node is equipped with $M = 4$ antennas. We assume the bandwidth $W = 1$. The transmit power for each node is set to 100. The path loss parameter $G_{\mathrm{Tx}(k)\mathrm{Rx}(k)} = [20\log_{10}(d) + 38.25]$ (in dB) [101], where d is the distance between $\mathrm{Tx}(k)$ and $\mathrm{Rx}(k)$. We assume the self-interference parameter $\alpha = 110$dB [7] and ambient noise power $P_n = -90$dBm. The number of time slots in a frame is $T = 4$. The worst case upper bound value for $A_l^q[t]$ is 7.3753. We set the maximum acceptable performance gap between the optimal objective values of $\mathbf{BR-TMP}_1$ and the linear approximation $\mathbf{BR-TMP}_3$ as $\psi = 0.01$, and apply piece wise linear approximation algorithm proposed in [42]. Based on (3.3.16), we compute $\epsilon = \frac{\ln 2 \cdot \psi}{W \cdot M \cdot L} = 3.4657 \times 10^{-4}$.

3.4.2 A Case Study

We will present the detail solution to throughput maximization problem for this network instance depicted in Fig. 3.4 for 4 different models: i) MIMO FD, ii)MIMO HD, iii)MIMO FD without IC, and iv)MIMO HD without IC.

MIMO FD. Fig 3.5 shows the set of active links and the number of data streams per link in each time slot in the solution. As shown in Fig. 3.5c, N_{21} in operating in FD mode in time slot 3 and the rest of the nodes in all other time slots are operating in half duplex. Fig 3.6 shows the combined results for all time slots. Table 3.4 shows the details of DoF allocation for SM, aggregated capacity over data streams, and session's rate (i.e., aggregate capacity averaged over a 4-time-slot frame) for all time slots for MIMO full duplex. As shown in Table 3.4, the objective value $r_{\min}$ (minimum session rate) for this network scenario is 2.99428 under MIMO FD. Table 3.3 shows the details of IC for time slot $T = 3$. Receivers are responsible to perform IC for strong interferring signals. The number of DoFs used for IC is shown in the last column of Table 3.3.

MIMO HD. To show the benefits of the full duplex MIMO communication in multi-hop network, we compare our solution to half duplex case. Figure 3.7 shows the set of active links and the number of data streams per link in each time slot in the solution. Figure 3.8 shows the combined



(a) Time slot 1

(b) Time slot 2

(c) Time slot 3

(d) Time slot 4

Figure 3.5: Full Duplex: Scheduled links, DoFs allocation on each link, and interference pattern in time slots 1 to 4, respectively.

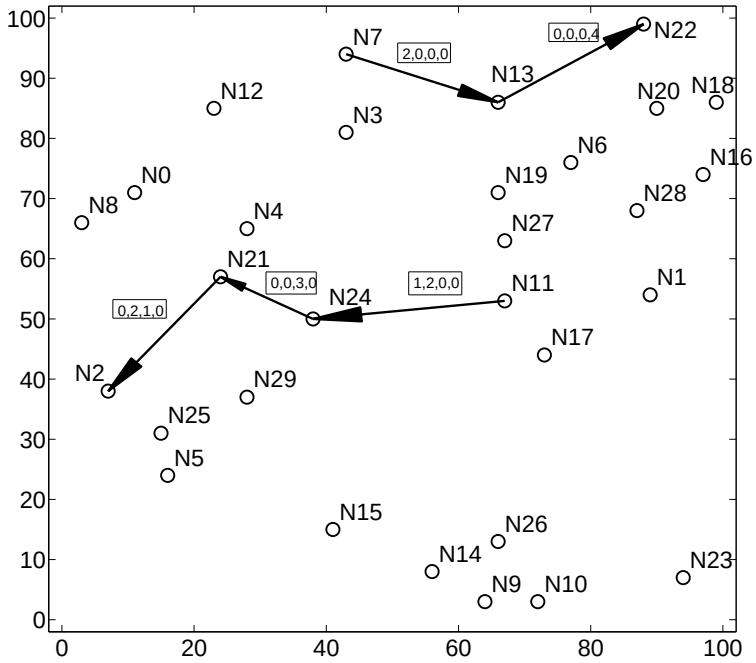


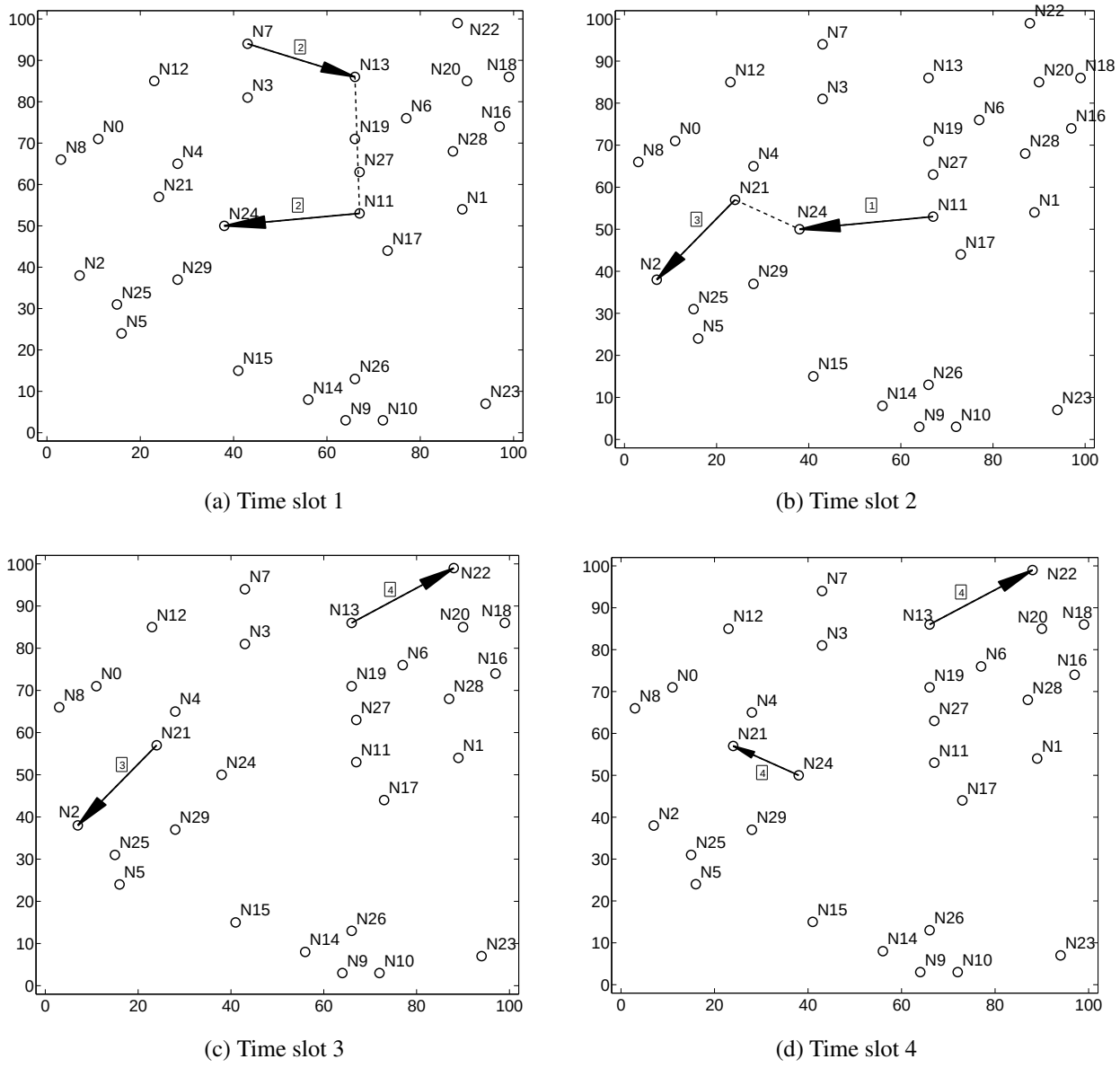
Figure 3.6: Full Duplex: Shows the combined results for all time slots (with the number of data streams for each time slot on the link shown in a box).

Table 3.3: Full Duplex: DoF allocation for SM/IC for time slot $T = 3$.

Node i	DoF for SM	DoF for IC
N_{24}	3	0
N_{21}	3 at Rx 1 at Tx	0
N_2	1	3

Table 3.4: Full Duplex: DOF allocation for SM, aggregated capacity over data streams, and session's rate for all time slots.

Session	Link	Time Slot	DoF for SM	Capacity	Session's Rate
1	$N_7 \rightarrow N_{13}$	1	2	18.8148	2.99428
		2	0	0	
		3	0	0	
		4	0	0	
	$N_{13} \rightarrow N_{22}$	1	0	0	
		2	0	0	
		3	0	0	
		4	4	18.5456	
2	$N_{11} \rightarrow N_{24}$	1	1	0.3924	2.99428
		2	2	17.7789	
		3	0	0	
		4	0	0	
	$N_{24} \rightarrow N_{21}$	1	0	0	
		2	0	0	
		3	3	12.1916	
		4	0	0	
	$N_{21} \rightarrow N_2$	1	0	0	
		2	2	0.9680	
		3	1	11.009	
		4	0	0	



(a) Time slot 1

(b) Time slot 2

(c) Time slot 3

(d) Time slot 4

Figure 3.7: Half Duplex: Scheduled links, DoFs allocation on each link, and interference pattern in time slots 1 to 4, respectively.

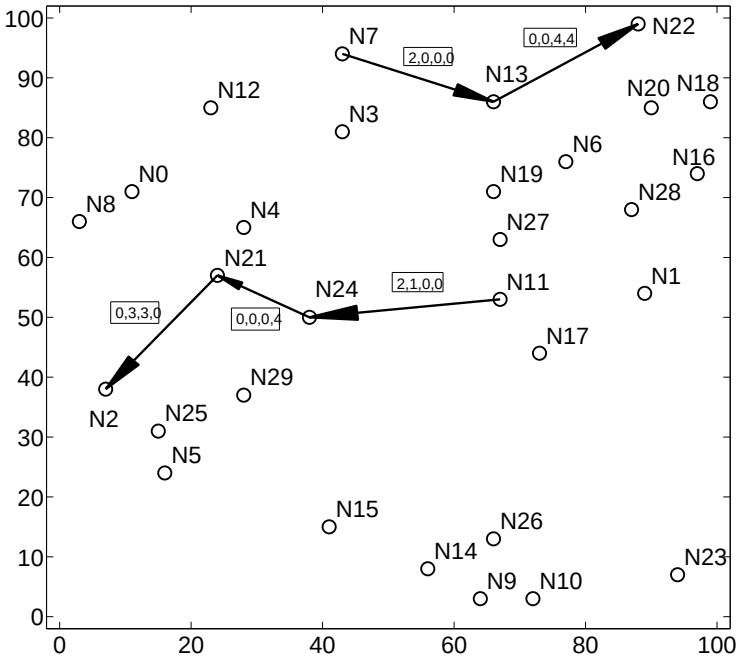


Figure 3.8: Half Duplex: Shows the combined results for all time slots (with the number of data streams for each time slot on the link shown in a box).

Table 3.5: Half Duplex: DoF allocation for SM/IC for time slot $T = 2$.

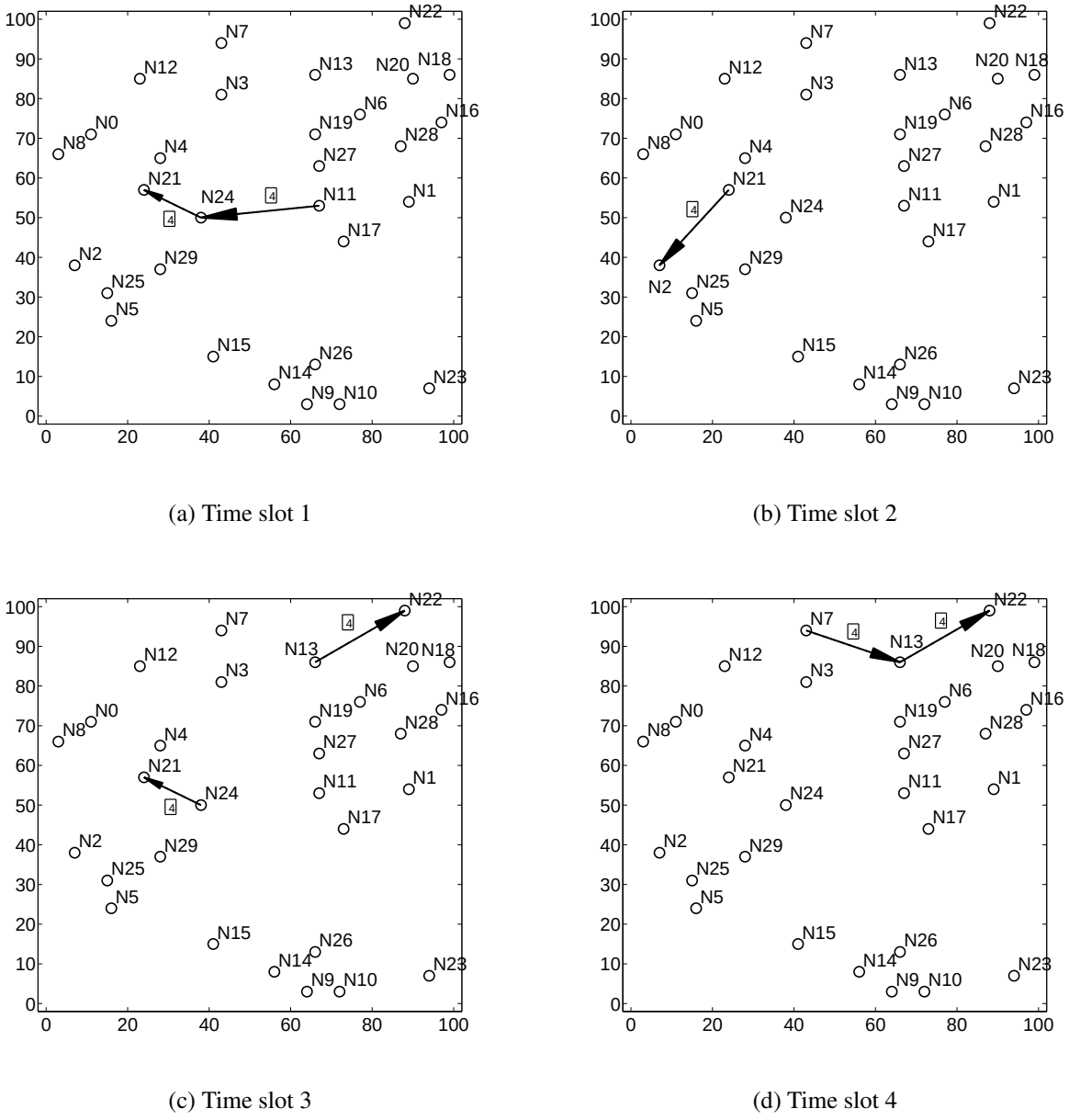
Node i	DoF for SM	DoF for IC
N_{11}	1	0
N_{24}	1	3
N_{21}	3	0
N_2	3	0

results for all time slots. Table 3.6 shows the details of DoF allocation for SM, aggregated capacity over data streams, and session's rate for all time slots for MIMO half duplex. Table 3.5 shows the details of IC for time slot $T = 2$. Receivers are responsible to perform IC for strong interfering signals. The number of DoFs used for IC is shown in the last column of Table 3.5. As shown in Table 3.6, the objective value $r_{\min}$ (minimum session rate) for this network scenario is 0.441784. Therefore, the ratio of the objective value of MIMO FD to MIMO HD is 6.78 (which is significantly

Table 3.6: Half Duplex: DOF allocation for SM, aggregated capacity over data streams, and session's rate for all time slots.

Session	Link	Time Slot	DoF for SM	Capacity	Session's Rate
1	$N_7 \rightarrow N_{13}$	1	2	11.1496	0.441784
		2	0	0	
		3	0	0	
		4	0	0	
	$N_{13} \rightarrow N_{22}$	1	0	0	
		2	0	0	
		3	4	1.52501	
		4	4	1.30248	
2	$N_{11} \rightarrow N_{24}$	1	2	0.419009	0.441784
		2	1	10.6223	
		3	0	0	
		4	0	0	
	$N_{24} \rightarrow N_{21}$	1	0	0	
		2	0	0	
		3	0	0	
		4	4	1.76713	
	$N_{21} \rightarrow N_2$	1	0	0	
		2	3	0.963369	
		3	3	1.32827	
		4	0	0	

larger than $2X$ expected benefit from FD). The joint modeling of full duplex, time slot scheduling, and IC unlock the potential of FD in MIMO multi-hop network as it opens the mathematical space for the optimization.



(a) Time slot 1

(b) Time slot 2

(c) Time slot 3

(d) Time slot 4

Figure 3.9: Full Duplex w/o DoF IC: Scheduled links, DoFs allocation on each link, and interference pattern in time slots 1 to 4, respectively.

MIMO FD without IC. The superior performance of MIMO FD compare to MIMO HD is mainly due to joint optimization of IC and FD, i.e., MIMO full duplex without IC does not have such a gain over MIMO half duplex. To show the above fact, we studied the performance of a

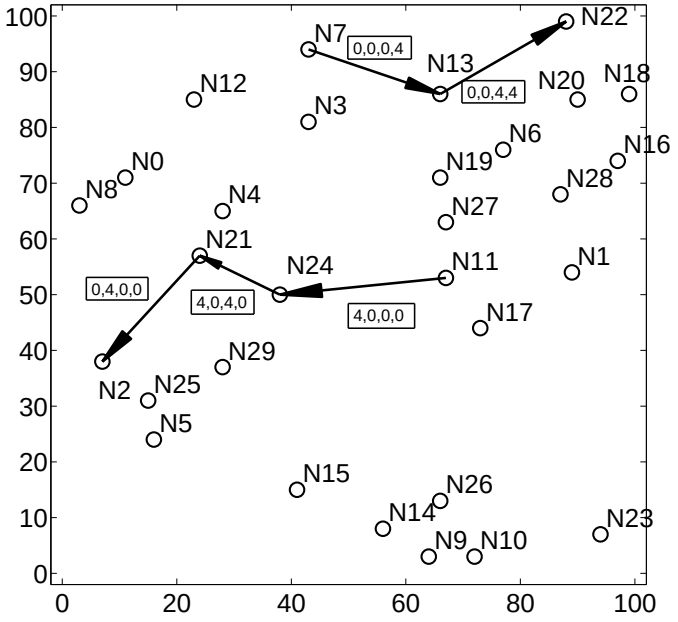


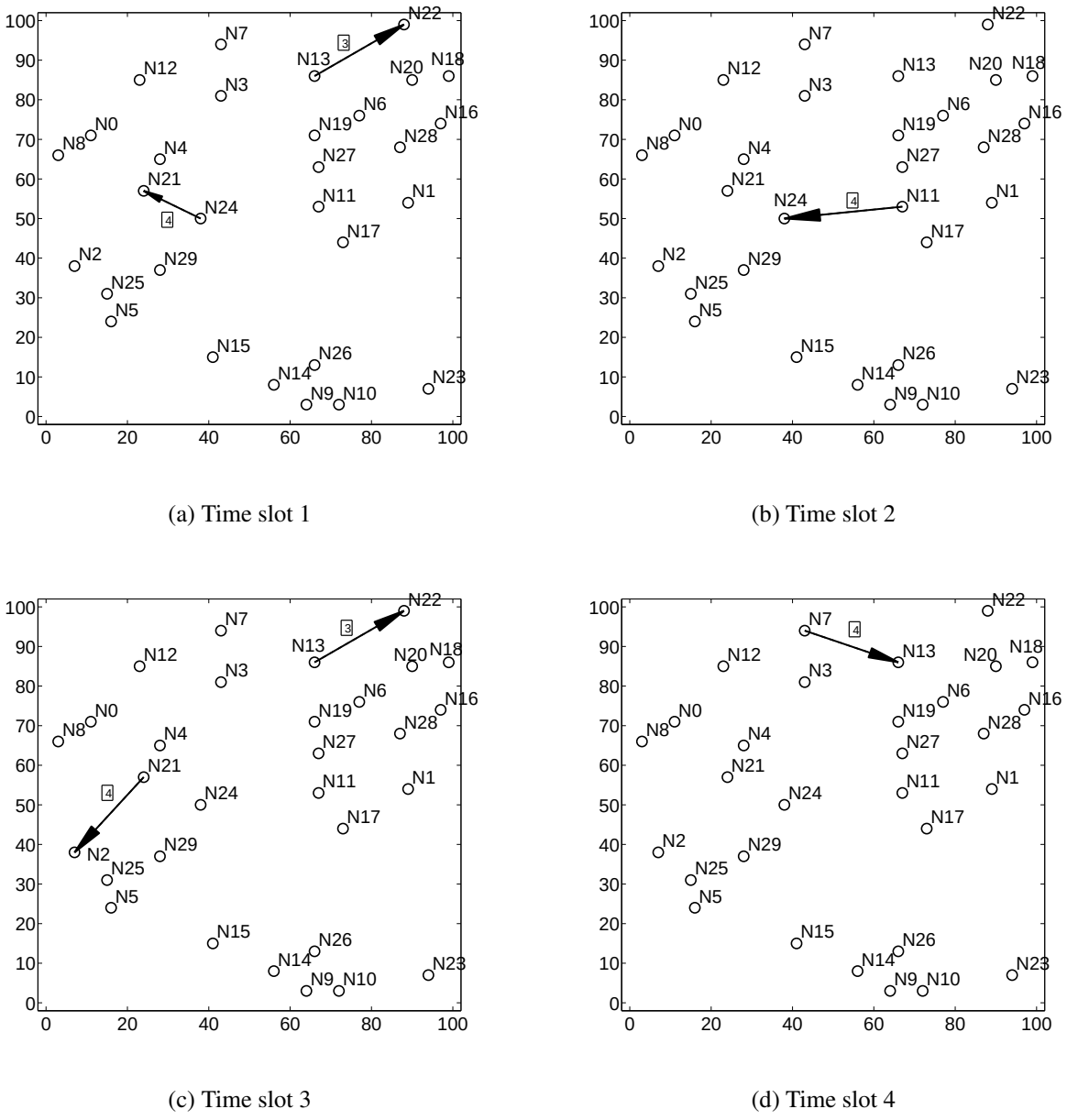
Figure 3.10: Full Duplex w/o DoF IC: Shows the combined results for all time slots (with the number of data streams for each time slot on the link shown in a box).

MIMO full duplex and MIMO half duplex without DoF IC under the same simulation settings. Figure 3.9 shows the set of active links and the number of data streams per link in each time slot in the solution. As shown in Fig. 3.9a and Fig. 3.9d, N_{24} in time slot 1 and N_{13} in time slot 4 are operating in full duplex mode and the rest of the nodes in all other time slots are operating in half duplex. Figure 3.10 shows the combined results for all time slots. Table 3.7 shows the details of DoF allocation for SM, aggregated capacity over data streams, and session's rate for all time slots for MIMO full duplex without IC. As shown in Table 3.7, the objective value $r_{\min}$ (minimum session rate) for this network scenario is 0.471389.

MIMO HD without IC. Figure 3.11 shows the set of active links and the number of data streams per link in each time slot in the solution. Figure 3.12 shows the combined results for all time slots. Table 3.8 shows the details of DoF allocation for SM, aggregated capacity over data streams, and session's rate (i.e., aggregate capacity averaged over a 4-time-slot frame) for all time slots for

Table 3.7: Full Duplex w/o IC: DOF allocation for SM, capacity, and session's rate for all time slots

Session	Link	Time Slot	DoF for SM	Capacity	Session's Rate
1	$N_7 \rightarrow N_{13}$	1	0	0	0.471389
		2	0	0	
		3	0	0	
		4	4	8.49779	
	$N_{13} \rightarrow N_{22}$	1	0	0	
		2	0	0	
		3	4	1.30248	
		4	4	0.583078	
2	$N_{11} \rightarrow N_{24}$	1	4	6.09706	0.471389
		2	0	0	
		3	0	0	
		4	0	0	
	$N_{24} \rightarrow N_{21}$	1	4	1.03962	
		2	0	0	
		3	4	1.75246	
		4	0	0	
	$N_{21} \rightarrow N_2$	1	0	0	
		2	4	11.0336	
		3	0	0	
		4	0	0	



(a) Time slot 1

(b) Time slot 2

(c) Time slot 3

(d) Time slot 4

Figure 3.11: Half Duplex w/o DoF IC: Scheduled links, DoFs allocation on each link, and interference pattern in time slots 1 to 4, respectively.

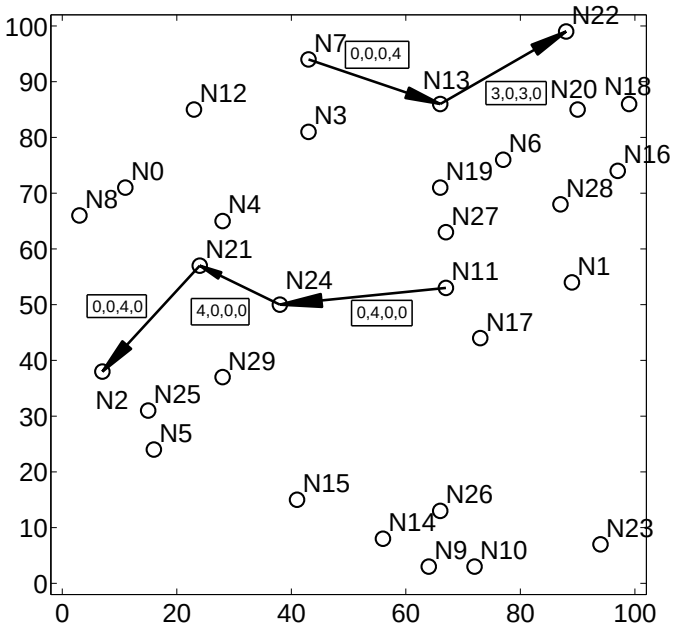


Figure 3.12: Half Duplex w/o DoF IC: Shows the combined results for all time slots (with the number of data streams for each time slot on the link shown in a box).

MIMO half duplex without IC. As shown in Table 3.8, the objective value $r_{\min}$ (minimum session rate) for this network scenario is 0.378504. Therefore, the ratio of the objective value of MIMO FD to MIMO HD is 1.2454. We can see that the benefits of FD without IC is marginal.

Case Study Summary. Although FD opens the scheduling space, it does not flourish the full potential of full duplex technology without integrating with IC to combat the additional interference introduced due to activation of full duplex links. Table 3.9 summarizes the findings of this case study. We can see there is a marginal increase in objective value of MIMO FD compare to MIMO HD when there is no IC. Moreover, the performance of MIMO HD with IC and MIMO FD w/o IC is similar. The true potential gain of FD is unleashed only when it is jointly integrated with IC in MIMO FD. For this case study, MIMO FD can performs more than 6 times better than MIMO HD. This increase is due to opening the mathematical space due to joint optimization of scheduling, IC, and FD.

Table 3.8: Half Duplex w/o IC: DOF allocation for SM, capacity, and session's rate for all time slots.

Session	Link	Time Slot	DoF for SM	Capacity	Session's Rate
1	$N_7 \rightarrow N_{13}$	1	0	0	0.378504
		2	0	0	
		3	0	0	
		4	4	11.1578	
	$N_{13} \rightarrow N_{22}$	1	3	1.24433	
		2	0	0	
		3	3	1.44765	
		4	0	0	
2	$N_{11} \rightarrow N_{24}$	1	0	0	0.378504
		2	4	10.6469	
		3	0	0	
		4	0	0	
	$N_{24} \rightarrow N_{21}$	1	4	1.76713	
		2	0	0	
		3	0	0	
		4	0	0	
	$N_{21} \rightarrow N_2$	1	0	0	
		2	0	0	
		3	4	1.51401	
		4	0	0	

Table 3.9: Objective Values of MIMO FD, MIMO HD, MIMO FD w/o IC, and MIMO HD w/o IC.

MIMO FD	MIMO HD	MIMO FD w/o IC	MIMO HD w/o IC
2.99428	0.441784	0.471389	0.378504

3.4.3 Complete Results

The previous section presents our results for a 30-node example network. In this section, we provide an additional set of results. We perform the same study for 30-node network, but for 50 randomly generated network instances, with each session's source and destination nodes being randomly selected among the nodes. Table 3.10 shows the results for both MIMO full duplex and half duplex. We find that the average ratio over the 50 instances is 2.2232. Fig 3.13 shows

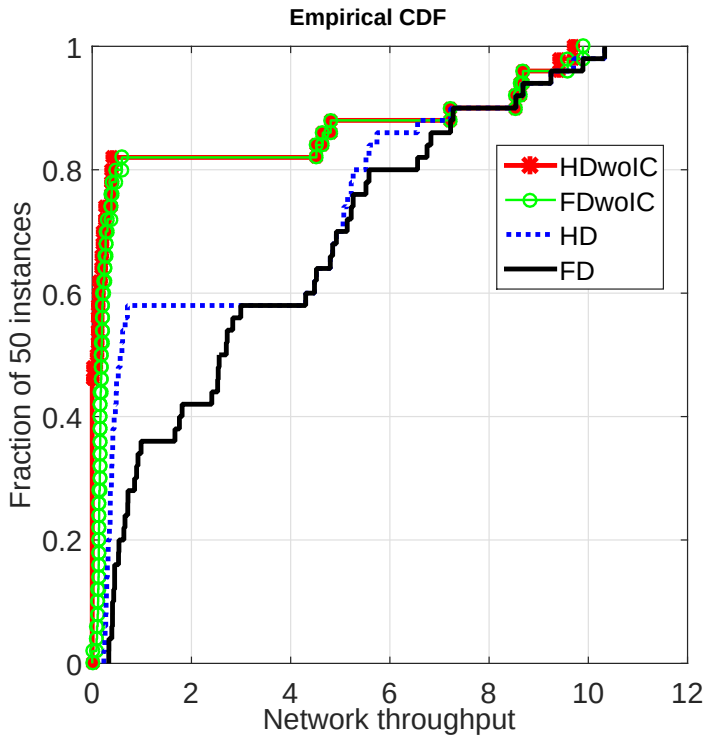


Figure 3.13: Cumulative distribution function of network throughput for Full Duplex, Half Duplex, Full Duplex without IC, and Half Duplex without IC

Table 3.10: Objective values under MIMO FD and MIMO HD for 30-node network over 50 instances.

IDX	HD	FD	Ratio	IDX	HD	FD	Ratio
1	0.490	1.757	3.587	26	5.057	6.749	1.335
2	5.148	7.273	1.413	27	0.322	0.452	1.403
3	9.703	9.891	1.019	28	6.558	6.558	1.000
4	0.393	2.559	6.510	29	5.739	6.827	1.190
5	0.347	0.409	1.179	30	0.319	2.410	7.567
6	0.599	0.716	1.196	31	0.361	0.661	1.832
7	0.235	0.409	1.741	32	5.071	5.140	1.014
8	4.921	4.921	1.000	33	8.539	8.539	1.000
9	8.683	8.683	1.000	34	0.292	0.638	2.182
10	0.525	0.895	1.705	35	0.416	0.539	1.296
11	4.799	4.799	1.000	36	0.260	0.395	1.517
12	7.224	7.224	1.000	37	0.305	0.452	1.481
13	0.442	2.994	6.778	38	0.280	2.721	9.723
14	0.484	2.528	5.227	39	0.387	0.721	1.865
15	4.480	4.480	1.000	40	5.579	5.579	1.000
16	0.253	0.530	2.093	41	0.563	2.833	5.030
17	0.286	0.334	1.166	42	0.613	1.668	2.721
18	5.220	5.220	1.000	43	0.363	0.984	2.710
19	5.260	5.260	1.000	44	4.302	4.302	1.000
20	9.246	9.246	1.000	45	0.709	0.922	1.300
21	0.272	0.336	1.234	46	0.664	2.703	4.073
22	0.413	0.854	2.069	47	0.398	1.812	4.557
23	0.467	2.540	5.442	48	4.519	4.519	1.000
24	4.846	4.846	1.000	49	10.329	10.329	1.000
25	5.518	5.518	1.000	50	0.364	0.433	1.189

the empirical cumulative distribution function of network throughput for MIMO FD, MIMO HD, MIMO FD w/o IC, and MIMO HD w/o IC. The CDF for MIMO FD w/o and HD w/o IC are very similar. By introducing IC, CDF for MIMO HD dramatically shifts down and to the right, which shows IC significantly improve the network performance. Finally, it can be seen that MIMO FD with DoF IC drags down the left portion of CDF even further (corresponding to low probability of observation of low network throughput). The latter shows the MIMO FD outperform all the other cases and significantly improve the network throughput. Fig. 3.14 shows the CDF for the ratio

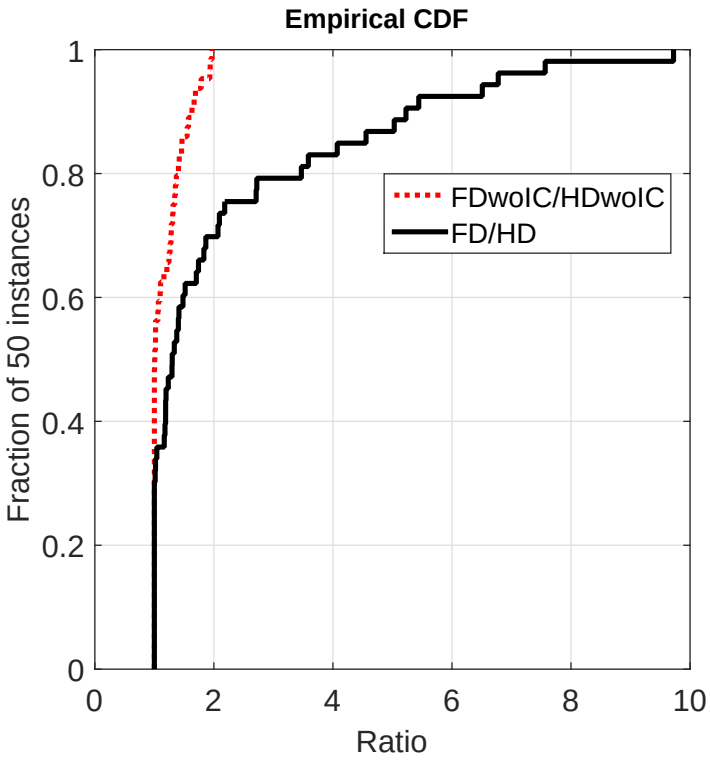


Figure 3.14: Cumulative distribution function for the ratio of FD/HD and FDwoIC/HDwoIC

of MIMO FD over MIMO HD and MIMO FD w/o IC over MIMO HD w/o IC. We can observe that the limit for the ratio of MIMO FD w/o IC over MIMO HD w/o IC is 2, however, there is a significant amount of network instances with ratios higher than 2 when we consider MIMO FD over MIMO HD (with IC). These results verify our claim in the previous case study; the true potential of FD is unleashed only when it is jointly optimized with IC to combat the additional interference caused by activating full duplex links.

3.5 Related Work

There are several proposed transceiver designs to implement a SISO full duplex in the literature. The simultaneous transmission and reception necessitates having both TX and RX radio frequency (RF) chains. There are two ways to separate these two RF chains [48]. One approach is to physically separate the transmit and receive antenna from one another in a separate-antenna system [15, 17, 66]. Alternatively, a circulator can be used in a shared-antenna system [37]. However, the above schemes are not capable to provide enough isolation to decode the signal of interest in the presence of the self interference caused by the signal leaking from its transmit RF chain to its receive RF chain. Therefore, the main challenge in implementing a full duplex system is how to cancel the self interference. Most of the self-interference cancellation schemes proposed for implementing full duplex systems usually consists of a combination of passive and active cancellation schemes. Passive schemes aim to suppress the self-interference before it enters the receive RF chain circuit by using various antenna techniques [3, 15, 17, 20, 21, 37, 44]. In active schemes the transmit signals is subtracted 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 [7, 8, 18, 17, 19, 39, 63]. Digital cancellation aims to subtract the estimate self-interference signal after ADC in the digital domain [1, 19, 45, 65]. The current state-of-art full duplex SISO transceiver design is proposed by Bharadia *et al.* in [7]. This design only requires a single antenna. It can achieve 110dB of self-interference cancellation (cancel the self interference to the noise level) and thus meets the requirements of 802.11 standards.

The benefit of SISO full duplex has been exploited in link layer by proposing various MAC protocols for both infrastructure networks and ad hoc networks. Several centralized MAC protocols are proposed for the infrastructure networks that manages the inter-user interference caused by simultaneous transmission and reception of AP and alleviate the hidden terminal problem [27, 39, 46, 66]. As for ad hoc networks, several distributed MAC protocols are proposed which support bi-directional links and two-hop relay transmissions [14, 32, 47, 88] as wells as multi-hop communications [75, 95].

Related work on cross-layer optimization for full duplex multi-hop networks includes [5, 22, 57, 64, 87]. In [5], Baranwal *et al.* studies the optimal number of multi-hop full duplex relay nodes which minimizes the outage probability. In [22], Fang *et al.* proposed distributed algorithms to solve user profit maximization problem and network power consumption minimization problem for multi-path routing in a multi-hop full duplex wireless network. Furthermore, the proofs of convergence and optimality gap for the proposed distributed algorithms are given. In [57], Mahboobi *et al.* investigated a joint power allocation and routing problem for a single source destination pair in a multihop full duplex relay network. Under optimal routing, the weighted sum of power in relay nodes is minimized while the end-to-end link outage probability is kept below a threshold. In [64], Ramirez *et al.* studied a joint routing and power allocation problem for a single source destination pair in a wireless full duplex network with imperfect self-interference cancellation. Xie *et al.* first showed that under protocol model, full-duplex cannot double network capacity (per-flow capacity) when the number of nodes goes to ∞ (asymptotic result) in [87]. Furthermore, the authors showed that the MAC-layer capacity (aggregated capacity of all links) gain of full duplex over half duplex is below 2 for a finite-sized network consists of single-hop bi-directional links.

3.6 Conclusions

The network throughput gain of FD radios over HD is unexpectedly marginal due to additional network interference introduced into the network, when both ends of the FD links are transmitting. In this paper, we exploit MIMO IC to combat the interference and unlock the true potential of FD in wireless MIMO multi-hop networks. By understanding the state-of-the-art MIMO transceiver architecture, we proposed a multi-hop MIMO FD model, where nodes can operate in FD mode and its spatial DoF resources can be used for SM and IC. We develop the necessary mathematical models to realize the proposed MIMO FD model. Together with scheduling and routing constraints, we develop a cross-layer optimization framework for a MIMO FD network. Our results show that MIMO FD with IC can have significant gain over MIMO HD in multi-hop network. However, MIMO FD w/o IC gain over MIMO HD w/o IC is limited by 2.

Chapter 4

Programmable Control Plane in Tactical Wireless Networking

4.1 Introduction

The control and data planes are two essential elements of any network that moves traffic reliably between source and destination. The control plane contains the signaling and the control rules for the network whereas the the actual data are carried through the data plane. In a classical network, the control and the data plane are coupled tightly. Moreover, these elements are embedded in the hardware of the network devices. Such a classical network removes the network flexibility to adapt to the end-user demands, wireless channel characteristics, and interference amongst the users , i.e., it is not possible to easily manipulate the control rules within the control plane within such a classical network. Fig. 4.1 illustrates the differences in the control plane for the classic vs the SDN networks.

On the other hand, the recent advances in the development of software defined networking (SDN) make a programmable control plane in wired network a reality by separation of the control and data plane [61, 99, 100, 58]. In SDN architecture, the control plane is completely removed

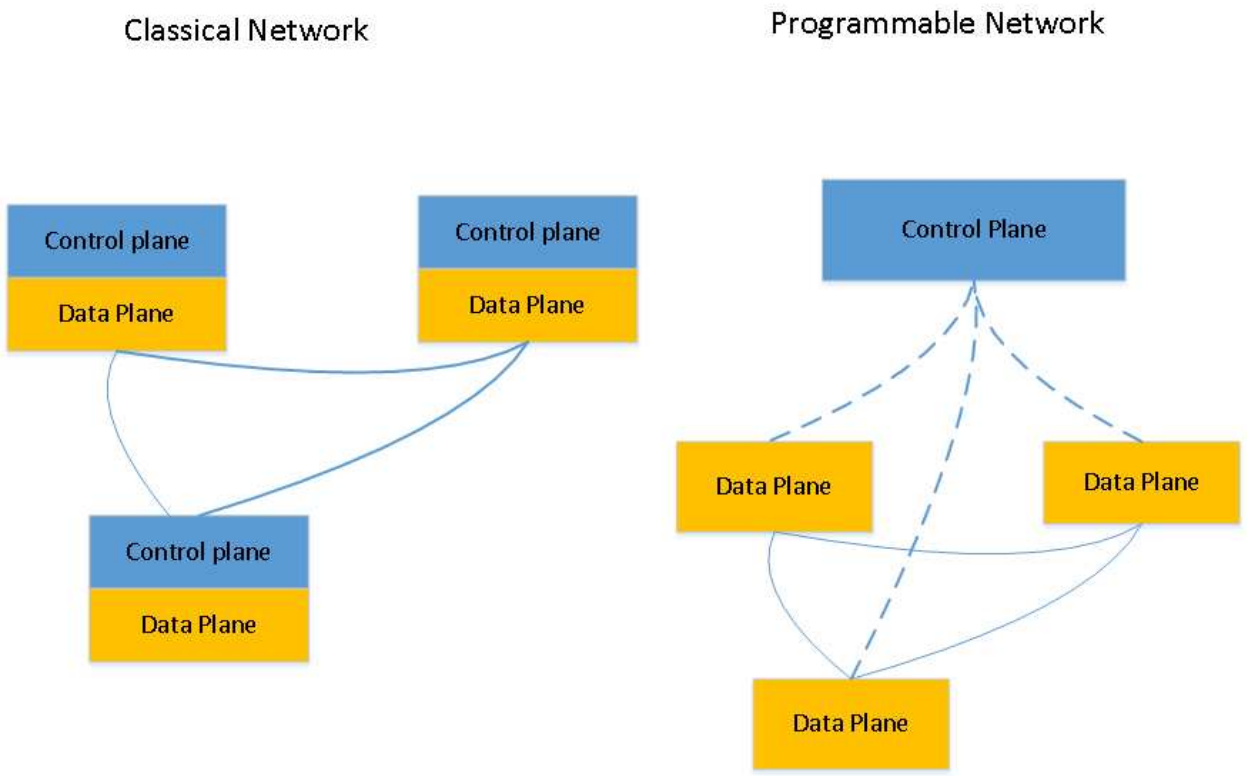


Figure 4.1: Difference in control plane implementation in classical and programmable networks.

from the hardware and externally centralized on a server called controller. The SDN control plane is completely free of proprietary code and will operate vendor independently.

Software defined wireless network (SDWN) recently has been proposed as a natural extension of SDN from wired to wireless network [99, 96, 13]. SDWN could potentially results in wireless networks with more flexible resource management, adaptable wireless network configuration base on the user demands/applications, and vendor-independence wireless network hardwares. SDWN will be different for wireless network from SDN for Internet. In addition to packet forwarding problem, SDWN has to also focus on wireless access and interference management with respect to physical layer design in complex radio environment[13]. The current research on SDWN is mostly limited to suitable network architectures [96, 13, 73].

Realizing a separated control plane for wireless network requires a clear theoretical understanding for modeling the various functionalities of such a control plane, which remains limited in

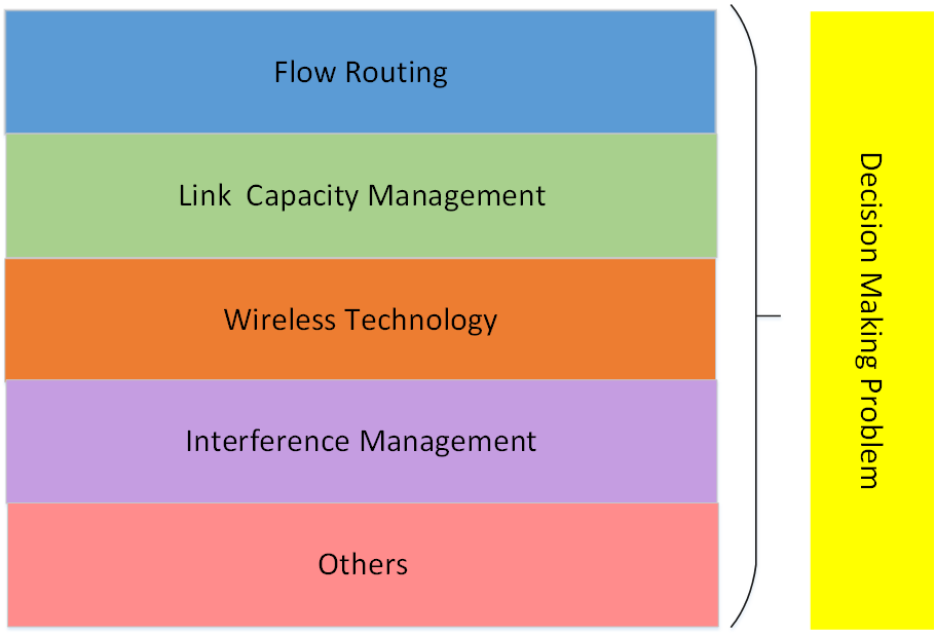


Figure 4.2: Wireless programmable control plane building blocks.

the current literature. In this paper, we will propose the novel idea of the programmable control plane for the tactical wireless network. The programmable control plane comprises of multiple virtual blocks which is in charge various functionalities as shown in Fig. 4.2. In our proposed programmable control plane, the network control layer functionalities can be dynamically configured to adapt to specific physical conditions, customized applications and/or certain tactical situations. Fig. 4.2 shows various building blocks for an envisioned separated programmable control plane in a wireless network. The flow routing block ensures the flow balance constraints hold in every node between source and destination for each session. This block is further responsible to find the optimal routing path for each active session. The link capacity management block manages the amount of flow on each wireless link and ensures it does not exceed the capacity of the link. The wireless technology block manages the signal processing, modulation and coding associated with different wireless technology. The interference management block handles the interference amongst the wireless node in the network. The control plane is programmable as the control layer functionalities in each virtual blocks can be modified or new virtual blocks can be introduced to the layer.

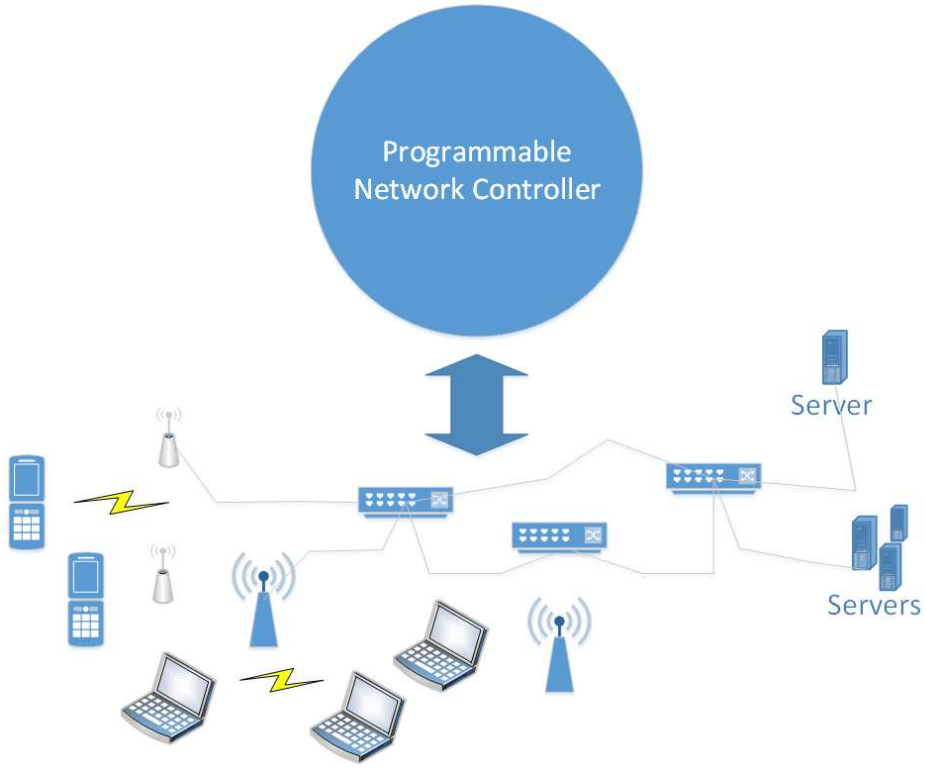


Figure 4.3: Programmable network architecture.

The proposed programmable control plane functionalities can be cast into a centralized optimization problem, which can be updated as needed. Such a programmable control plane decides the optimal value of the each network parameter across various building blocks to achieve the optimal performance for the network. In this work, we develop necessary mathematical model to realize a programmable control plane for wireless networks. We develop a cross-layer optimization framework, which characterizes the interaction between physical, link, and network layer. By applying the framework on a throughput maximization problem, we will show how an envisioned control plane programming can solve complex problems in a tactical networks. Fig. 4.3 shows one possible centralized architecture for implementation of a programmable control plane. The centralized programmable network controller forms the decision making problem with respect to the network objectives and rules in every building block and determines the value of every network variables.

Table 4.1: Notation

Symbol	Definition
N	Number of nodes in the network
$\mathcal{N}$	Set of nodes in the network
$\mathcal{I}_i$	Set of nodes within node i's strong interference range
M	Number of antennas at node each node
G_{ij}	path loss from the transmit node i to the receive node j
$P_{\max}$	Total transmit power of an active transmitter
Q	Total number of transmit power levels
$p_l^q[t]$	Transmit power for the qth data stream at the transmit node of link l in time slot t
$c_l^q[t]$	rate of qth data stream of link l in time slot t
$c_l[t]$	aggregate data rate at link l over all data streams in time slot t
W	bandwidth of the channel
P_n	Noise power
L	Total number of links in the network
$\mathcal{L}$	Set of potential link in the network
$\mathcal{L}_i^{\text{in}}$	Set of incoming links at node i
$\mathcal{L}_i^{\text{out}}$	Set of outgoing links at node i
$A_l^q[t]$	power scaling coefficient of the interference and the noise for the qth data stream of link l
$\text{Rx}(l)$	Receiver of link l
$\text{Tx}(l)$	Transmitter of link l
$r(f)$	Rate of session f
$r_l(f)$	Rate for session f on link l
$x_i[t]$	Indicator variable to show if node i is a transmitter in time slot t
$y_i[t]$	Indicator variable to show if node i is a receiver in time slot t
$z_l[t]$	Number of data streams on link l in time slot t

The remainder of this chapter is organized as following. In Section 4.2, we will propose our general cross-layered model, which captures the most essential rules of the programmable control plane for wireless networks. In Section 4.4, we will provide our numerical results. Section 4.5 concludes this chapter.

4.2 Cross-layered Programmable Control Plane Model

Consider a multi-hop MIMO network consisting of a set of nodes $\mathcal{N}$ which has N elements. Each node is assumed to have M antennas. Suppose that there are L possible links in the network.

Denote $\mathrm{Tx}(l)$ and $\mathrm{Rx}(l)$ as the transmit and receive nodes of link l, $1 \leq l \leq L$. We consider a time-slotted scheduling, where a time frame consists of T time slots. Depending on link scheduling, a subset of links will be active in time slot t, $1 \leq t \leq T$. We assume that there is a set $\mathcal{F}$ of active communication sessions in the network. Each session involves a single-source single-destination unicast flow. Let $r(f)$ be the data rate of session f. Table 4.1 shows the notation used in this section.

Strong Interference vs Weak Interference. We first introduce a concept of "strong" interference range D. For a receive node j, it may be interfered by all the unintended transmit nodes in the network. We distinguish an interference as either a "strong" interference or a "weak" interference through strong interference range D. Specifically, if the distance from a transmit node to its unintended receive node is less than or equal to D, we consider this interference as strong interference; otherwise, we consider it as week interference. In our model, we only avoid strong interference, while weak interference are not avoided. Instead, weak interference will be treated as noise at the receive node when calculating its achievable data rate. We assume $\mathcal{I}_i$ is the set of node strongly interfered by node i. Denote $\mathcal{I}_i$ as the set of nodes that are located within the strong interference range of transmitter i.

Node's SM Constraints. Denote $x_i[t]$ as a binary variable to indicate whether node $i \in \mathcal{N}$ is transmitting in time slot t, i.e., $x_i[t] = 1$ if node i is a transmitter in time slot t and 0 otherwise. Similarly, denote $y_i[t]$ as a binary variable to indicate whether node $i \in \mathcal{N}$ is receiving in time slot t, i.e., $y_i[t] = 1$ if node i is a receiver in time slot t and 0 otherwise. Denote $\mathcal{L}_i^{\mathrm{in}}$ and $\mathcal{L}_i^{\mathrm{out}}$ as the set of potential incoming and outgoing links at node i, respectively. Denote $z_{(l)}[t]$ as the number of data streams over link l. If node i is not a transmitter, then we have $\sum_{l \in \mathcal{L}_i^{\mathrm{out}}} z_l[t] = 0$. Otherwise, the total number of outgoing streams should be positive and lesser than the number of antennas, i.e., $1 \leq \sum_{l \in \mathcal{L}_i^{\mathrm{out}}} z_l[t] \leq M$. These two cases can be expressed in a compact form as follows:

$$x_i[t] \leq \sum_{l \in \mathcal{L}_i^{\mathrm{out}}} z_l[t] \leq M x_i[t], (1 \leq i \leq N, 1 \leq t \leq T) \,. \tag{4.2.1}$$

Similarly, depending on whether node i is an active receiver, we have the following constraint:

$$y_i[t] \leq \sum_{l \in \mathcal{L}_i^{\text{in}}} z_l[t] \leq M y_i[t], (1 \leq i \leq N, 1 \leq t \leq T) \ . \tag{4.2.2}$$

Half Duplex Constraints. Although there has been significant advance on full duplex, it has not yet become part of any wireless standards. Therefore, we assume half duplex on a node in this paper. For half-duplex within the same band, we have the following constraint:

$$x_i[t] + y_i[t] \leq 1, \quad (1 \leq i \leq N, 1 \leq t \leq T) \ . \tag{4.2.3}$$

Protocol Model Constraints. Denote $\lambda_l[t]$ as a binary variable to indicate whether link l is active in time slot t, i.e., $\lambda_l[t] = 1$ if link l is active in time slot t and 0 otherwise. In a protocol model, the transmitter of link k is not supposed to transmit in the interference range of the receiver of link l. The above constraint can be expressed in mathematical form as following.

$$\lambda_l[t] \leq z_l[t] \leq M \lambda_l[t], (1 \leq l \leq L, 1 \leq t \leq T) \ , \tag{4.2.4}$$

$$\lambda_l[t] + \lambda_k[t] \leq 1, (1 \leq l, k \leq L, \text{Tx}(k) \in \mathcal{I}_{\text{Rx}(l)}, \text{Tx}(k) \neq \text{Tx}(l), 1 \leq t \leq T) \ . \tag{4.2.5}$$

Data Stream Based Power Allocation Constraints. For power control, we assume that the transmission power allocated to qth outgoing data stream at transmit node of link l can be tuned to a finite number of levels between 0 and $P_{\max}$. We further assume that the total transmit power of a node is P_s. To model this discrete data stream power control, we introduce an integer parameter Q that represents the total number of power levels to which a transmitter can be adjusted, i.e., transmission power can be $0, \frac{1}{Q}P_{\max}, \frac{2}{Q}P_{\max}, \ldots, P_{\max}$. Denote $p_l^q \in \{0, 1, \ldots, Q\}$ the integer levels for transmission power allocated to qth outgoing data stream at transmit node of link l. If node i is an active transmitter (i.e., $x_i(t) = 1$), then its total transmit power over all data streams is $P_{\max}$. Otherwise (i.e., $x_i(t) = 0$), it transmit power for each data stream is zero. Then we have

$$\sum_{l \in \mathcal{L}_i^{\text{out}}} \sum_{q=1}^{z_l[t]} p_l^q[t] = Q \cdot x_i[t], (1 \leq i \leq N, 1 \leq t \leq T) \ . \tag{4.2.6}$$

Data Stream Capacity Constraints. Denote $c_l^q[t]$ as the achievable rate of qth data stream of link l in time slot t. Denote $\gamma_l^q[t]$ as the effective SINR at the receive node of link l for receiving the kth data streams of link l. For $(1 \leq l \leq L, 1 \leq q \leq z_l[t], 1 \leq t \leq T)$,

$$c_l^q[t] = W \cdot \log_2(1 + \gamma_l^q[t]) \ , \tag{4.2.7}$$

where W is the system bandwidth.

We know calculate the effective SINR $\gamma_l^q[t]$. The DoF IC cancels the strong interfering signals and weak interference signals is treated as noise. Therefore, the effective SINR at the receive node of link l for receiving the qth data streams of link l for $(1 \leq l \leq L, 1 \leq q \leq z_l[t], 1 \leq t \leq T)$ can be written as

$$\gamma_l^q[t] = \frac{G_{\mathrm{Tx}(l)\mathrm{Rx}(l)} \frac{p_l^q[t]}{Q} P_{\max}}{A_l^q[t]\big[\sum\limits_{i \in \mathcal{N} \setminus \mathcal{I}_{\mathrm{Rx}(l)}}^{i \neq \mathrm{Tx}(l)} G_{i\mathrm{Rx}(l)} P_{\max} x_i[t] + P_n\big]} \ , \tag{4.2.8}$$

where G_{ij} is the path loss from the transmit node i to the receive node j; P_n is the noise power at the receiver; $A_l^q[t]$ is a power scaling coefficient of the interference and the noise for the qth data stream of link l, which is determined by the channel matrices. The value of α characterizes the performance of self-interference cancellation, which is consist of the residue of self-talk and cross-talk – the smaller the value of α, the cleaner the cancellation of self interference.

Denote $c_l[t]$ as the aggregate data rate at link l over its $z_l[t]$ data streams in time slot t. Then we have

$$c_l[t] = \sum_{q=1}^{z_l[t]} c_l^q[t], (1 \leq l \leq L, 1 \leq t \leq T) \ . \tag{4.2.9}$$

Link Capacity Constraints. Denote $r_l(f)$ as the amount of data rate on link l that is attributed to session $f \in \mathcal{F}$. Since the aggregate data rate on link l cannot exceed the link's average rate, we have

$$\sum_{f \in \mathcal{F}} r_l(f) \leq \frac{1}{T} \sum_{t=1}^{T} c_l[t], \quad (1 \leq l \leq L) \ , \tag{4.2.10}$$

where the right-hand-side represents the average throughput on link l over a frame (T time slots).

Flow Routing Constraints. Suppose there is a set of active sessions $\mathcal{F}$. Denote $r(f)$ as the rate of session $f \in \mathcal{F}$ and $r_{\min}$ as the minimum session rate, i.e., $r_{\min} = \min_{f\in\mathcal{F}} r(f)$. Denote $s(f)$ and $d(f)$ as the source and destination nodes of session $f \in \mathcal{F}$, respectively. Then at source node, $s(f)$, $f \in \mathcal{F}$, we have the following flow balance:

$$\sum_{l\in\mathcal{L}_i^{\text{out}}} r_l(f) = r(f), \quad (i = s(f), f \in \mathcal{F}) \ . \tag{4.2.11}$$

At any intermediate relay node, we have

$$\sum_{l\in\mathcal{L}_i^{\text{in}}} r_l(f) = \sum_{l\in\mathcal{L}_i^{\text{out}}} r_l(f), (1 \le i \le N, i \ne s(f), i \ne d(f), f \in \mathcal{F}) \ . \tag{4.2.12}$$

At a destination node, we have

$$\sum_{l\in\mathcal{L}_i^{\text{in}}} r_l(f) = r(f), \quad (i = d(f), f \in \mathcal{F}) \ . \tag{4.2.13}$$

4.2.1 Problem Formulation

Denote $r_{\min}$ as the minimum session rate, i.e., $r_{\min} = \min_{f\in\mathcal{F}} r(f)$. Putting all the constraints together, we have the following formulation for the throughput maximization problem:

$$\begin{aligned}
\mathbf{TMP}_1 \quad \max \quad & r_{\min} \\
\text{s.t.} \quad & r_{\min} \le r(f) \qquad (f \in \mathcal{F}); \\
& \text{SM Constraints: (4.2.1),(4.2.2);} \\
& \text{Half Duplex Constraint: (4.2.3);} \\
& \text{Protocol Model Constraints: (4.2.4),(4.2.5);} \\
& \text{Data Stream Power Allocation Constraints: (4.2.6);} \\
& \text{Data Stream Capacity Constraints: (4.2.7)–(4.2.9);} \\
& \text{Link Capacity Constraints: (4.2.10);} \\
& \text{Flow balance constraints: (4.2.11),(4.2.12).}
\end{aligned}$$

In $\mathbf{TMP}_1$, constraints (4.2.6)–(4.2.9) are non-linear. Therefore, $\mathbf{TMP}_1$ is a mixed-integer non-linear program (MINLP), which in general is NP-hard [28]. MINLP problems are known to be

difficult due to the combinatorial nature of mixed-integer programs and the difficulty in solving nonlinear programs. Note that there exist some techniques to address *general* MINLP problems (e.g., outer-approximation methods [23], branch-and-bound [34], extended cutting plane methods [84], and generalized Benders' decomposition [30]). However, these techniques do not exploit our problem-specific structures and properties, and hence can only handle small-sized problems. In this paper, we exploit the unique mathematical structure of our MINLP problem and develop a novel near-optimal solution procedure, with a performance guarantee. We first reformulate the ranges for constraints (4.2.6)–(4.2.9) to only include the constants appear in the range. Next we use reformulation linearizion technique (RLT [70]) to linearize constraint (4.2.8). Similar to the approach proposed by Jiang et al. [42], we adopt a piece-wise linear approximation technique proposed by Jiang et al. [42] to transform the nonlinear constraint (4.2.7) to a set of linear constraints. The main idea of the piece-wise linear approximation is to use a set of linear segments to approximate the log term in (4.2.7) while ensuring that the linear approximation error does not exceed a specific threshold. Since MILP problems are relatively easier to solve than MINLP problems, we can efficiently apply a solver such as CPLEX [98] to obtain a solution efficiently. The solution to this optimization problem gives us the optimal decision on when and how to perform link scheduling for interference management, SM, routing, etc. to ensure the optimal performance for the wireless network.

4.3 Reformulation and Linearization

In $\mathbf{TMP}_1$, the nonlinear constraints are (4.2.6)–(4.2.9). We first reformulate (4.2.6)–(4.2.9) using Reformulation Linearization Technique (RLT) [70] and then linearize the logarithmic function in (4.2.7) via near optimal approximation. We will show that the resulting optimization problem does not involve any nonlinear constraints and provides a near optimal solution to the original problem (i.e., $\mathbf{TMP}_1$).

Reformulation of (4.2.6) and (4.2.9). (4.2.6) and (4.2.9) are non-linear constraints since $z_l[t]$

is not a constant rather an optimization variable. In order to linearize these two constraints, we introduce $M - z_l[t]$ dummy outgoing data streams for node $\mathrm{Tx}(l)$ (transmitter of link l) and add constraints to force the data rate of the dummy data streams to zero. Denote $\lambda_l^q[t]$ as a binary variable to indicate whether or not outgoing data stream q from the transmitter of link l (node $\mathrm{Tx}(l)$) is dummy at time slot t. Specifically, $\lambda_l^q[t] = 0$ if data stream k is dummy and 1 if not. Therefore, we have

$$\sum_{q=1}^{M} \lambda_l^q[t] = z_l[t], \quad (1 \le l \le L, 1 \le t \le T) \; . \tag{4.3.1}$$

Consider the outgoing data streams at $\mathrm{Tx}(l)$ (transmitter of link l). For data stream q, If it is dummy (i.e., $\lambda_l^q[t] = 0$), then the transmit power allocated for this data stream should be 0. Otherwise (i.e., $\lambda_l^q[t] = 1$), the transmit power allocated for this data stream is limited by the total transmit power $P_{\max}$. Therefore, we have

$$0 \le p_l^q[t] \le Q \cdot \lambda_l^q[t], \quad (1 \le l \le L, 1 \le q \le M, 1 \le t \le T) \; . \tag{4.3.2}$$

Similarly, we use the following constraints to force the achievable data rate of a dummy data stream to zero.

$$0 \le c_l^q[t] \le A \cdot \lambda_l^q[t], \quad (1 \le l \le L, 1 \le q \le M, 1 \le t \le T) \; . \tag{4.3.3}$$

where A is a large enough constant.

The new constraints (4.3.1)–(4.3.3) ensure that a dummy data stream has zero transmit power and zero data rate. Therefore, it is equivalent to reformulate (4.2.6) to

$$\sum_{l \in \mathcal{L}_i^{\mathrm{out}}} \sum_{q=1}^{M} p_l^q[t] = Q \cdot x_i[t], \qquad (1 \le i \le N, 1 \le t \le T) \; . \tag{4.3.4}$$

and reformulate (4.2.9) to

$$c_l[t] = \sum_{q=1}^{M} c_l^q[t], \qquad (1 \le l \le L, 1 \le t \le T) \; . \tag{4.3.5}$$

Reformulation of (4.2.8). Based on the new constraints (4.3.1)–(4.3.3), we can rewrite constraint (4.2.8) for $(1 \leq l \leq L, 1 \leq q \leq M, 1 \leq t \leq T)$ as

$$\sum_{i \in \mathcal{N} \setminus \mathcal{I}_{\mathrm{Rx}(l)}}^{i \neq \mathrm{Tx}(l)} G_{i\mathrm{Rx}(l)} P_{\max} x_i[t] \gamma_l^q[t] + P_n \gamma_l^q[t] = \frac{G_{\mathrm{Tx}(l)\mathrm{Rx}(l)}}{A_l^q[t]} \frac{p_l^q[t]}{Q} P_{\max} \ .$$

This constraint is nonlinear due to nonlinear term $x_i[t]\gamma_l^q[t]$. We can use reformulation linearizion technique (RLT [70]) to linearize $x_i[t]\gamma_l^q[t]$. Note since this term is the product of a integer and continuous variable, therefore, the linearizion does not compromise the optimality. We define new variable $\omega_{il}^q[t] = x_i[t]\gamma_l^q[t]$. Then this constraint can be written for $(1 \leq l \leq L, 1 \leq q \leq M, 1 \leq t \leq T)$ as

$$\sum_{i \in \mathcal{N} \setminus \mathcal{I}_{\mathrm{Rx}(l)}}^{i \neq \mathrm{Tx}(l)} G_{i\mathrm{Rx}(l)} P_{\max} \omega_{il}^q[t] + P_n \gamma_l^q[t] = \frac{G_{\mathrm{Tx}(l)\mathrm{Rx}(l)}}{A_l^q[t]} \frac{p_l^q[t]}{Q} P_{\max} \ . \tag{4.3.6}$$

To ensure that $\omega_{il}^q[t] = x_i[t]\gamma_l^q[t]$ always holds, we add the following two linear constraints in the formulation for $(1 \leq l \leq L, i \in \mathcal{N} \setminus \mathcal{I}_{\mathrm{Rx}(l)}, 1 \leq q \leq M, 1 \leq t \leq T)$

$$0 \leq \omega_{il}^q[t] \leq \frac{G_{\mathrm{Tx}(l)\mathrm{Rx}(l)} P_{\max}}{A_l^q[t] P_n} x_i[t] \ , \tag{4.3.7}$$

$$\frac{G_{\mathrm{Tx}(l)\mathrm{Rx}(l)} P_{\max}}{A_l^q[t] P_n} x_i[t] + \gamma_l^q[t] - \frac{G_{\mathrm{Tx}(l)\mathrm{Rx}(l)} P_{\max}}{A_l^q[t] P_n} \leq \omega_{il}^q[t] \leq \gamma_l^q[t] \ . \tag{4.3.8}$$

Reformulation of (4.2.7). Based on (4.3.1)–(4.3.8), we know that the dummy data stream has zero transmit power, zero SINR, and zero data rate. Therefore, it is equivalent to translate (4.2.7) to the following constraint

$$c_l^q[t] = W \cdot \log_2(1 + \gamma_l^q[t]), (1 \leq l \leq L, 1 \leq q \leq M, 1 \leq t \leq T) \ . \tag{4.3.9}$$

In summary, we replace (4.2.6), and (4.2.7)–(4.2.9) with (4.3.1)–(4.3.9). The resulting optimization problem which we denote $\mathbf{BR-TMP}_2$, can be written as

$$
\begin{aligned}
\mathbf{TMP}_2 \quad \max \quad & r_{\min} \\
\text{s.t.} \quad & r_{\min} \le r(f) \qquad (f \in \mathcal{F}); \\
& \text{SM Constraints: (4.2.1),(4.2.2);} \\
& \text{Half Duplex Constraint: (4.2.3);} \\
& \text{Protocol Model Constraint: (4.2.4),(4.2.5);} \\
& \text{Dummy Data Stream Constraints: (4.3.1)–(4.3.3);} \\
& \text{Data Stream Power Allocation Constraints: (4.3.4);} \\
& \text{Data Stream Capacity Constraints: (4.3.5)–(4.3.9);} \\
& \text{Link Capacity Constraints: (4.2.10);} \\
& \text{Flow balance constraints: (4.2.11), (4.2.12);} \\
& \text{Variables: } x_i[t], y_i[t], z_l[t], p_l^q[t], \gamma_l^q[t], c_l^q[t], \lambda_l^q[t], \\
& \omega_{il}^q[t], c_l[t], r_l(f), r(f).
\end{aligned}
$$

where $x_i[t]$, $y_i[t]$, and $\lambda_l^q[t]$ are binary variables; $z_l[t]$ and $p_l^q[t]$ are integer variables; $\gamma_l^q[t]$, $c_l^q[t]$, $\omega_{il}^q[t]$, $c_l[t]$, $r_l(f)$, and $r(f)$ are continuous variables, and all the other parameters are constant. Denote $r^*_{\min}(\mathbf{TMP}_1)$ as the optimal objective value of $\mathbf{TMP}_1$ and $r^*_{\min}(\mathbf{TMP}_2)$ as the optimal objective value of $\mathbf{TMP}_2$. Then we have the following lemma:

Lemma 2. $\mathbf{TMP}_1$ *and* $\mathbf{TMP}_2$ *have the same objective value,i.e.,*

$$r^*_{\min}(\mathbf{TMP}_1) = r^*_{\min}(\mathbf{TMP}_2).$$

The proof is straightforward and we omit its discussion.

4.3.1 Near Optimal Approximation

In this paper, we exploit the unique mathematical structure of our MINLP problem and develop a novel near-optimal solution procedure, with a performance guarantee. Note that in Problem $\mathbf{TMP}_2$, the only nonlinear constraints are the link capacity constraints (4.3.9), which involve the log function. To address this problem, we adopt a piece-wise linear approximation technique

proposed by Jiang et al. [42] to transform the nonlinear constraints to linear constraints. The main idea is as follows. We first use a set of linear segments to approximate the log term in (4.3.9) while ensuring that the linear approximation error does not exceed a specific threshold ϵ. Subsequently, the nonlinear constraints in $\mathbf{TMP}_2$ are replaced by a set of linear constraints. Denote the linearized optimization problem as $\mathbf{TMP}_3$, which is a mixed-integer linear problem (MILP). Since MILP problems are relatively easier to solve than MINLP problems, we can efficiently apply a solver such as CPLEX [98] to obtain a solution efficiently.

We will show that solving $\mathbf{TMP}_3$ gives us a near-optimal solution to the original problem $\mathbf{TMP}_1$. Denote ψ as the desired performance gap for the near-optimal solution, i.e., the difference in the objective values between the optimal solution and the near-optimal solution to OPT. We analyze the relationship between the performance gap ψ and the linear approximation error ϵ. Specifically, for a desired performance gap ψ, we compute the maximum allowed linear approximation error ϵ, and accordingly, we derive the linear approximation constraints and construct $\mathbf{TMP}_3$. The solution to $\mathbf{TMP}_3$ then provides a near-optimal solution with the performance guarantee ψ. We summarize the above steps in Fig. 4.4, and We provide details for the steps in the remainder of this section.

Piece-wise Linear Approximation. We will employ the scheme proposed in [42] by Jiang et al. for piece-wise linear approximation. Denote $\gamma_l^q[t]^{\max} = \frac{G_{\mathrm{Tx}(l)\mathrm{Rx}(l)}P_{\max}}{A_l^q[t]P_n}$ as the maximum effective SINR for the data stream q of link l in time slot t. This scheme introduces a set of consecutive linear segments to approximate $\ln(1+\gamma_l^q[t])$ for $\gamma_l^q[t] \in [0, \gamma_l^q[t]^{\max}]$(see Fig. 4.5). We can rewrite the nonlinear constraints in (4.3.9) as follows:

$$c_l^q[t] = \frac{W}{\ln 2} \cdot \ln(1+\gamma_l^q[t]), (1 \le l \le L, 1 \le q \le M, 1 \le t \le T) \ . \tag{4.3.10}$$

The nonlinear term in (4.3.10) is $\ln(1+\gamma_l^q[t])$. The range of $\gamma_l^q[t]$ is $[0, \gamma_l^q[t]^{\max}]$. Denote ϵ as the maximum allowed error for this linear approximation and let $K_l^q[t]$ be the number of linear segments needed to meet this error requirement. Denote $\gamma_l^q[t]^{(k)}$, $k = 0, 1, \ldots, K_l^q[t]$ as the $\gamma_l^q[t]$-axis values of the endpoints for these $K_l^q[t]$ segments, with $\gamma_l^q[t]^{(0)} \equiv 0$ and $\gamma_l^q[t]^{(K_l^q[t])} \equiv \gamma_l^q[t]^{\max}$.

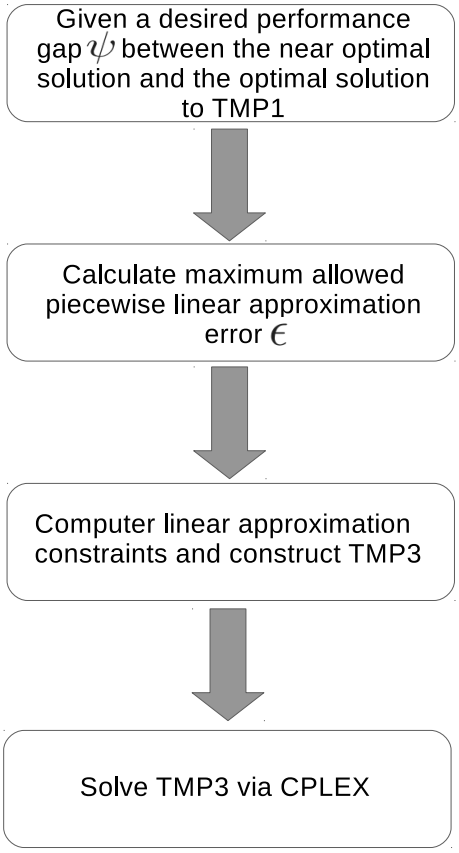


Figure 4.4: A flow chart for obtaining a near optimal solution to TMP1.

Denote $a_l^q[t]^{(k)}$ as the slope of the k-th linear segments, i.e.,

$$a_l^q[t]^{(k)} = \frac{\ln(1+\gamma_l^q[t]^{(k)}) - \ln(1+\gamma_l^q[t]^{(k-1)})}{\gamma_l^q[t]^{(k)} - \gamma_l^q[t]^{(k-1)}} \ . \tag{4.3.11}$$

Denote $g_l^q[t]^{(k)}(\gamma_l^q[t])$ as the k-th linear approximation segment (see Fig. 4.6), which can be represented as follows. For $\gamma_l^q[t]^{(k-1)} \leq \gamma_l^q[t] \leq \gamma_l^q[t]^{(k)}$,

$$g_l^q[t]^{(k)}(\gamma_l^q[t]) = a_l^q[t]^{(k)} \cdot (\gamma_l^q[t] - \gamma_l^q[t]^{(k-1)}) + \ln(1+\gamma_l^q[t]^{(k-1)}). \tag{4.3.12}$$

The values of $\gamma_l^q[t]^{(0)}, \ldots \gamma_l^q[t]^{(K_l^q[t])}$ can by computed sequentially (for a give ϵ) using the algorithm proposed by Jiang et al. in [42]. The algorithm iteratively computes $a_l^q[t]^{(k)}$ that satisfies

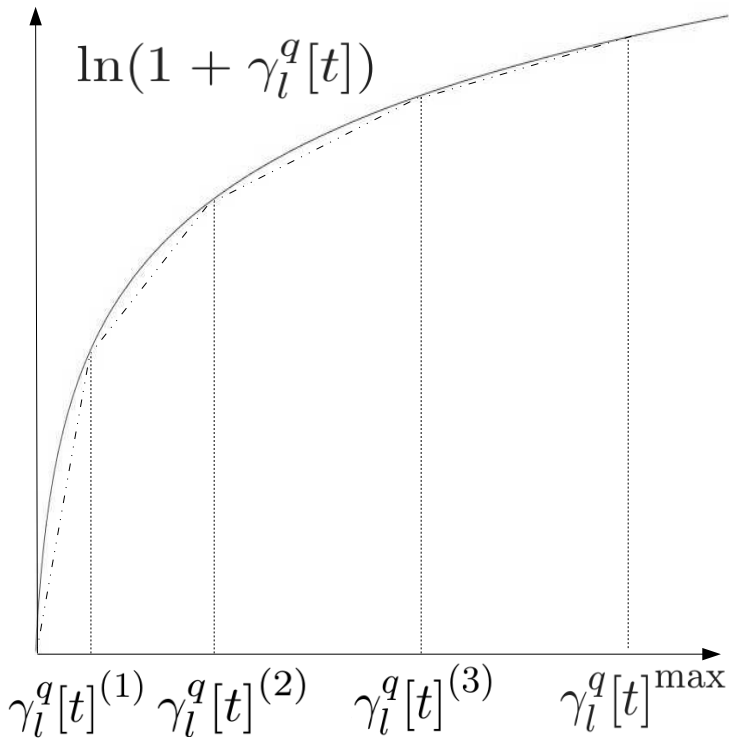


Figure 4.5: An illustration of piece-wise linear approximation with four line segments.

$-\ln(a_l^q[t]^{(k)}) + a_l^q[t]^{(k)}(1+\gamma_l^q[t]^{(k-1)}) - 1 - \ln(1+\gamma_l^q[t]^{(k-1)}) = \epsilon$ and with $a_l^q[t]^{(k)}$, computes $\gamma_l^q[t]^{(k)}$ that satisfies (4.3.11). The proposed algorithm in [42] further guarantees that the maximum approximation error of each linear segment is at most ϵ and *minimizes* the number of linear segments to approximate $\ln(1+\gamma_l^q[t])$, for a given approximation error bound ϵ for each linear segment. After piece-wise linear approximation of $\ln(1+\gamma_l^q[t])$, Constraint (4.3.9) can be replaced by the following set of constraints:

$$c_l^q[t] \leq \frac{W}{\ln 2} g_l^q[t]^{(k)}(\gamma_l^q[t]), (k = 0, 1, \ldots, K_l^q[t], 1 \leq l \leq L, 1 \leq q \leq M, 1 \leq t \leq T) \ ,$$

where $g_l^q[t]^{(k)}(\gamma_l^q[t])$ is given by (4.3.12). Substituting (4.3.12) into the above equation, for $(k = 0, 1, \ldots, K_l^q[t], 1 \leq l \leq L, 1 \leq q \leq M, 1 \leq t \leq T)$ we have

$$c_l^q[t] \leq \frac{W}{\ln 2}[a_l^q[t]^{(k)} \cdot (\gamma_l^q[t] - \gamma_l^q[t]^{(k-1)}) + \ln(1+\gamma_l^q[t]^{(k-1)})] \ . \tag{4.3.13}$$

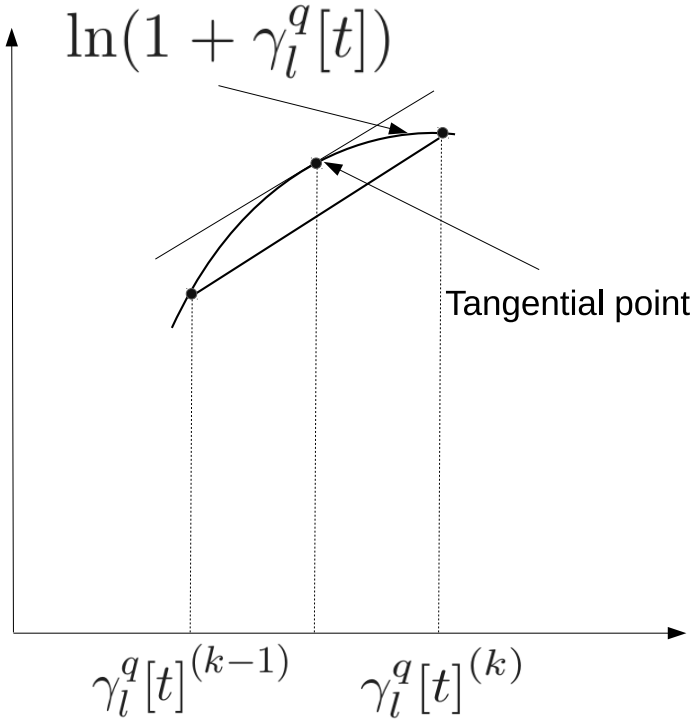


Figure 4.6: An illustration of maximum approximation error for the k-th linear segment.

By replacing the nonlinear constraints in (4.3.9) with the set of linear constraints in (4.3.13), we obtain the following revised formulation for $\mathbf{TMP}_2$, which we denote as $\mathbf{TMP}_3$

$\mathbf{TMP}_3$ max $r_{\min}$

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

SM Constraints: (4.2.1),(4.2.2);

Half Duplex Constraint: (4.2.3);

Protocol Model Constraint: (4.2.4),(4.2.5);

Dummy Data Stream Constraints: (4.3.1)–(4.3.3);

Data Stream Power Allocation Constraints: (4.3.4);

Data Stream Capacity Constraints: (4.3.5)–(4.3.8),(4.3.13);

Link Capacity Constraints: (4.2.10);

Flow balance constraints: (4.2.11), (4.2.12);

Variables: $x_i[t], y_i[t], z_l[t], p_l^q[t], \gamma_l^q[t], c_l^q[t], \lambda_l^q[t]$,
$\omega_{il}^q[t], c_l[t], r_l(f), r(f)$.

Theorem 2. *The gap between the optimal objective values of* $\mathbf{TMP}_3$ *and* $\mathbf{TMP}_1$ *is no more than* $\frac{W \cdot M \cdot L \cdot \epsilon}{\ln 2}$.

Based on Theorem 2, the following algorithm prescribes a near-optimal solution to $\mathbf{TMP}_1$ with a performance guarantee.

Algorithm 4.3.1 An algorithm for obtaining near optimal solution.

Input: Given a desired performance gap psi for the solution to $\mathbf{TMP}_1$.

1. Compute epsilon based on
$$\psi = \frac{W \cdot M \cdot L \cdot \epsilon}{\ln 2} \ . \tag{4.3.14}$$
2. Compute $a_l^q[t]^{(k)}$ and $\gamma_l^q[t]^{(k)}$ by the proposed Algorithm in [42].
3. Construct $\mathbf{TMP}_3$ based on $a_l^q[t]^{(k)}$ and $\gamma_l^q[t]^{(k)}$.
4. Solve $\mathbf{TMP}_3$ optimally using an MILP package (e.g.,CPLEX).

Upon the completion of Algorithm 4.3.1, we have a near-optimal solution to $\mathbf{TMP}_1$ with a guaranteed performance bound (no more than ψ from the optimal objective value).

4.3.2 Near-Optimality Proof

In this section, we present the proof to Theorem 2. In $\mathbf{TMP}_3$, we linearize constraint (4.2.7) in $\mathbf{TMP}_1$ by a piece-wise linear approximation scheme proposed in [42]. In this proof, we provide a bound for the gap between the optimal objective values of $\mathbf{TMP}_1$ and $\mathbf{TMP}_3$.

To better illustrate the idea, we first introduce the following notations. In $\mathbf{TMP}_1$, we assume that there exists a scheduling and power assignment solution $\boldsymbol{\varphi} = (\mathbf{x}[t], \mathbf{y}[t], \mathbf{z}[t], \mathbf{p}[t])$ that satisfies constraints (4.2.1)–(4.2.6), where $\boldsymbol{\pi}[t]$, $\mathbf{x}[t]$, $\mathbf{y}[t]$, and $\mathbf{z}[t]$ represent the vectors $[x_1[t],\ x_2[t], \ldots, x_N[t]]$, $[y_1[t],\ y_2[t], ...,\ y_N[t]]$, and $[z_1[t],\ z_2[t], ...,\ z_L[t]]$; $\mathbf{p}[t]$ represents matrix $[p_l^q[t]]_{L\times Q}$. Then we define $\bar{\boldsymbol{\phi}} = (\boldsymbol{\varphi}, \bar{r}_l(f), \bar{r}(f))$ as a feasible solution to $\mathbf{TMP}_1$, where $(\bar{r}_l(f), \bar{r}(f))$ is the optimal solution to the following linear program $\mathbf{TMP}_1(\boldsymbol{\varphi})$.

$$
\begin{aligned}
\mathbf{TMP}_1(\boldsymbol{\varphi}) \quad \max \quad & r_{\min} \\
\text{s.t.} \quad & r_{\min} \le r(f) && (f \in \mathcal{F}) \\
& \sum_{l\in\mathcal{L}_i^{\text{out}}} r_l(f) = r(f) && (i = s(f), f \in \mathcal{F}) \\
& \sum_{l\in\mathcal{L}_i^{\text{in}}} r_l(f) = \sum_{l\in\mathcal{L}_i^{\text{out}}} r_l(f) && (1\le i\le N, i\ne s(f), i\ne d(f), f \in \mathcal{F}) \\
& \sum_{f\in\mathcal{F}} r_l(f) \le \bar{c}_l && (1 \le l \le L) \\
& r_{\min}, r_l(f), r(f) \ge 0 && (l \in \mathcal{L}, f \in \mathcal{F})
\end{aligned}
$$

where $\bar{c}_l = \frac{1}{T}\sum_{t=1}^{T} \bar{c}_l[t]$. Note that once we fix the scheduling and power variables in $\mathbf{TMP}_1$ to values in $\boldsymbol{\varphi}$, we can get the value of $\bar{c}_l$ by constraints (4.2.7)–(4.2.9). Therefore, given $\boldsymbol{\varphi}$, $\mathbf{TMP}_1(\boldsymbol{\varphi})$ is an LP.

Based on the feasible solution $\bar{\boldsymbol{\phi}} = (\boldsymbol{\varphi}, \bar{r}_l(f), \bar{r}(f))$ to $\mathbf{TMP}_1$, we define a feasible solution $\hat{\boldsymbol{\phi}} = (\boldsymbol{\varphi}, \hat{r}_l(f), \hat{r}(f))$ to $\mathbf{TMP}_3$, where $(\hat{r}_l(f), \hat{r}(f))$ is the optimal solution to $\mathbf{TMP}_3$ when the scheduling and power variables are fixed to values in $\boldsymbol{\varphi}$. That is, $(\hat{r}_l(f), \hat{r}(f))$ is the optimal solution to the following LP.

$$
\begin{aligned}
\mathbf{TMP}_3(\boldsymbol{\varphi}) \quad \max \quad & r_{\min} \\
\text{s.t.} \quad & r_{\min} \le r(f) && (f \in \mathcal{F}) \\
& \sum_{l \in \mathcal{L}_i^{\text{out}}} r_l(f) = r(f) && (i = s(f), f \in \mathcal{F}) \\
& \sum_{l \in \mathcal{L}_i^{\text{in}}} r_l(f) = \sum_{l \in \mathcal{L}_i^{\text{out}}} r_l(f) && (1 \le i \le N, i \ne s(f), i \ne d(f), f \in \mathcal{F}) \\
& \sum_{f \in \mathcal{F}} r_l(f) \le \hat{c}_l && (1 \le l \le L) \\
& r_{\min}, r_l(f), r(f) \ge 0 && (l \in \mathcal{L}, f \in \mathcal{F})
\end{aligned}
$$

where $\hat{c}_l = \frac{1}{T} \sum_{t=1}^{T} \hat{c}_l[t]$. According to constraint (4.3.5), we know that $\hat{c}_l[t] = \sum_{q=1}^{M} \hat{c}_l^q[t]$, where $\hat{c}_l^q[t]$ is the linear approximation of link l's achievable rate of qth data stream in time slot t. Recall that in $\mathbf{TMP}_3$, we use constraint (4.3.13) to replace constraint (4.2.7) in $\mathbf{TMP}_1$. When scheduling variable $\mathbf{x}[t]$ and link l's power assignment $p_l^q[t]$ are fixed at the values given in $\boldsymbol{\varphi}$, we can obtain the value of $\gamma_l^q[t]$ by constraint (4.2.8). Therefore, we can determine which line segment is used in the linear approximation of $\ln(1 + \gamma_l^q[t])$. Suppose that the k-th segment is used, then $\hat{c}_l^q[t]$ can be obtained as $\hat{c}_l^q[t] = \frac{W}{\ln 2} \cdot g_l^q[t]^{(k)}(\gamma_l^q[t])$. As a result, with given $\boldsymbol{\varphi}$, $\mathbf{TMP}_3(\boldsymbol{\varphi})$ is an LP.

Now, we quantify the gap between the optimal solutions to $\mathbf{TMP}_1$ and $\mathbf{TMP}_3$ by the following two steps. First, we show that for any feasible scheduling and power assignment solution $\boldsymbol{\varphi}$, the gap between the feasible objective values corresponding to $\bar{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\phi}}$ is $\frac{W \cdot L \cdot M \cdot \epsilon}{\ln 2}$. Then, we show that the gap between the optimal objective values of $\mathbf{TMP}_1$ and $\mathbf{TMP}_3$ is also bounded by $\frac{W \cdot L \cdot M \cdot \epsilon}{\ln 2}$. This gap is characterized in the dual domain of problem $\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{TMP}_3(\boldsymbol{\varphi})$ as follows.

Step 1: For a given $\boldsymbol{\varphi}$, we denote $\bar{r}$ as the feasible objective value corresponding to solution $\bar{\boldsymbol{\phi}}$ to $\mathbf{TMP}_1$, and $\hat{r}$ as the feasible objective value corresponding to solution $\hat{\boldsymbol{\phi}}$ to $\mathbf{TMP}_3$. In this step, we will show that $\bar{r} - \hat{r} \le \frac{W \cdot L \cdot M \cdot \epsilon}{\ln 2}$.

Denote the dual problem of $\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{TMP}_3(\boldsymbol{\varphi})$ as $\mathbf{D}_1(\boldsymbol{\varphi})$ and $\mathbf{D}_3(\boldsymbol{\varphi})$, respectively. Since the only difference between the formulation of $\mathbf{TMP}_1(\boldsymbol{\varphi})$ and
$\mathbf{TMP}_3(\boldsymbol{\varphi})$ is the constant term $\bar{c}_l$ and $\hat{c}_l$, it is easy to see that $\mathbf{D}_1(\boldsymbol{\varphi})$ and $\mathbf{D}_3(\boldsymbol{\varphi})$ have the same

constraints, but different objective functions.

Denote the dual variables corresponding to the first set of constraints in $\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{TMP}_3(\boldsymbol{\varphi})$ as $w(f), f \in \mathcal{F}$. Denote the dual variables corresponding to the second set of constraints in $\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{TMP}_3(\boldsymbol{\varphi})$ as $v(f), f \in \mathcal{F}$. Denote the dual variables corresponding to the third set of constraints in $\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{TMP}_3(\boldsymbol{\varphi})$ as $y_i(f), f \in \mathcal{F}, i \in \mathcal{N}, i \neq s(f), d(f)$. Denote the dual variables corresponding to the fourth set of constraints in $\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{TMP}_3(\boldsymbol{\varphi})$ as $h_l, l \in \mathcal{L}$. Then $\mathbf{D}_1(\boldsymbol{\varphi})$ can be written as follows:

$$
\begin{array}{lll}
\mathbf{D}_1(\boldsymbol{\varphi}) & \min & \sum_{l\in\mathcal{L}} \bar{c}_l h_l \\
& \text{s.t.} & w(f) \geq 1 \qquad\qquad (f \in \mathcal{F}) \\
& & -w(f) - v(f) \geq 0 \qquad (f \in \mathcal{F}) \\
& & v(f) + y_j(f) + h_l \geq 0 \quad (f \in \mathcal{F}, l \equiv (s(f), j) \in \mathcal{L}, j \neq d(f)) \\
& & y_j(f) - y_i(f) + h_l \geq 0 \\
& & (f \in \mathcal{F}, l \equiv (i, j) \in \mathcal{L}, i \neq s(f), i \neq d(f), j \neq s(f), j \neq d(f)) \\
& & -y_i(f) + h_l \geq 0 \qquad (f \in \mathcal{F}, l \equiv (i, d(f)) \in \mathcal{L}, i \neq s(f)) \\
& & v(f) + h_l \geq 0 \qquad (f \in \mathcal{F}, l \equiv (s(f), d(f)) \in \mathcal{L}) \\
& & w(f), h_l \geq 0, v(f) \text{ and } y_i(f) \text{ unrestricted } (f \in \mathcal{F}, i \in N, i \neq s(f), d(f), l \in \mathcal{L})
\end{array}
$$

The dual problem $\mathbf{D}_3(\boldsymbol{\varphi})$ can be written as follows:

$$
\begin{array}{lll}
\mathbf{D}_3(\boldsymbol{\varphi}) & \min & \sum_{l\in\mathcal{L}} \hat{c}_l h_l \\
& \text{s.t.} & \text{Same constraints as in } \mathbf{D}_1(\boldsymbol{\varphi})
\end{array}
$$

Since $\mathbf{D}_1(\boldsymbol{\varphi})$ and $\mathbf{D}_3(\boldsymbol{\varphi})$ share the same constraints, they have the same feasible region. If h_l^* is (part of) an optimal solution to $\mathbf{D}_3(\boldsymbol{\varphi})$, then since $\bar{r}$ and $\hat{r}$ is the optimal objective value of $\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{TMP}_3(\boldsymbol{\varphi})$ respectively, we have

$$
\bar{r} - \hat{r} \leq \sum_{l\in\mathcal{L}} \bar{c}_l h_l^* - \sum_{l\in\mathcal{L}} \hat{c}_l h_l^* = \sum_{l\in\mathcal{L}} (\bar{c}_l - \hat{c}_l) h_l^* \ .
$$

Now we quantify the gap between $\bar{c}_l$ and $\hat{c}_l$. Since the maximum error of our linear approximation is ϵ, we know that

$$\bar{c}_l^q[t] - \hat{c}_l^q[t] \leq \frac{W}{\ln 2}\epsilon \ .$$

Then based on constraint (4.3.5), we have

$$\bar{c}_l - \hat{c}_l = \frac{1}{T}\sum_{t=1}^{T}\sum_{q=1}^{M}(\bar{c}_l^q[t] - \hat{c}_l^q[t]) \leq \frac{W \cdot M}{\ln 2}\epsilon \ .$$

Therefore,

$$\bar{r} - \hat{r} \leq \frac{W \cdot M}{\ln 2}\epsilon \sum_{l \in \mathcal{L}} h_l^* \ . \tag{4.3.15}$$

According to the definition of marginal rate of change of dual variables in [6], h_l^* is upper bounded by the largest possible change of the optimal objective value $\hat{r}$ of $\mathbf{TMP}_3(\boldsymbol{\varphi})$ with respect to the right-hand side $\hat{c}_l$. Since the objective function is the achievable session rate, and a small marginal $\triangle$-change (say, increase) in the capacity of a link can at most increase $\triangle$ units of session rate, we have

$$h_l^* \leq 1 \ \ (l \in L) \ . \tag{4.3.16}$$

Combining (4.3.15) and (4.3.16), we have

$$\bar{r} - \hat{r} \leq \frac{W \cdot M \cdot L \cdot \epsilon}{\ln 2} \ .$$

This completes Step 1 of the proof.

Step 2: Denote $\boldsymbol{\phi}_1^*$ and r_1^* as the optimal solution and its corresponding optimal objective value of $\mathbf{TMP}_1$. Denote $\boldsymbol{\phi}_3^*$ and r_3^* as the optimal solution and its corresponding optimal objective value of $\mathbf{TMP}_3$. In this step, we show that $r_1^* - r_3^* \leq \frac{W \cdot M \cdot L \cdot \epsilon}{\ln 2}$.

Since the optimal solution $\boldsymbol{\phi}_1^*$ to $\mathbf{TMP}_1$ is a particular case of feasible solution $\bar{\boldsymbol{\phi}}$, we know that there exists a corresponding feasible solution to $\mathbf{TMP}_3$, denoted as $\tilde{\boldsymbol{\phi}}_3$, with a corresponding objective value $\tilde{r}_3$. From Step 1, we know that

$$r_1^* - \tilde{r}_3 \leq \frac{W \cdot M \cdot L \cdot \epsilon}{\ln 2} \ .$$

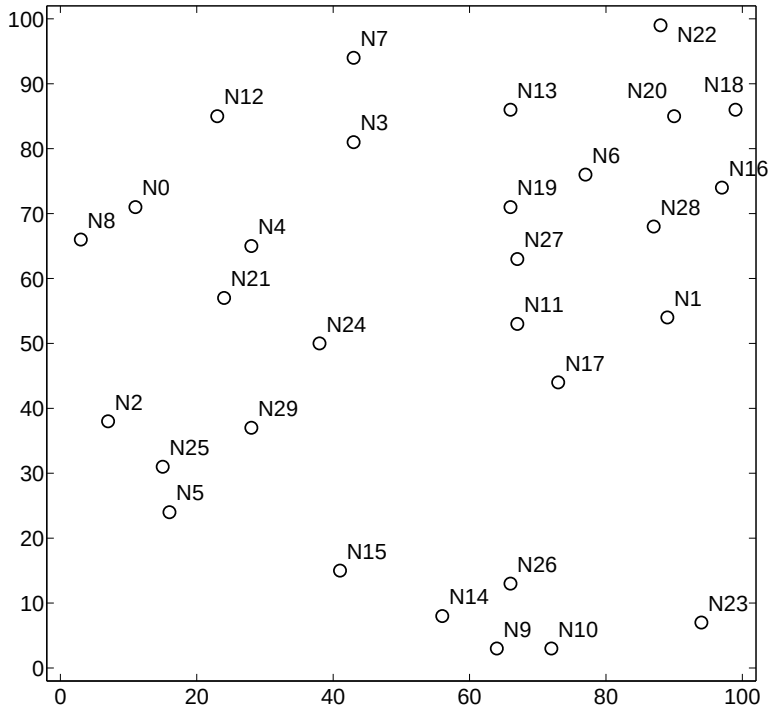


Figure 4.7: A 30-node network in a 100×100 area.

Since r_3^* is the optimal objective value of $\mathbf{TMP}_3$, we know that $r_3^* \geq \tilde{r}_3$. Therefore,

$$r_1^* - r_3^* \leq \frac{W \cdot M \cdot L \cdot \epsilon}{\ln 2} \ .$$

This completes the proof. □

4.4 Numerical Results

We consider a randomly generated multi-hop wireless network with 30 nodes that are distributed in a 100×100 area. For generality, we normalize all units for distance, data rate, bandwidth, and power with appropriate dimensions. At the network layer, minimum-hop routing is employed. The topology of the network is shown in Fig. 4.7. There are 2 active sessions in the network with each session's source node and destination node given in Table 4.2. Each node is equipped with $M = 4$ antennas. We assume the bandwidth $W = 1$. The transmit power for each node is set to 100.

Table 4.2: Source node and destination node in the 30-node network

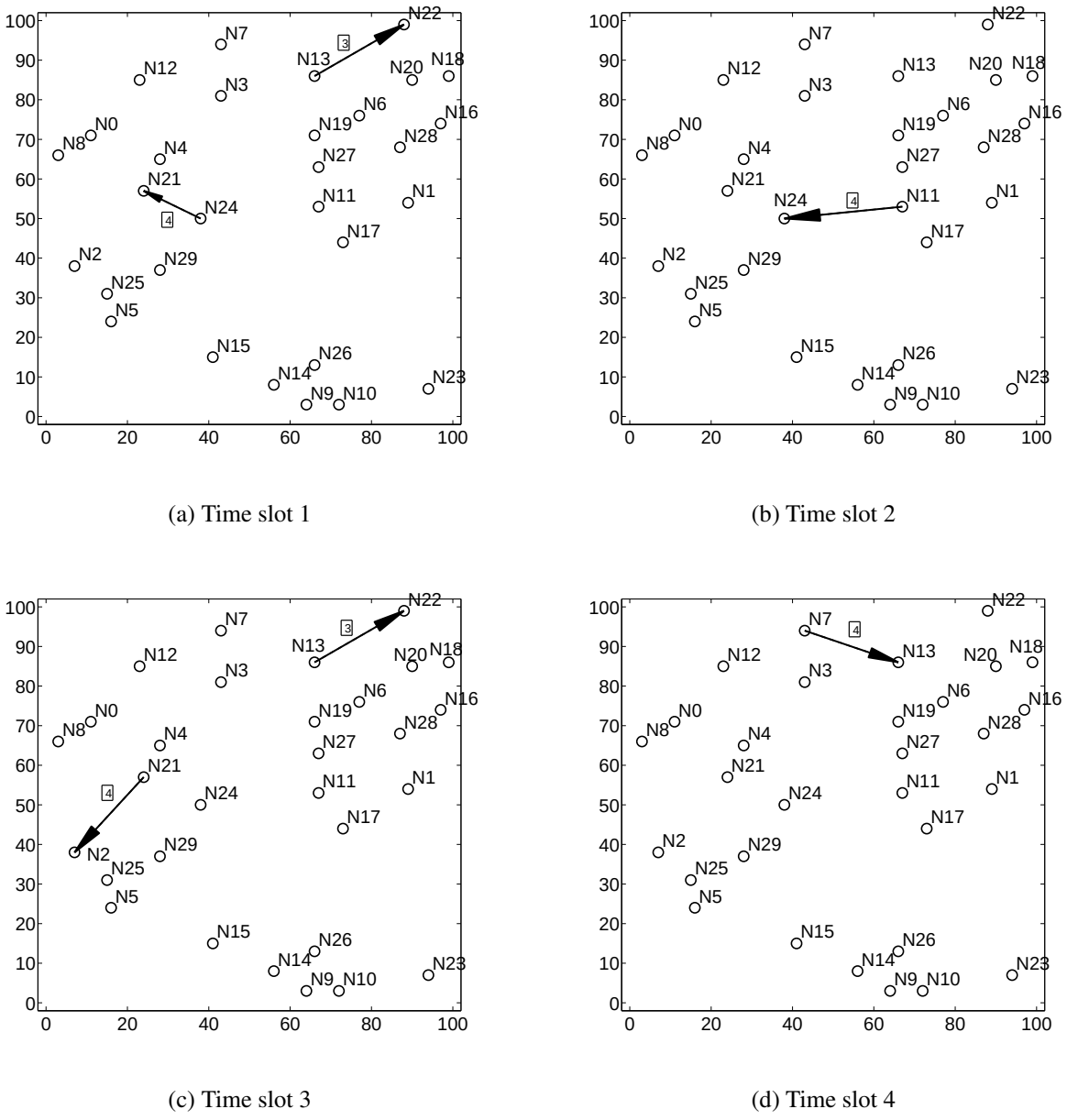
Session f	Source Node $s(f)$	Dest. Node $d(f)$
1	7	22
2	11	2

The path loss parameter $G_{\mathrm{Tx}(k)\mathrm{Rx}(k)} = [20\log_{10}(d) + 38.25]$ (in dB) [101], where d is the distance between $\mathrm{Tx}(k)$ and $\mathrm{Rx}(k)$. The number of time slots in a frame is $T = 4$. The worst case upper bound value for $A_l^q[t]$ is 7.3753.

Figure 4.8 shows the set of active links and the number of data streams per link in each time slot in the solution. Figure 4.9 shows the combined results for all time slots. Table 4.3 shows the details of DoF allocation for SM, aggregated capacity over data streams, and session's rate (i.e., aggregate capacity averaged over a 4-time-slot frame) for all time slots. As shown in Table 4.3, the objective value $r_{\min}$ (minimum session rate) for this network scenario is 0.378504.

4.5 Conclusions

With the recent development in SDN and wide spread availability of high performance computing resources, the idea of centralized separated control plane is not longer a myth. The separation of the data and control plane is a key element to design a highly programmable network. Such a programmable wireless network can address varying end-user demands, manage the interference amongst the users, and modify its physical layer transmission technology with respect to wireless channel characteristic. Realizing a separated control plane for wireless network requires a clear theoretical understanding for modeling the various functionalities of such control plane. In this paper, we proposed the idea of the programmable control plane for the tactical wireless network. In the proposed programmable control plane, the network control layer functionalities can be dynamically configured to adapt to specific physical conditions, customized applications and/or certain



(a) Time slot 1

(b) Time slot 2

(c) Time slot 3

(d) Time slot 4

Figure 4.8: Scheduled links, DoFs allocation on each link, and interference pattern in time slots 1 to 4, respectively.

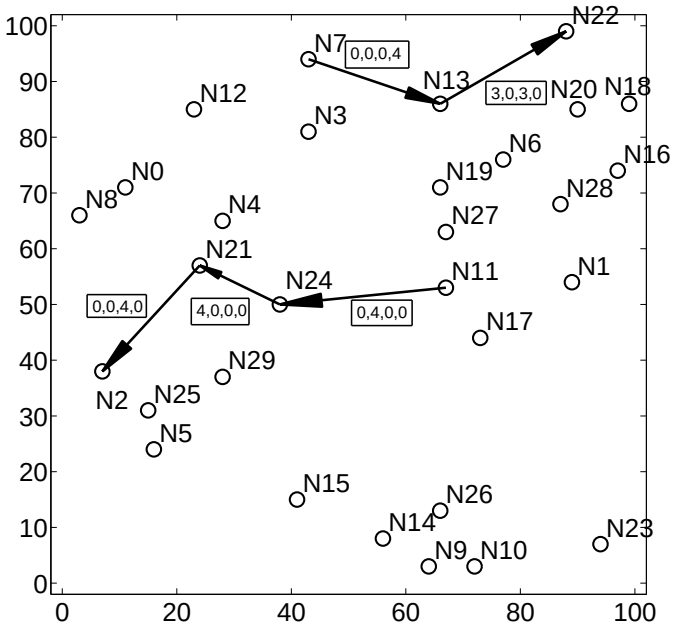


Figure 4.9: Shows the combined results for all time slots (with the number of data streams for each time slot on the link shown in a box).

Table 4.3: DOF allocation for SM, aggregated capacity over data streams, and session's rate for all time slots.

Session	Link	Time Slot	DoF for SM	Capacity	Session's Rate
1	$N_7 \rightarrow N_{13}$	1	0	0	0.378504
		2	0	0	
		3	0	0	
		4	4	11.1578	
	$N_{13} \rightarrow N_{22}$	1	3	1.24433	
		2	0	0	
		3	3	1.44765	
		4	0	0	
2	$N_{11} \rightarrow N_{24}$	1	0	0	0.378504
		2	4	10.6469	
		3	0	0	
		4	0	0	
	$N_{24} \rightarrow N_{21}$	1	4	1.76713	
		2	0	0	
		3	0	0	
		4	0	0	
	$N_{21} \rightarrow N_2$	1	0	0	
		2	0	0	
		3	4	1.51401	
		4	0	0	

tactical situations. The programmable control plane functionalities can be cast into a centralized optimization problem, which can be updated as needed. Specifically, we develop a cross-layer optimization framework, which characterizes the interaction between the physical, link, and network layers. By applying the framework to a throughput maximization problem, we show how the envisioned control plane programming framework can solve complex problems in tactical

Chapter 5

Unified Programmable Control Plane for Heterogeneous Wireless Networks

5.1 Introduction

The recent advances in the development of software defined networking (SDN) make the programmable control plane in the wired network a reality by separating the control and data plane[61, 99, 100, 58]. In SDN architecture, the control plane is completely removed from the hardware and externally centralized on a server called controller. Fig. 5.1 illustrates the differences in the control plane for the traditional and programmable control plane.

Software defined wireless network (SDWN) recently has been proposed as a natural extension of SDN from wired to wireless network[99, 96, 13]. SDWN envisioned to provide flexible resource management, end-user/application aware control plane , and vendor-independent wireless network hardwares. In addition to packet forwarding problem, SDWN has to also focus on wireless access and interference management with respect to physical layer design in complex radio environment[13]. The current research efforts on SDWN[96, 13, 73] mainly focuses on proposing network architecture. To the best of our knowledge, the previous work does not show how to

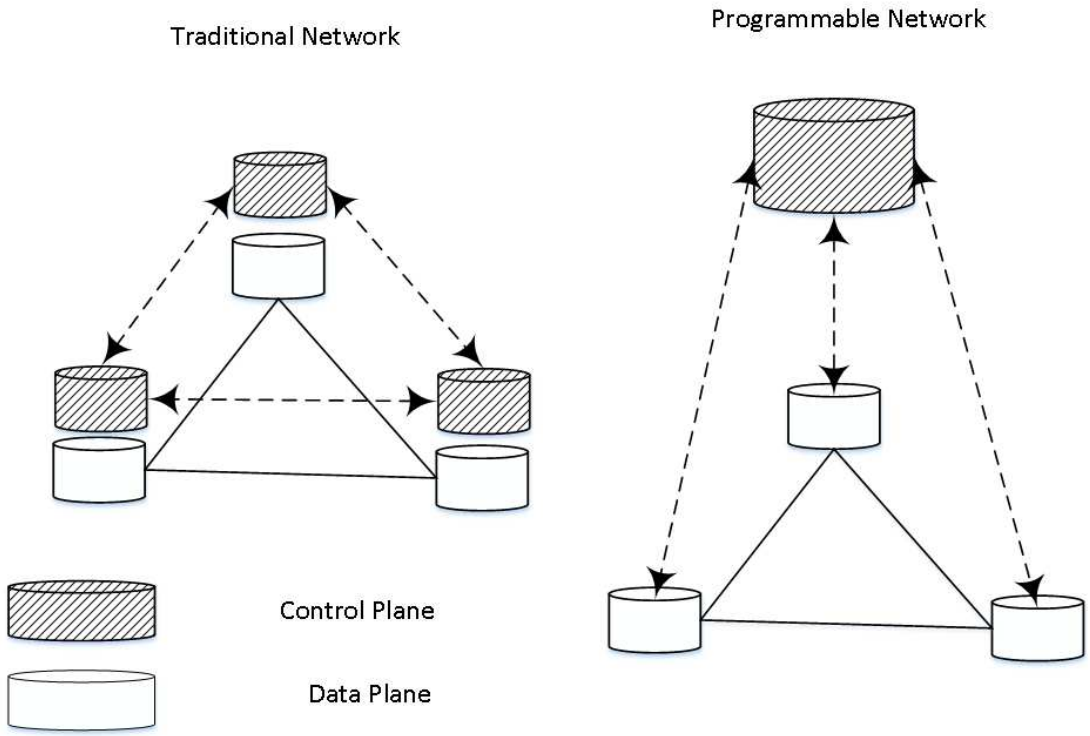


Figure 5.1: Control plane in traditional vs programmable networks.

clearly model and solve the complex functionality of the control plane.

In Fig. 5.2, we demonstrate the logical blocks of the unified programmable control plane for heterogeneous wireless network. The network objective corresponds to the dynamic objective which sought to be achieved by the network. The network layer block ensures the flow balance constraints hold in every node between source and destination for each session. This block is further responsible to find the optimal routing path for each active session. The MAC/link layer has two main functions: i) resolve the medium access contention by link scheduling over different time slots. ii) control the amount of flow on each wireless link to ensure it does not exceed the capacity of the link. The physical layer block manages the signal processing, power management, modulation and coding. This block is also responsible for interference management. The programmable nature of the unified control plane allows the modification of the network objective, constraints within each individual block or even introducing new blocks. The interoperability

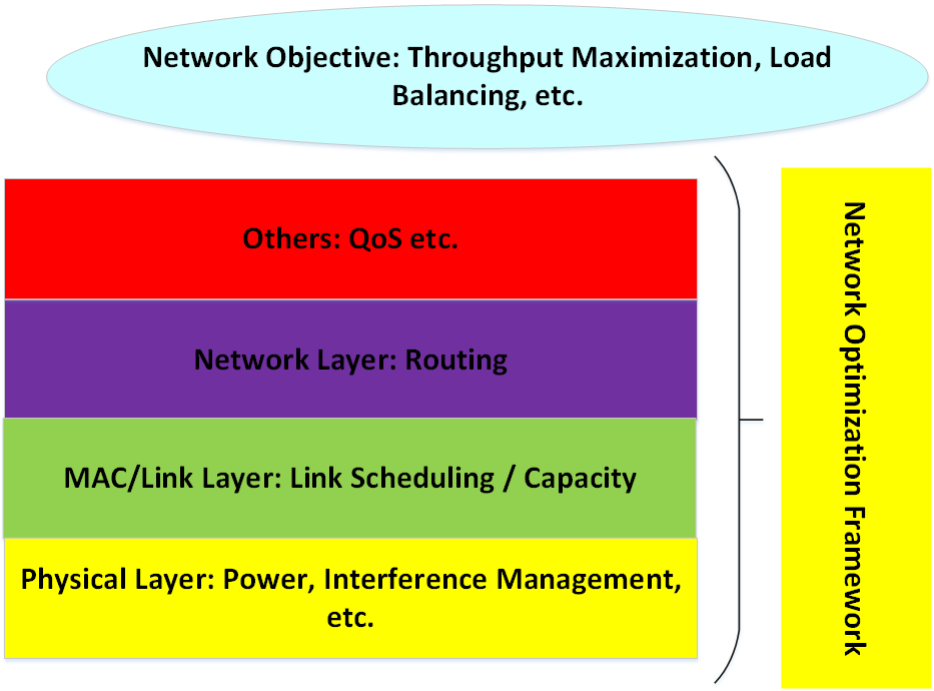


Figure 5.2: Wireless programmable control plane building blocks.

between different technologies for wireless heterogeneous network can be captured by having necessary constraints in these building blocks.

We claim that the functionality of the unified programmable control plane can be modeled as a centralized network optimization problem, in which the optimal network decisions are to be identified and the optimal wireless network resources are to be allocated to achieve the optimal objective for the network at a give time. Fig. 5.3 demonstrates a realization of a network architecture for the unified programmable control plane. The centralized network control center performs the control plane tasks and the data will be carried through the core and wireless network.

In this paper, we develop necessary mathematical model to realize the unified programmable control plane in wireless heterogeneous network. We develop a cross-layer optimization framework, which characterizes the interaction between physical, link, and network layer. By applying the framework on a throughput maximization problem, we will show an application of the model to solve practical issues in a tactical network and gain some theoretical insight on the optimal

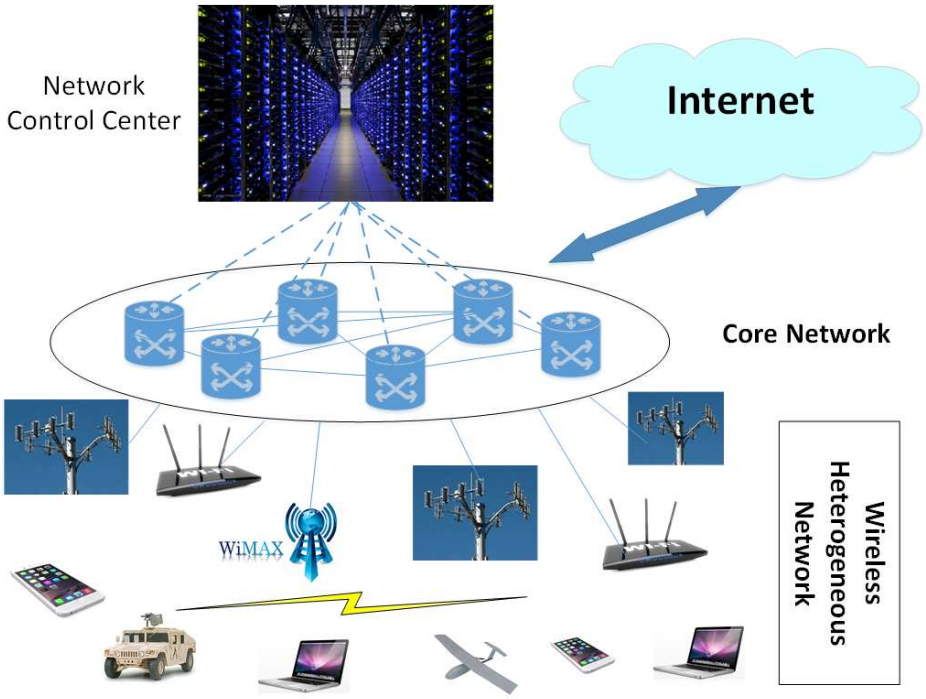


Figure 5.3: Programmable network architecture.

behavior of the unified programmable control plane for a heterogeneous wireless network.

The remainder of this chapter is organized as following. In Section 5.2, we propose necessary mathematical model to realize the unified programmable control plane for heterogeneous wireless networks. In 5.3, we develop a cross-layer optimization framework, which characterizes the interaction between physical, link, and network layer for the unified programmable control plane in a heterogeneous wireless network. In Section 5.4, we will show an application of the model to solve practical issues in a tactical network and gain some theoretical insight on the optimal behavior of the unified programmable control plane for a heterogeneous wireless network. Section 5.5 concludes this chapter.

Table 5.1: Notation

Symbol	Definition
$\mathcal{N}$	Set of nodes in the network
$\mathcal{B}$	set of frequency bands
$\mathcal{I}_i$	Set of nodes within node i's strong interference range
$\mathcal{L}$	Set of potential link in the network
$\mathcal{L}_i^{\text{in}}$	Set of incoming links at node i
$\mathcal{L}_i^{\text{out}}$	Set of outgoing links at node i
M	Number of antennas at node each node
G_{ij}	path loss from the transmit node i to the receive node j
$P_{\max}$	Total transmit power of an active transmitter
Q	Total number of transmit power levels
$p_l^{b,q}[t]$	Transmit power for the kth data stream at the transmit node of link l in time slot t and freq. band b
$c_l^{b,q}[t]$	rate of qth data stream of link l in time slot t and freq. band b
$c_l^b[t]$	aggregate data rate at link l over all data streams in time slot t and freq. band b
W	bandwidth of the channel
P_n	Noise power
$A_l^q[t]$	power scaling coefficient of the interference and the noise for the kth data stream of link l
$\text{Rx}(l)$	Receiver of link l
$\text{Tx}(l)$	Transmitter of link l
$r(f)$	Rate of session f
$r_l(f)$	Rate for session f on link l
$x_i[t]$	Indicator variable to show if node i is a transmitter in time slot t
$y_i[t]$	Indicator variable to show if node i is a receiver in time slot t
$z_l[t]$	Number of data streams on link l in time slot t
$\theta_{ji}[t]$	A binary variable to indicate whether node $i$is placed after node j in $\pi[t]$

5.2 Modeling Unified Control Plane for Wireless Heterogeneous Network

In this section, we will present the functionality of each individual module in the unified control plane for wireless heterogeneous network in mathematical programming. Table 5.1 shows the notation used in this paper.

Consider a multi-hop MIMO network consisting of a set of nodes $\mathcal{N}$ which has N elements. Each node is assumed to have M antennas. Denote $\mathcal{L}$ as the set of all possible links in the network. Denote $\text{Tx}(l)$ and $\text{Rx}(l)$ as the transmit and receive nodes of link l, $l \in \mathcal{L}$. We consider a time-

slotted scheduling, where a time frame consists of T time slots. Depending on link scheduling, a subset of links will be active in time slot t, $1 \leq t \leq T$. We assume each heterogeneous wireless network (e.g. WiFi, LTE, WiMAX, etc.) are operating on orthogonal frequency bands $b \in \mathcal{B}$, where $\mathcal{B}$ is the set of frequency bands.

5.2.1 Mac and Link Layer Constraints

Half-Duplex Constraint. Although there has been significant advance on full duplex for single antenna node, there remain significant challenges to have a practical design for full duplex on a MIMO node. Therefore, we assume half duplex on a MIMO node in this paper. Denote $x_i^b[t]$ as a binary variable to indicate whether node $i \in \mathcal{N}$ is transmitting in time slot t and frequency band b, i.e., $x_i^b[t] = 1$ if node i is a transmitter in time slot t and 0 otherwise. Similarly, denote $y_i^b[t]$ as a binary variable to indicate whether node $i \in \mathcal{N}$ is receiving in time slot t and frequency band b, i.e., $y_i^b[t] = 1$ if node i is a receiver in time slot t and 0 otherwise. For half-duplex, we have the following constraint:

$$x_i^b[t] + y_i^b[t] \leq 1, \qquad (i \in \mathcal{N}, b \in \mathcal{B}, 1 \leq t \leq T) \ . \tag{5.2.1}$$

Node's SM Constraints. Denote $z_l^b[t]$ as the number of data streams over link l at frequency band b. If node i is not a transmitter, then we have $\sum_{l \in \mathcal{L}_i^{\text{out}}} z_l^b[t] = 0$. Otherwise, the total number of outgoing streams should be positive and lesser than the number of antennas, i.e., $1 \leq \sum_{l \in \mathcal{L}_i^{\text{out}}} z_l^b[t] \leq M$. These two cases can be expressed in a compact form as follows:

$$x_i^b[t] \leq \sum_{l \in \mathcal{L}_i^{\text{out}}} z_l^b[t] \leq M x_i^b[t], \quad (i \in \mathcal{N}, b \in \mathcal{B}, 1 \leq t \leq T) \ . \tag{5.2.2}$$

Similarly, depending on whether node i is an active receiver, we have the following constraint:

$$y_i^b[t] \leq \sum_{l \in \mathcal{L}_i^{\text{in}}} z_l^b[t] \leq M y_i^b[t], \quad (i \in \mathcal{N}, b \in \mathcal{B}, 1 \leq t \leq T) \ . \tag{5.2.3}$$

5.2.2 Interference Management

Strong Interference vs Weak Interference. We first introduce a concept of "strong" interference range. For a receive node j, it may be interfered by all the unintended transmit nodes in the network. We distinguish an interference as either a "strong" interference or a "weak" interference through strong interference range. Specifically, if the distance from a transmit node to its unintended receive node is less than or equal to interference range, we consider this interference as strong interference; otherwise, we consider it as week interference. In our model, only strong interference will be considered for IC, while weak interference will not be considered for IC. Instead, weak interference will be treated as noise at the receive node when calculating its achievable data rate. Denote $\mathcal{I}_i$ as the set of nodes that are located within the strong interference range of transmitter i.

Ordering Constraint. In a multi-hop MIMO network, to avoid duplication in IC while ensuring feasibility of DoF allocation, Shi *et al.* [49] introduced a novel IC scheme among the nodes based on a node ordering concept. Under this scheme, all nodes in the network are put into a logical list with the position of the node in the list representing its order. Specifically, denote $\pi^b[t]$ as an ordered list of nodes in the network in time slot t and frequency band b and denote $\pi_i^b[t]$ as the position of node $i \in \mathcal{N}$ in $\pi^b[t]$. Then we have:

$$1 \leq \pi_i^b[t] \leq N, \qquad (i \in \mathcal{N}, b \in \mathcal{B}, 1 \leq t \leq T)\ . \tag{5.2.4}$$

To model the relative ordering between any two nodes i and j in $\pi^b[t]$, we define an indicator variable $\theta_{ji}[t]$ as follows:

$$\theta_{ji}^b[t] = \begin{cases} 1 & \text{if node } j \text{ is before node } i \text{ in } \pi^b[t], \\ 0 & \text{otherwise.} \end{cases}$$

Denote $\mathcal{I}_i$ as the set of nodes that are located within the interference range of transmitter i. Then the ordering relationship between any two nodes in the network can be represented by the following mathematical programming constraints[49]. For $(i \in \mathcal{N}, j \in \mathcal{I}_i, b \in \mathcal{B}, 1 \leq t \leq T)$,

$$\pi_i^b[t] - N \cdot \theta_{ji}^b[t] + 1 \leq \pi_j^b[t] \leq \pi_i^b[t] - N \cdot \theta_{ji}^b[t] + N - 1\ . \tag{5.2.5}$$

Based on $\pi^b[t]$, each node in this list has the following responsibility in IC:

- *Transmit node.* If this node is a transmit node, then it only needs to cancel its interference to those receive nodes that are before itself in the ordered node list. It does not need to consume DoFs to cancel its interference to those receive nodes that are after itself in the ordered node list. Interference from this transmit node to receive nodes after itself will be canceled by those receive nodes later. The number of DoFs consumed at this transmit node for IC is equal to the total number of desired data streams received by those receive nodes.

- *Receive node.* If this node is a receive node, then it only needs to cancel interference from those transmit nodes that are before itself in the ordered node list. It does not need to cancel interference from those transmit nodes that are after itself in the ordered node list. Interference from transmit nodes after this node will be canceled by those transmit nodes later. The number of DoFs consumed at this receive node for IC is equal to the total number of data streams transmitted by those transmit nodes.

The above IC rules can also be cast into mathematical programming constraints. Then the DoF constraint at a transmit node and a receive node for $(i \in \mathcal{N}, b \in \mathcal{B}, 1 \leq t \leq T)$ can be written as follows [49]:

$$\sum_{l \in \mathcal{L}_i^{\text{out}}} z_l^b[t] + \sum_{j \in \mathcal{I}_i} \theta_{ji}^b[t] \sum_{k \in \mathcal{L}_j^{\text{in}}}^{\text{Tx}(k) \neq i} z_k^b[t] \leq M x_i^b[t] + (1 - x_i^b[t]) B_i, \tag{5.2.6}$$

$$\sum_{k \in \mathcal{L}_i^{\text{in}}} z_k^b[t] + \sum_{j \in \mathcal{I}_i} \theta_{ji}^b[t] \sum_{l \in \mathcal{L}_j^{\text{out}}}^{\text{Rx}(l) \neq i} z_l^b[t] \leq M \cdot y_i^b[t] + (1 - y_i^b[t]) B_i, \tag{5.2.7}$$

where B_i is a large constant and is no small than $\sum_{j \in \mathcal{I}_i} \theta_{ji}^b[t] \sum_{k \in \mathcal{L}_j^{\text{in}}}^{\text{Tx}(k) \neq i} z_k^b[t]$ and $\sum_{j \in \mathcal{I}_i} \theta_{ji}^b[t] \sum_{l \in \mathcal{L}_j^{\text{out}}}^{\text{Rx}(l) \neq i} z_l^b[t]$. For example, we can set $B_i = M \cdot |\mathcal{I}_i|$.

5.2.3 Physical Layer Constraints

Power Allocation Constraints. For power control, we assume that the transmission power allocated to qth outgoing data stream at transmit node of link l can be tuned to a finite number of

levels between 0 and $P_{\max}$. We further assume that the total transmit power of a node is P_s. To model this discrete data stream power control, we introduce an integer parameter Q that represents the total number of power levels to which a transmitter can be adjusted, i.e., transmission power can be $0, \frac{1}{Q}P_{\max}, \frac{2}{Q}P_{\max}, \ldots, P_{\max}$. Denote $p_l^{b,q} \in \{0, 1, \ldots, Q\}$ the integer levels for transmission power allocated to qth outgoing data stream at transmit node of link l. If node i is an active transmitter (i.e., $x_i(t) = 1$), then its total transmit power over all data streams is $P_{\max}$. Otherwise (i.e., $x_i(t) = 0$), its transmit power for each data stream is zero. Then we have

$$\sum_{l \in \mathcal{L}_i^{\text{out}}} \sum_{q=1}^{z_l^b[t]} p_l^{b,q}[t] = Q \cdot x_i^b[t], (i \in \mathcal{N}, b \in \mathcal{B}, 1 \leq t \leq T) \ . \tag{5.2.8}$$

SINR Constraints. Denote $c_l^{b,q}[t]$ as the achievable rate of qth data stream of link l in time slot t. Denote $\gamma_l^{b,q}[t]$ as the effective SINR at the receive node of link l for receiving the kth data streams of link l. Then the data stream capacity constraints for $(l \in \mathcal{L}, b \in \mathcal{B}, 1 \leq q \leq z_l^b[t], 1 \leq t \leq T)$ can be written as follows:

$$c_l^{b,q}[t] = W \cdot \log_2(1 + \gamma_l^{b,q}[t]) \ , \tag{5.2.9}$$

where W is the system bandwidth.

We know calculate the effective SINR $\gamma_l^{b,q}[t]$. The DoF IC cancels the strong interfering signals and weak interference signals is treated as noise. Therefore, the effective SINR at the receive node of link l for receiving the qth data streams of link l for $(i \in \mathcal{N}, b \in \mathcal{B}, 1 \leq q \leq z_l^b[t], 1 \leq t \leq T)$ can be written as

$$\gamma_l^{b,q}[t] = \frac{G_{\text{Tx}(l)\text{Rx}(l)} \frac{p_l^{b,q}[t]}{Q} P_{\max}}{A_l^q[t] \big[\sum\limits_{i \in \mathcal{N} \setminus \mathcal{I}_{\text{Rx}(l)}}^{i \neq \text{Tx}(l)} G_{i\text{Rx}(l)} P_{\max} x_i^b[t] + P_n \big]} \ , \tag{5.2.10}$$

where G_{ij} is the path loss from the transmit node i to the receive node j; P_n is the noise power at the receiver; $A_l^q[t]$ is a power scaling coefficient of the interference and the noise for the qth data stream of link l, which is determined by the channel matrices.

Link Capacity Constraints. Denote $c_l^b[t]$ as the aggregate data rate at link l over its $z_l^b[t]$ data

streams in time slot t. Then we have

$$c_l^b[t] = \sum_{q=1}^{z_l^b[t]} c_l^{b,q}[t], \qquad (i \in \mathcal{N}, b \in \mathcal{B}, 1 \le t \le T) \ . \tag{5.2.11}$$

Denote $r_l(f)$ as the amount of data rate on link l that is attributed to session $f \in \mathcal{F}$. Since the aggregate data rate on link l cannot exceed the link's average rate, we have

$$\sum_{f \in \mathcal{F}} r_l(f) \le \frac{1}{T \cdot |\mathcal{B}|} \sum_{t=1}^{T} \sum_{b \in \mathcal{B}} c_l^b[t], \quad (l \in \mathcal{L}). \tag{5.2.12}$$

where the right-hand-side represents the average throughput on link l over a frame (T time slots).

5.2.4 Network Layer Constraints

Flow Routing Constraints. Suppose there is a set of active sessions $\mathcal{F}$. Denote $r(f)$ as the rate of session $f \in \mathcal{F}$ and $r_{\min}$ as the minimum session rate, i.e., $r_{\min} = \min_{f \in \mathcal{F}} r(f)$. Denote $s(f)$ and $d(f)$ as the source and destination nodes of session $f \in \mathcal{F}$, respectively. Then at source node, $s(f)$, $f \in \mathcal{F}$, we have the following flow balance:

$$\sum_{l \in \mathcal{L}_i^{\text{out}}} r_l(f) = r(f), \quad (i = s(f), f \in \mathcal{F}) \ . \tag{5.2.13}$$

At any intermediate relay node, we have

$$\sum_{l \in \mathcal{L}_i^{\text{in}}} r_l(f) = \sum_{l \in \mathcal{L}_i^{\text{out}}} r_l(f), (1 \le i \le N, i \ne s(f), i \ne d(f), f \in \mathcal{F}) \ . \tag{5.2.14}$$

At a destination node, we have

$$\sum_{l \in \mathcal{L}_i^{\text{in}}} r_l(f) = r(f), \quad (i = d(f), f \in \mathcal{F}) \ . \tag{5.2.15}$$

5.3 Problem Formulation and Solution

In this section, we show how the unified heterogeneous wireless network model developed in the previous section can be used to study tactical networking problems. Let's consider a typical

throughput maximization problem in a multi-hop MIMO network. It can be easily verified that if (5.2.13) and (5.2.14) are satisfied, then (5.2.15) is also satisfied. Therefore, it is sufficient to have (5.2.13) and (5.2.14).

Putting all the constraints together, we have the following formulation for the throughput maximization problem:

$$
\begin{aligned}
\mathbf{TMP}_1 \quad \max \quad & r_{\min} \\
\text{s.t.} \quad & r_{\min} \le r(f) \qquad (f \in \mathcal{F}); \\
& \text{Ordering Constraints.:(5.2.4),(5.2.5);} \\
& \text{Half Duplex Constraints.(5.2.1);} \\
& \text{SM Constraints.: (5.2.2),(5.2.3);} \\
& \text{DoF IC Constraints.: (5.2.6),(5.2.7);} \\
& \text{Power Allocation Constraints.: (5.2.8);} \\
& \text{SINR Constraints.: (5.2.9)–(5.2.10);} \\
& \text{Link Capacity Constraints.: (5.2.11),(5.2.12);} \\
& \text{Flow Balance Constraints.: (5.2.13), (5.2.14).}
\end{aligned}
$$

In $\mathbf{TMP}_1$, constraints (5.2.6)–(5.2.11) are non-linear. Therefore, $\mathbf{TMP}_1$ is a mixed-integer nonlinear program (MINLP), which in general is NP-hard [28]. MINLP problems are known to be difficult due to the combinatorial nature of mixed-integer programs and the difficulty in solving nonlinear programs. Note that there exist some techniques to address *general* MINLP problems (e.g., outer-approximation methods [23], branch-and-bound [34], extended cutting plane methods [84], and generalized Benders' decomposition [30]). However, these techniques do not exploit our problem-specific structures and properties, and hence can only handle small-sized problems.

Through reformulation on (5.2.6)–(5.2.11) as well as linearization of the logarithmic function in (5.2.9), $\mathbf{TMP}_1$ can be reformulated into a mixed integer linear program (MILP). Although the theoretical worst-case complexity to a general MILP problem is exponential [28, 68], there exist highly efficient heuristics (e.g., sequential fixing algorithm [38, Chapter 10]) to solve it. Another approach is to apply an off-the-shelf solver (CPLEX [98]), which we found can handle up to a

moderate-sized network successfully. Since the main goal of this paper is to develop necessary mathematical model to to realize the unified programmable control plane for heterogeneous wireless networks, it is sufficient to demonstrate our results with moderate-sized networks. Therefore, we will use CPLEX to solve MILP. The solution to this optimization problem gives us the optimal decision on when and how to perform scheduling, SM, and IC in different time-slots.

5.3.1 Problem Reformulation

In $\mathbf{TMP}_1$, the nonlinear constraints are (5.2.6)–(5.2.11). We first reformulate (5.2.6)–(5.2.11) using Reformulation Linearization Technique (RLT) [70] and then linearize the logarithmic function in (5.2.9) via near optimal approximation. We will show that the resulting optimization problem does not involve any nonlinear constraints and provides a near optimal solution to the original problem (i.e., $\mathbf{TMP}_1$).

Reformulation of (5.2.6) and (5.2.7). We have the nonlinear terms $\theta^b_{ji}[t]\sum_{k\in\mathcal{L}^{\rm in}_j}^{{\rm Tx}(k)\neq i} z^b_k[t]$ and $\theta^b_{ji}[t]\sum_{l\in\mathcal{L}^{\rm out}_j}^{{\rm Rx}(l)\neq i} z^b_l[t]$ in (5.2.6) and (5.2.7). To reformulate these nonlinear terms, we introduce new variables and adding new constraints. Specifically, we define new integer variable $\phi^b_{ji}[t] = \theta^b_{ji}[t]\sum_{k\in\mathcal{L}^{\rm in}_j}^{{\rm Tx}(k)\neq i} z^b_k[t]$. Then (5.2.6) can be rewritten as

$$\sum_{l\in\mathcal{L}^{\rm out}_i} z^b_l[t] + \sum_{j\in\mathcal{I}_i}\phi^b_{ji}[t] \leq M\cdot x^b_i[t] + (1-x^b_i[t])B_i \quad (i\in\mathcal{N}, 1\leq t\leq T)\ , \tag{5.3.1}$$

along with new constraints for $\phi^b_{ji}[t]$ for $(i\in\mathcal{N}, j\in\mathcal{I}_i, 1\leq t\leq T)$.

$$\phi^b_{ji}[t] \leq \sum_{k\in\mathcal{L}^{\rm in}_j}^{{\rm Tx}(k)\neq i} z^b_k[t]\ , \tag{5.3.2}$$

$$\phi^b_{ji}[t] \leq M\cdot\theta^b_{ji}[t]\ , \tag{5.3.3}$$

$$\phi^b_{ji}[t] \geq M\cdot\theta^b_{ji}[t] + \sum_{k\in\mathcal{L}^{\rm in}_j}^{{\rm Tx}(k)\neq i} z^b_k[t] - M\ . \tag{5.3.4}$$

Similarly, for (5.2.7), we define new variable $\xi^b_{ji}[t] = \theta^b_{ji}[t]\sum_{l\in\mathcal{L}^{\rm out}_j}^{{\rm Rx}(l)\neq i} z^b_l[t]$. Then (5.2.7) can be

rewritten as:

$$\sum_{l\in\mathcal{L}_i^{\text{in}}} z_l^b[t] + \sum_{j\in\mathcal{I}_i} \xi_{ji}^b[t] \le M\cdot y_i^b[t] + (1-y_i^b[t])B_i \quad (i\in\mathcal{N}, 1\le t\le T)\ , \tag{5.3.5}$$

along with new constraints for $\xi_{ji}^b[t]$ for $(i\in\mathcal{N}, j\in\mathcal{I}_i, 1\le t\le T)$,

$$\xi_{ji}^b[t] \le \sum_{l\in\mathcal{L}_j^{\text{out}}}^{\text{Rx}(l)\neq i} z_l^b[t]\ , \tag{5.3.6}$$

$$\xi_{ji}^b[t] \le M\cdot\theta_{ji}^b[t]\ , \tag{5.3.7}$$

$$\xi_{ji}^b[t] \ge M\cdot\theta_{ji}^b[t] + \sum_{l\in\mathcal{L}_j^{\text{out}}}^{\text{Rx}(l)\neq i} z_l^b[t] - M\ . \tag{5.3.8}$$

Reformulation of (5.2.8) and (5.2.11). (5.2.8) and (5.2.11) are non-linear constraints since $z_l^b[t]$ is not a constant rather an optimization variable. In order to linearize these two constraints, we introduce $M - z_l^b[t]$ dummy outgoing data streams for node $\text{Tx}(l)$ (transmitter of link l) and add constraints to force the data rate of the dummy data streams to zero. Denote $\lambda_l^{b,q}[t]$ as a binary variable to indicate whether or not outgoing data stream q from the transmitter of link l (node $\text{Tx}(l)$) is dummy at time slot t. Specifically, $\lambda_l^{b,q}[t] = 0$ if data stream k is dummy and 1 if not. Therefore, we have

$$\sum_{q=1}^{M} \lambda_l^{b,q}[t] = z_l^b[t], \quad (l\in\mathcal{L}, 1\le t\le T)\ . \tag{5.3.9}$$

Consider the outgoing data streams at $\text{Tx}(l)$ (transmitter of link l). For data stream q, If it is dummy (i.e., $\lambda_l^{b,q}[t] = 0$), then the transmit power allocated for this data stream should be 0. Otherwise (i.e., $\lambda_l^{b,q}[t] = 1$), the transmit power allocated for this data stream is limited by the total transmit power $P_{\max}$. Therefore, we have

$$0 \le p_l^{b,q}[t] \le Q\cdot\lambda_l^{b,q}[t], \quad (l\in\mathcal{L}, 1\le q\le M, 1\le t\le T)\ . \tag{5.3.10}$$

Similarly, we use the following constraints to force the achievable data rate of a dummy data stream to zero.

$$0 \le c_l^{b,q}[t] \le A\cdot\lambda_l^{b,q}[t], \quad (l\in\mathcal{L}, 1\le q\le M, 1\le t\le T)\ . \tag{5.3.11}$$

where A is a large enough constant.

The new constraints (5.3.9)–(5.3.11) ensure that a dummy data stream has zero transmit power and zero data rate. Therefore, it is equivalent to reformulate (5.2.8) to

$$\sum_{l\in\mathcal{L}_i^{\text{out}}}\sum_{q=1}^{M} p_l^{b,q}[t] = Q\cdot x_i^b[t], \qquad (i\in\mathcal{N}, 1\le t\le T)\ . \tag{5.3.12}$$

and reformulate (5.2.11) to

$$c_l^b[t] = \sum_{q=1}^{M} c_l^{b,q}[t], \qquad (l\in\mathcal{L}, 1\le t\le T)\ . \tag{5.3.13}$$

Reformulation of (5.2.10). Based on the new constraints (5.3.9)–(5.3.11), we can rewrite constraint (5.2.10) for $(l\in\mathcal{L}, 1\le q\le M, 1\le t\le T)$ as

$$\sum_{i\in\mathcal{N}\setminus\mathcal{I}_{\text{Rx}(l)}}^{i\neq\text{Tx}(l)} G_{i\text{Rx}(l)}P_{\max}x_i^b[t]\gamma_l^{b,q}[t] + P_n\gamma_l^{b,q}[t] = \frac{G_{\text{Tx}(l)\text{Rx}(l)}}{A_l^q[t]}\frac{p_l^{b,q}[t]}{Q}P_{\max}\ .$$

This constraint is nonlinear due to nonlinear term $x_i^b[t]\gamma_l^{b,q}[t]$. We can use reformulation linearizion technique (RLT [70]) to linearize $x_i^b[t]\gamma_l^{b,q}[t]$. Note since this term is the product of an integer and a continuous variable, the linearizion guarantees optimality. We define new variable $\omega_{il}^{b,q}[t] = x_i[t]\gamma_l^{b,q}[t]$. Then this constraint can be written for $(l\in\mathcal{L}, 1\le q\le M, 1\le t\le T)$ as

$$\sum_{i\in\mathcal{N}\setminus\mathcal{I}_{\text{Rx}(l)}}^{i\neq\text{Tx}(l)} G_{i\text{Rx}(l)}P_{\max}\omega_{il}^{b,q}[t] + P_n\gamma_l^{b,q}[t] = \frac{G_{\text{Tx}(l)\text{Rx}(l)}}{A_l^q[t]}\frac{p_l^{b,q}[t]}{Q}P_{\max}\ . \tag{5.3.14}$$

To ensure that $\omega_{il}^{b,q}[t] = x_i^b[t]\gamma_l^{b,q}[t]$ always holds, we add the following two linear constraints in the formulation for $(l\in\mathcal{L}, i\in\mathcal{N}\setminus\mathcal{I}_{\text{Rx}(l)}, 1\le q\le M, 1\le t\le T)$

$$0\le\omega_{il}^{b,q}[t]\le\frac{G_{\text{Tx}(l)\text{Rx}(l)}P_{\max}}{A_l^q[t]P_n}x_i^b[t]\ , \tag{5.3.15}$$

$$\frac{G_{\text{Tx}(l)\text{Rx}(l)}P_{\max}}{A_l^q[t]P_n}x_i^b[t] + \gamma_l^{b,q}[t] - \frac{G_{\text{Tx}(l)\text{Rx}(l)}P_{\max}}{A_l^q[t]P_n}\le\omega_{il}^{b,q}[t]\le\gamma_l^{b,q}[t]\ . \tag{5.3.16}$$

Reformulation of (5.2.9). Based on (5.3.9)–(5.3.16), we know that the dummy data stream has zero transmit power, zero SINR, and zero data rate. Therefore, it is equivalent to translate (5.2.9)

to the following constraint

$$c_l^{b,q}[t] = W \cdot \log_2(1+\gamma_l^{b,q}[t]), (l \in \mathcal{L}, 1 \le q \le M, 1 \le t \le T) \ . \tag{5.3.17}$$

In summary, we replace (5.2.6),(5.2.7), (5.2.8), and (5.2.9)–(5.2.11) with (5.3.9)–(5.3.17). The resulting optimization problem which we denote $\mathbf{TMP}_2$, can be written as

$$\begin{aligned}
\mathbf{TMP}_2 \quad \max \quad & r_{\min} \\
\text{s.t.} \quad & r_{\min} \le r(f) \qquad (f \in \mathcal{F}); \\
& \text{Ordering Constraints.:(5.2.4),(5.2.5);} \\
& \text{Half Duplex Constraint: (5.2.1);} \\
& \text{SM Constraints: (5.2.2),(5.2.3);} \\
& \text{DoF IC Constraints: (5.3.1)–(5.3.8);} \\
& \text{Dummy Data Stream Constraints: (5.3.9)–(5.3.11);} \\
& \text{Power Allocation Constraints: (5.3.12);} \\
& \text{SINR Constraints: (5.3.14)–(5.3.17);} \\
& \text{Link Capacity Constraints.: (5.3.13),(5.2.12);} \\
& \text{Flow balance constraints: (5.2.13), (5.2.14);} \\
& \text{Variables: } x_i^b[t], y_i^b[t], z_l^b[t], \pi_i^b[t], \theta_{ji}^b[t], p_l^{b,q}[t], \gamma_l^{b,q}[t], c_l^{b,q}[t], \lambda_l^{b,q}[t], \\
& \phi_{ji}^b[t], \xi_{ji}^b[t], \omega_{il}^{b,q}[t], c_l^b[t], r_l(f), r(f).
\end{aligned}$$

where $x_i^b[t], y_i^b[t], \pi_i^b[t], \theta_{ji}^b[t]$, and $\lambda_l^{b,q}[t]$ are binary variables; $z_l^b[t], p_l^{b,q}[t], \phi_{ji}^b[t]$, and $\xi_{ji}^b[t]$ are integer variables; $\gamma_l^{b,q}[t], c_l^{b,q}[t], \omega_{il}^{b,q}[t], c_l^b[t], r_l(f)$, and $r(f)$ are continuous variables, and all the other parameters are constant. Denote $r_{\min}^*(\mathbf{TMP}_1)$ as the optimal objective value of $\mathbf{TMP}_1$ and $r_{\min}^*(\mathbf{TMP}_2)$ as the optimal objective value of $\mathbf{TMP}_2$. Then we have the following lemma:

Lemma 3. $\mathbf{TMP}_1$ *and* $\mathbf{TMP}_2$ *have the same objective value,i.e.,*

$$r_{\min}^*(\mathbf{TMP}_1) = r_{\min}^*(\mathbf{TMP}_2).$$

The proof is straightforward and we omit its discussion.

5.3.2 Near Optimal Approximation

In this paper, we exploit the unique mathematical structure of our MINLP problem and develop a novel near-optimal solution procedure, with a performance guarantee. Note that in Problem $\mathbf{TMP}_2$, the only nonlinear constraints are the link capacity constraints (5.3.17), which involve the log function. To address this problem, we adopt a piece-wise linear approximation technique proposed by Jiang et al. [42] to transform the nonlinear constraints to linear constraints. The main idea is as follows. We first use a set of linear segments to approximate the log term in (5.3.17) while ensuring that the linear approximation error does not exceed a specific threshold ϵ. Subsequently, the nonlinear constraints in $\mathbf{TMP}_2$ are replaced by a set of linear constraints. Denote the linearized optimization problem as $\mathbf{TMP}_3$, which is a mixed-integer linear problem (MILP). Since MILP problems are relatively easier to solve than MINLP problems, we can efficiently apply a solver such as CPLEX [98] to obtain a solution efficiently.

We will show that solving $\mathbf{TMP}_3$ gives us a near-optimal solution to the original problem $\mathbf{TMP}_1$. Denote ψ as the desired performance gap for the near-optimal solution, i.e., the difference in the objective values between the optimal solution and the near-optimal solution to OPT. We analyze the relationship between the performance gap ψ and the linear approximation error ϵ. Specifically, for a desired performance gap ψ, we compute the maximum allowed linear approximation error ϵ, and accordingly, we derive the linear approximation constraints and construct $\mathbf{TMP}_3$ (see details in Section 5.3.2). The solution to $\mathbf{TMP}_3$ then provides a near-optimal solution with the performance guarantee ψ. We summarize the above steps in Fig. 5.4, and We provide details for the steps in the remainder of this section.

Piece-wise Linear Approximation. We will employ the scheme proposed in [42] by Jiang et al. for piece-wise linear approximation. Denote $\gamma_l^{b,q}[t]^{\max} = \frac{G_{\mathrm{Tx}(l)\mathrm{Rx}(l)}P_{\max}}{A_l^q[t]P_n}$ as the maximum effective SINR for the data stream q of link l in time slot t. This scheme introduces a set of consecutive linear segments to approximate $\ln(1+\gamma_l^{b,q}[t])$ for $\gamma_l^{b,q}[t] \in [0, \gamma_l^{b,q}[t]^{\max}]$(see Fig. 5.5). We can

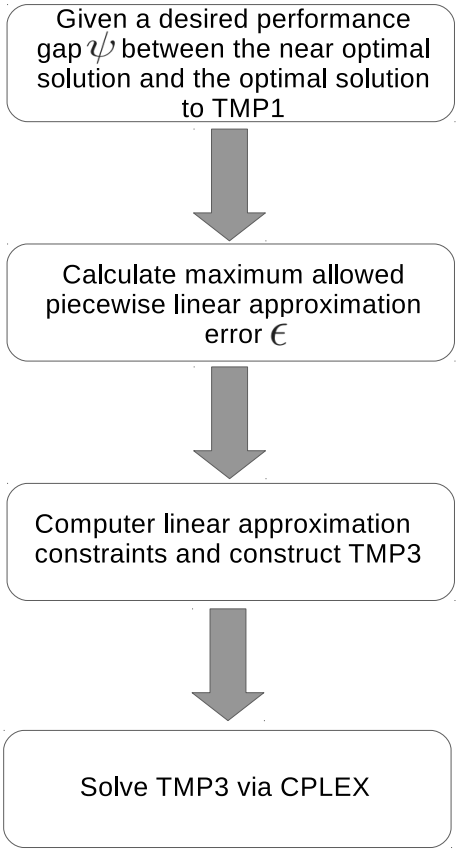


Figure 5.4: A flow chart for obtaining a near optimal solution to TMP1.

rewrite the nonlinear constraints in (5.3.17) as follows:

$$c_l^{b,q}[t] = \frac{W}{\ln 2} \cdot \ln(1 + \gamma_l^{b,q}[t]), (l \in \mathcal{L}, 1 \leq q \leq M, 1 \leq t \leq T) \ . \tag{5.3.18}$$

The nonlinear term in (5.3.18) is $\ln(1 + \gamma_l^{b,q}[t])$. The range of $\gamma_l^{b,q}[t]$ is $[0, \gamma_l^{b,q}[t]^{\max}]$. Denote ϵ as the maximum allowed error for this linear approximation and let $K_l^q[t]$ be the number of linear segments needed to meet this error requirement. Denote $\gamma_l^{b,q}[t]^{(k)}$, $k = 0, 1, \ldots, K_l^q[t]$ as the $\gamma_l^{b,q}[t]$-axis values of the endpoints for these $K_l^q[t]$ segments, with $\gamma_l^{b,q}[t]^{(0)} \equiv 0$ and $\gamma_l^{b,q}[t]^{(K_l^q[t])} \equiv$

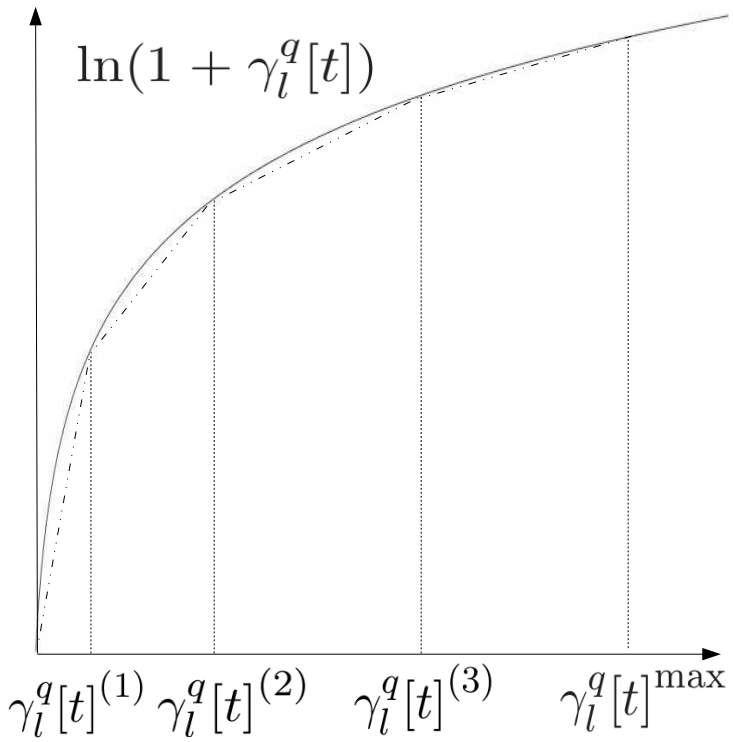


Figure 5.5: An illustration of piece-wise linear approximation with four line segments.

$\gamma_l^{b,q}[t]^{\max}$. Denote $a_l^q[t]^{(k)}$ as the slope of the k-th linear segments, i.e.,

$$a_l^q[t]^{(k)} = \frac{\ln(1+\gamma_l^{b,q}[t]^{(k)}) - \ln(1+\gamma_l^{b,q}[t]^{(k-1)})}{\gamma_l^{b,q}[t]^{(k)} - \gamma_l^{b,q}[t]^{(k-1)}} \ . \tag{5.3.19}$$

Denote $g_l^q[t]^{(k)}(\gamma_l^{b,q}[t])$ as the k-th linear approximation segment (see Fig. 5.6), which can be represented as follows. For $\gamma_l^{b,q}[t]^{(k-1)} \leq \gamma_l^{b,q}[t] \leq \gamma_l^{b,q}[t]^{(k)}$,

$$g_l^q[t]^{(k)}(\gamma_l^{b,q}[t]) = a_l^q[t]^{(k)} \cdot (\gamma_l^{b,q}[t] - \gamma_l^{b,q}[t]^{(k-1)}) + \ln(1+\gamma_l^{b,q}[t]^{(k-1)}). \tag{5.3.20}$$

The values of $\gamma_l^{b,q}[t]^{(0)}, \ldots \gamma_l^{b,q}[t]^{(K_l^q[t])}$ can by computed sequentially (for a give ϵ) using the algorithm proposed by Jiang et al. in [42]. The algorithm iteratively computes $a_l^q[t]^{(k)}$ that satisfies $-\ln(a_l^q[t]^{(k)}) + a_l^q[t]^{(k)}(1+\gamma_l^{b,q}[t]^{(k-1)}) - 1 - \ln(1+\gamma_l^{b,q}[t]^{(k-1)}) = \epsilon$ and with $a_l^q[t]^{(k)}$, computes $\gamma_l^{b,q}[t]^{(k)}$ that satisfies (5.3.19). The proposed algorithm in [42] further guarantees that the maximum approximation error of each linear segment is at most ϵ and *minimizes* the number of

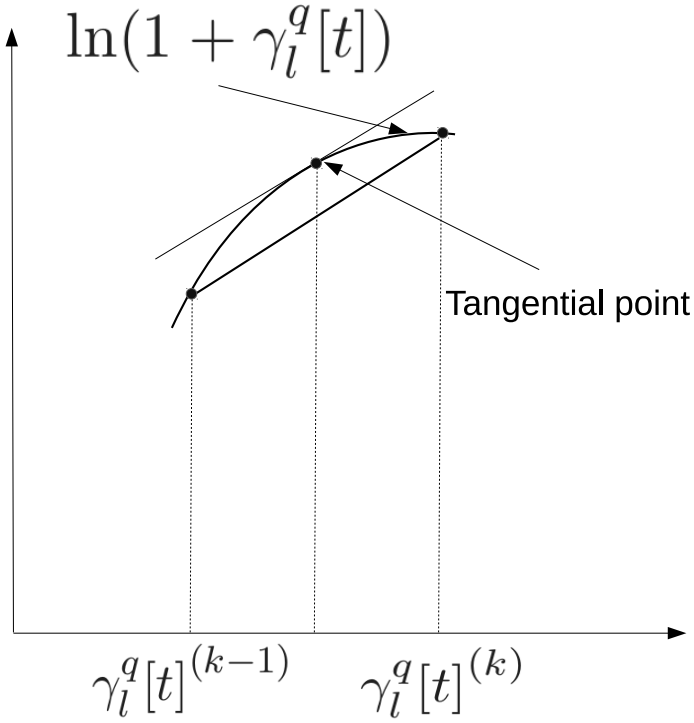


Figure 5.6: An illustration of maximum approximation error for the k-th linear segment.

linear segments to approximate $\ln(1+\gamma_l^{b,q}[t])$, for a given approximation error bound ϵ for each linear segment. After piece-wise linear approximation of $\ln(1+\gamma_l^{b,q}[t])$, Constraint (5.3.17) can be replaced by the following set of constraints:

$$c_l^{b,q}[t] \leq \frac{W}{\ln 2} g_l^q[t]^{(k)}(\gamma_l^{b,q}[t]), (k=0,1,\ldots,K_l^q[t], l \in \mathcal{L}, 1 \leq q \leq M, 1 \leq t \leq T) \ ,$$

where $g_l^q[t]^{(k)}(\gamma_l^{b,q}[t])$ is given by (5.3.20). Substituting (5.3.20) into the above equation, for $(k = 0,1,\ldots,K_l^q[t], l \in \mathcal{L}, 1 \leq q \leq M, 1 \leq t \leq T)$ we have

$$c_l^{b,q}[t] \leq \frac{W}{\ln 2}[a_l^q[t]^{(k)} \cdot (\gamma_l^{b,q}[t] - \gamma_l^{b,q}[t]^{(k-1)}) + \ln(1+\gamma_l^{b,q}[t]^{(k-1)})] \ . \tag{5.3.21}$$

By replacing the nonlinear constraints in (5.3.17) with the set of linear constraints in (5.3.21), we obtain the following revised formulation for $\mathbf{TMP}_2$, which we denote as $\mathbf{TMP}_3$

$\mathbf{TMP}_3$ max $r_{\min}$

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

SM Constraints: (5.2.2),(5.2.2);

Half Duplex Constraint: (5.2.1);

DoF IC Constraints: (5.3.1)–(5.3.8);

Dummy Data Stream Constraints: (5.3.9)–(5.3.11);

Data Stream Power Allocation Constraints: (5.3.12);

SINR Constraints: (5.3.14)–(5.3.16),(5.3.21);

Link Capacity Constraints.: (5.3.13),(5.2.12);

Flow balance constraints: (5.2.13), (5.2.14);

Variables: $x_i^b[t], y_i^b[t], z_l^b[t], \pi_i^b[t], \theta_{ji}^b[t], p_l^{b,q}[t], \gamma_l^{b,q}[t], c_l^{b,q}[t], \lambda_l^{b,q}[t]$, $\phi_{ji}^b[t], \xi_{ji}^b[t], \omega_{il}^{b,q}[t], c_l^b[t], r_l(f), r(f)$.

Theorem 3. *The gap between the optimal objective values of* $\mathbf{TMP}_3$ *and* $\mathbf{TMP}_1$ *is no more than* $\frac{W \cdot M \cdot L \cdot \epsilon}{\ln 2}$.

The proof for the theorem is provided in appendix 5.3.3.

Based on Theorem 3, the following algorithm prescribes a near-optimal solution to $\mathbf{TMP}_1$ with a performance guarantee.

Algorithm 5.3.1 An algorithm for obtaining near optimal solution.

Input: Given a desired performance gap psi for the solution to $\mathbf{TMP}_1$.

1. Compute epsilon based on
$$\psi = \frac{W \cdot M \cdot L \cdot \epsilon}{\ln 2} \,. \tag{5.3.22}$$
2. Compute $a_l^q[t]^{(k)}$ and $\gamma_l^{b,q}[t]^{(k)}$ by the proposed Algorithm in [42].
3. Construct $\mathbf{TMP}_3$ based on $a_l^q[t]^{(k)}$ and $\gamma_l^{b,q}[t]^{(k)}$.
4. Solve $\mathbf{TMP}_3$ optimally using an MILP package (e.g.,CPLEX).

Upon the completion of Algorithm 5.3.1, we have a near-optimal solution to $\mathbf{TMP}_1$ with a guaranteed performance bound (no more than ψ from the optimal objective value).

5.3.3 Near-Optimality Proof

In this section, we present the proof to Theorem 3. In $\mathbf{TMP}_3$, we linearize constraint (5.2.9) in $\mathbf{TMP}_1$ by a piece-wise linear approximation scheme proposed in [42]. In this proof, we provide a bound for the gap between the optimal objective values of $\mathbf{TMP}_1$ and $\mathbf{TMP}_3$.

To better illustrate the idea, we first introduce the following notations. In $\mathbf{TMP}_1$, we assume that there exists a scheduling and power allocation solution $\boldsymbol{\varphi} = (\boldsymbol{\pi}^b[t], \boldsymbol{\theta}[t], \mathbf{x}[t], \mathbf{y}[t], \mathbf{z}[t], \boldsymbol{\eta}[t], \mathbf{p}[t])$ that satisfies constraints (5.2.1)–(5.2.8), where $\boldsymbol{\pi}^b[t]$, $\mathbf{x}[t]$, $\mathbf{y}[t]$, and $\mathbf{z}[t]$ represent the vectors $[\pi_1[t],\ \pi_2[t], ...,\ \pi_N[t]]$, $[x_1[t],\ x_2[t], ...,\ x_N[t]]$, $[y_1[t],\ y_2[t], ...,\ y_N[t]]$, and $[z_1[t],\ z_2[t], ...,\ z_L[t]]$; $\boldsymbol{\theta}[t]$, $\boldsymbol{\eta}[t]$ and $\mathbf{p}[t]$ represent matrices $[\theta^b_{ji}[t]]_{N\times N}$, $[\eta_{ji}[t]]_{N\times N}$ and $[p^q_l[t]]_{L\times Q}$. Then we define $\bar{\boldsymbol{\phi}} = (\boldsymbol{\varphi}, \bar{r}_l(f), \bar{r}(f))$ as a feasible solution to $\mathbf{TMP}_1$, where $(\bar{r}_l(f), \bar{r}(f))$ is the optimal solution to the following linear program $\mathbf{TMP}_1(\boldsymbol{\varphi})$.

$$
\begin{aligned}
\mathbf{TMP}_1(\boldsymbol{\varphi}) \quad \max \quad & r_{\min} & & \\
\text{s.t.} \quad & r_{\min} \le r(f) & & (f \in \mathcal{F}) \\
& \sum_{l\in\mathcal{L}_i^{\text{out}}} r_l(f) = r(f) & & (i = s(f), f \in \mathcal{F}) \\
& \sum_{l\in\mathcal{L}_i^{\text{in}}} r_l(f) = \sum_{l\in\mathcal{L}_i^{\text{out}}} r_l(f) & & (1\le i\le N, i\ne s(f), i\ne d(f), f \in \mathcal{F}) \\
& \sum_{f\in\mathcal{F}} r_l(f) \le \bar{c}_l^b & & (l \in \mathcal{L}) \\
& r_{\min}, r_l(f), r(f) \ge 0 & & (l \in \mathcal{L}, f \in \mathcal{F})
\end{aligned}
$$

where $\bar{c}_l^b = \frac{1}{T\cdot|\mathcal{B}|}\sum_{t=1}^{T}\sum_{b\in\mathcal{B}} \bar{c}_l^b[t]$. Note that once we fix the scheduling and power variables in $\mathbf{TMP}_1$ to values in $\boldsymbol{\varphi}$, we can get the value of $\bar{c}_l^b$ by constraints (5.2.9)–(5.2.11). Therefore, given $\boldsymbol{\varphi}$, $\mathbf{TMP}_1(\boldsymbol{\varphi})$ is an LP.

Based on the feasible solution $\bar{\boldsymbol{\phi}} = (\boldsymbol{\varphi}, \bar{r}_l(f), \bar{r}(f))$ to $\mathbf{TMP}_1$, we define a feasible solution

$\hat{\boldsymbol{\phi}} = (\boldsymbol{\varphi}, \hat{r}_l(f), \hat{r}(f))$ to $\mathbf{TMP}_3$, where $(\hat{r}_l(f), \hat{r}(f))$ is the optimal solution to $\mathbf{TMP}_3$ when the scheduling and power variables are fixed to values in $\boldsymbol{\varphi}$. That is, $(\hat{r}_l(f), \hat{r}(f))$ is the optimal solution to the following LP.

$$
\begin{aligned}
\mathbf{TMP}_3(\boldsymbol{\varphi}) \quad \max \quad & r_{\min} \\
\text{s.t.} \quad & r_{\min} \le r(f) && (f \in \mathcal{F}) \\
& \sum_{l \in \mathcal{L}_i^{\text{out}}} r_l(f) = r(f) && (i = s(f), f \in \mathcal{F}) \\
& \sum_{l \in \mathcal{L}_i^{\text{in}}} r_l(f) = \sum_{l \in \mathcal{L}_i^{\text{out}}} r_l(f) && (1 \le i \le N, i \ne s(f), i \ne d(f), f \in \mathcal{F}) \\
& \sum_{f \in \mathcal{F}} r_l(f) \le \hat{c}_l^b && (l \in \mathcal{L}) \\
& r_{\min}, r_l(f), r(f) \ge 0 && (l \in \mathcal{L}, f \in \mathcal{F})
\end{aligned}
$$

where $\hat{c}_l^b = \frac{1}{T \cdot |\mathcal{B}|} \sum_{t=1}^{T} \sum_{b \in \mathcal{B}} \hat{c}_l^b[t]$. According to constraint (5.3.13), we know that $\hat{c}_l^b[t] = \sum_{q=1}^{M} \hat{c}_l^{b,q}[t]$, where $\hat{c}_l^{b,q}[t]$ is the linear approximation of link l's achievable rate of qth data stream in time slot t. Recall that in $\mathbf{TMP}_3$, we use constraint (3.3.15) to replace constraint (5.2.9) in $\mathbf{TMP}_1$. When scheduling variable $\mathbf{x}[t]$ and link l's power assignment $p_l^{b,q}[t]$ are fixed at the values given in $\boldsymbol{\varphi}$, we can obtain the value of $\gamma_l^{b,q}[t]$ by constraint (5.2.10). Therefore, we can determine which line segment is used in the linear approximation of $\ln(1 + \gamma_l^{b,q}[t])$. Suppose that the k-th segment is used, then $\hat{c}_l^{b,q}[t]$ can be obtained as $\hat{c}_l^{b,q}[t] = \frac{W}{\ln 2} \cdot g_l^q[t]^{(k)}(\gamma_l^{b,q}[t])$. As a result, with given $\boldsymbol{\varphi}$, $\mathbf{TMP}_3(\boldsymbol{\varphi})$ is an LP.

Now, we quantify the gap between the optimal solutions to $\mathbf{TMP}_1$ and $\mathbf{TMP}_3$ by the following two steps. First, we show that for any feasible scheduling and power assignment solution $\boldsymbol{\varphi}$, the gap between the feasible objective values corresponding to $\bar{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\phi}}$ is $\frac{W \cdot L \cdot M \cdot \epsilon}{\ln 2}$. Then, we show that the gap between the optimal objective values of $\mathbf{TMP}_1$ and $\mathbf{TMP}_3$ is also bounded by $\frac{W \cdot L \cdot M \cdot \epsilon}{\ln 2}$. This gap is characterized in the dual domain of problem $\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{TMP}_3(\boldsymbol{\varphi})$ as follows.

Step 1: For a given $\boldsymbol{\varphi}$, we denote $\bar{r}$ as the feasible objective value corresponding to solution $\bar{\boldsymbol{\phi}}$ to $\mathbf{TMP}_1$, and $\hat{r}$ as the feasible objective value corresponding to solution $\hat{\boldsymbol{\phi}}$ to $\mathbf{TMP}_3$. In this

step, we will show that $\bar{r} - \hat{r} \leq \frac{W \cdot L \cdot M \cdot \epsilon}{\ln 2}$.

Denote the dual problem of $\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{TMP}_3(\boldsymbol{\varphi})$ as $\mathbf{D}_1(\boldsymbol{\varphi})$ and $\mathbf{D}_3(\boldsymbol{\varphi})$, respectively. Since the only difference between the formulation of $\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{TMP}_3(\boldsymbol{\varphi})$ is the constant term $\bar{c}_l^b$ and $\hat{c}_l^b$, it is easy to see that $\mathbf{D}_1(\boldsymbol{\varphi})$ and $\mathbf{D}_3(\boldsymbol{\varphi})$ have the same constraints, but different objective functions.

Denote the dual variables corresponding to the first set of constraints in $\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{TMP}_3(\boldsymbol{\varphi})$ as $w(f), f \in \mathcal{F}$. Denote the dual variables corresponding to the second set of constraints in $\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{TMP}_3(\boldsymbol{\varphi})$ as $v(f), f \in \mathcal{F}$. Denote the dual variables corresponding to the third set of constraints in $\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{TMP}_3(\boldsymbol{\varphi})$ as $y_i(f), f \in \mathcal{F}, i \in \mathcal{N}, i \neq s(f), d(f)$. Denote the dual variables corresponding to the fourth set of constraints in $\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{TMP}_3(\boldsymbol{\varphi})$ as $h_l, l \in \mathcal{L}$. Then $\mathbf{D}_1(\boldsymbol{\varphi})$ can be written as follows:

$$
\begin{array}{lll}
\mathbf{D}_1(\boldsymbol{\varphi}) & \min & \sum_{l \in \mathcal{L}} \bar{c}_l^b h_l \\
& \text{s.t.} & w(f) \geq 1 \qquad\qquad (f \in \mathcal{F}) \\
& & -w(f) - v(f) \geq 0 \qquad (f \in \mathcal{F}) \\
& & v(f) + y_j(f) + h_l \geq 0 \quad (f \in \mathcal{F}, l \equiv (s(f), j) \in \mathcal{L}, j \neq d(f)) \\
& & y_j(f) - y_i(f) + h_l \geq 0 \\
& & (f \in \mathcal{F}, l \equiv (i, j) \in \mathcal{L}, i \neq s(f), i \neq d(f), j \neq s(f), j \neq d(f)) \\
& & -y_i(f) + h_l \geq 0 \qquad (f \in \mathcal{F}, l \equiv (i, d(f)) \in \mathcal{L}, i \neq s(f)) \\
& & v(f) + h_l \geq 0 \qquad (f \in \mathcal{F}, l \equiv (s(f), d(f)) \in \mathcal{L}) \\
& & w(f), h_l \geq 0, v(f) \text{ and } y_i(f) \text{ unrestricted } (f \in \mathcal{F}, i \in N, i \neq s(f), d(f), l \in \mathcal{L})
\end{array}
$$

The dual problem $\mathbf{D}_3(\boldsymbol{\varphi})$ can be written as follows:

$$
\begin{array}{lll}
\mathbf{D}_3(\boldsymbol{\varphi}) & \min & \sum_{l \in \mathcal{L}} \hat{c}_l^b h_l \\
& \text{s.t.} & \text{Same constraints as in } \mathbf{D}_1(\boldsymbol{\varphi})
\end{array}
$$

Since $\mathbf{D}_1(\boldsymbol{\varphi})$ and $\mathbf{D}_3(\boldsymbol{\varphi})$ share the same constraints, they have the same feasible region. If h_l^* is (part of) an optimal solution to $\mathbf{D}_3(\boldsymbol{\varphi})$, then since $\bar{r}$ and $\hat{r}$ is the optimal objective value of

$\mathbf{TMP}_1(\boldsymbol{\varphi})$ and $\mathbf{TMP}_3(\boldsymbol{\varphi})$ respectively, we have

$$\bar{r} - \hat{r} \leq \sum_{l \in \mathcal{L}} \bar{c}_l^b h_l^* - \sum_{l \in \mathcal{L}} \hat{c}_l^b h_l^* = \sum_{l \in \mathcal{L}} (\bar{c}_l^b - \hat{c}_l^b) h_l^* \ .$$

Now we quantify the gap between $\bar{c}_l^b$ and $\hat{c}_l^b$. Since the maximum error of our linear approximation is ϵ, we know that

$$\bar{c}_l^{b,q}[t] - \hat{c}_l^{b,q}[t] \leq \frac{W}{\ln 2}\epsilon \ .$$

Then based on constraint (5.3.13), we have

$$\bar{c}_l^b - \hat{c}_l^b = \frac{1}{T \cdot |\mathcal{B}|} \sum_{t=1}^{T} \sum_{b \in \mathcal{B}} \sum_{q=1}^{M} (\bar{c}_l^{b,q}[t] - \hat{c}_l^{b,q}[t]) \leq \frac{W \cdot M}{\ln 2}\epsilon \ .$$

Therefore,

$$\bar{r} - \hat{r} \leq \frac{W \cdot M}{\ln 2}\epsilon \sum_{l \in \mathcal{L}} h_l^* \ . \tag{5.3.23}$$

According to the definition of marginal rate of change of dual variables in [6], h_l^* is upper bounded by the largest possible change of the optimal objective value $\hat{r}$ of $\mathbf{TMP}_3(\boldsymbol{\varphi})$ with respect to the right-hand side $\hat{c}_l^b$. Since the objective function is the achievable session rate, and a small marginal $\triangle$-change (say, increase) in the capacity of a link can at most increase $\triangle$ units of session rate, we have

$$h_l^* \leq 1 \ \ (l \in L) \ . \tag{5.3.24}$$

Combining (5.3.23) and (5.3.24), we have

$$\bar{r} - \hat{r} \leq \frac{W \cdot M \cdot L \cdot \epsilon}{\ln 2} \ .$$

This completes Step 1 of the proof.

Step 2: Denote $\boldsymbol{\phi}_1^*$ and r_1^* as the optimal solution and its corresponding optimal objective value of $\mathbf{TMP}_1$. Denote $\boldsymbol{\phi}_3^*$ and r_3^* as the optimal solution and its corresponding optimal objective value of $\mathbf{TMP}_3$. In this step, we show that $r_1^* - r_3^* \leq \frac{W \cdot M \cdot L \cdot \epsilon}{\ln 2}$.

Since the optimal solution $\boldsymbol{\phi}_1^*$ to $\mathbf{TMP}_1$ is a particular case of feasible solution $\bar{\boldsymbol{\phi}}$, we know that there exists a corresponding feasible solution to $\mathbf{TMP}_3$, denoted as $\tilde{\boldsymbol{\phi}}_3$, with a corresponding

objective value $\tilde{r}_3$. From Step 1, we know that

$$r_1^* - \tilde{r}_3 \leq \frac{W \cdot M \cdot L \cdot \epsilon}{\ln 2} .$$

Since r_3^* is the optimal objective value of $\mathbf{TMP}_3$, we know that $r_3^* \geq \tilde{r}_3$. Therefore,

$$r_1^* - r_3^* \leq \frac{W \cdot M \cdot L \cdot \epsilon}{\ln 2} .$$

This completes the proof. □

5.4 Numerical Results

In this section, we present some numerical results to study the performance of the unified control plane for wireless heterogeneous network described in Section 5.2. The goal of this effort is twofold. First, we want to show how a solution to the $\mathbf{TMP}_1$ formulation looks like for an example network. By studying the details of our solution for an example network, we will develop some quantitative understanding on how the intelligent network decision made in the unified control plane can significantly improve the network throughput. Second, we will show how the intelligent centralized interference management scheme implemented in the unified control plane for the wireless heterogeneous network can cancel the interference and improves the network throughput compared to the case without such an intelligent interference management in the unified control plane.

5.4.1 Simulation Settings

We consider a randomly generated multi-hop wireless network with 30 nodes that are distributed in a 100×100 area. For generality, we normalize all units for distance, data rate, bandwidth, and power with appropriate dimensions. At the network layer, minimum-hop routing is employed. The topology of the network is shown in Fig. 5.7. There are 2 active sessions in the network with each session's source node and destination node given in Table 5.2. Each node is equipped with

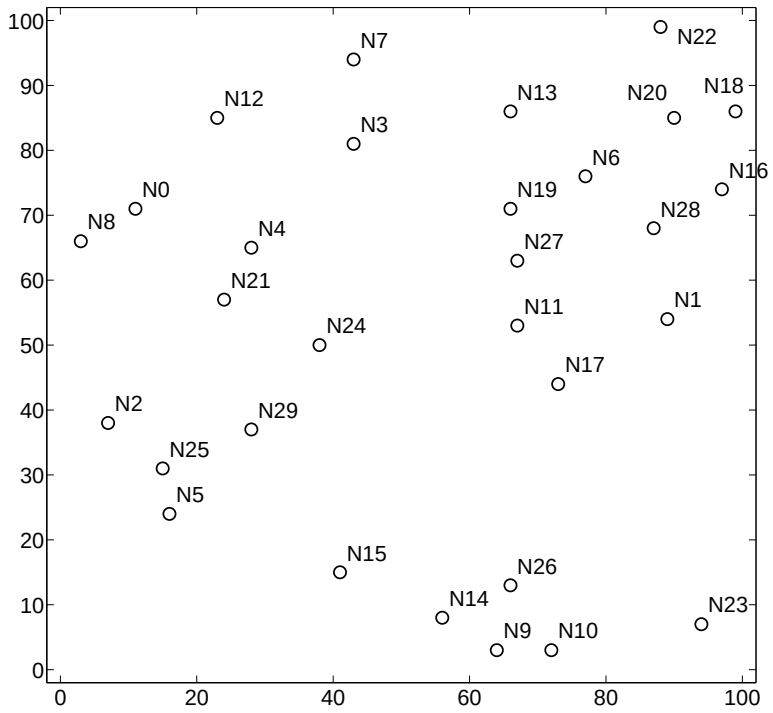


Figure 5.7: A 30-node network in a 100×100 area.

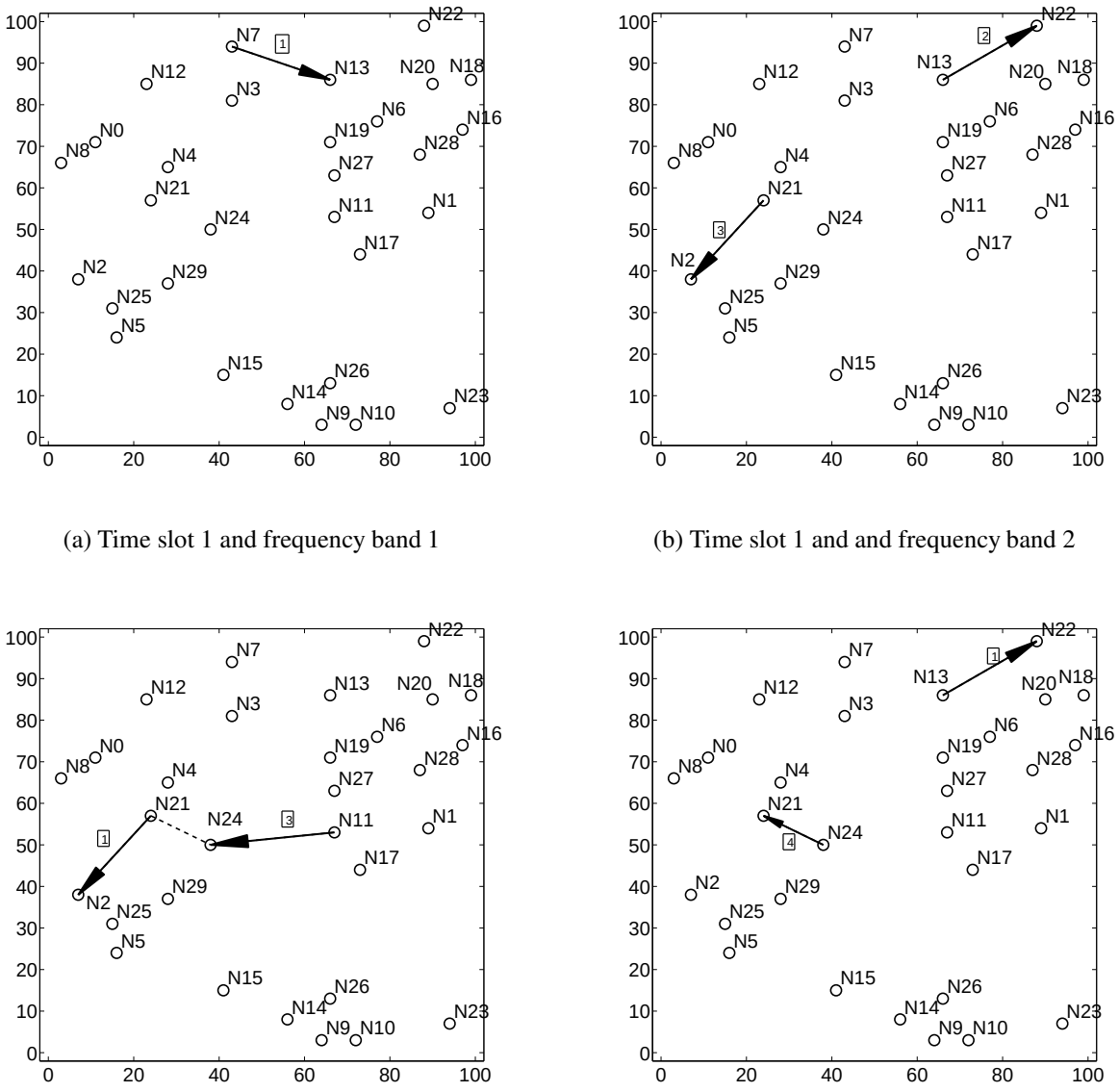
Table 5.2: Source node and destination node in the 30-node network.

Session f	Source Node $s(f)$	Dest. Node $d(f)$
1	7	22
2	11	2

Table 5.3: DOF allocation for SM, aggregated capacity over data streams, and session's rate for all time slots and frequency bands.

Session	Link	(Time Slot,Freq. Band)	DoF for SM	Capacity	Session's Rate
1	$N_7 \to N_{13}$	(1,1)	1	11.1414	0.441784
		(1,2)	0	0	
		(2,1)	0	0	
		(2,2)	0	0	
	$N_{13} \to N_{22}$	1(1,1)	0	0	
		(1,2)	2	1.15608	
		(2,1)	0	0	
		(2,2)	1	1.01216	
2	$N_{11} \to N_{24}$	(1,1)	0	0	0.441784
		(1,2)	0	0	
		(2,1)	3	10.6387	
		(2,2)	0	0	
	$N_{24} \to N_{21}$	(1,1)	0	0	
		(1,2)	0	0	
		(2,1)	0	0	
		(2,2)	4	1.76713	
	$N_{21} \to N_2$	(1,1)	0	0	
		(1,2)	3	1.1577	
		(2,1)	1	0.845915	
		(2,2)	0	0	

$M = 4$ antennas. We assume the bandwidth $W = 1$. The transmit power for each node is set to 100. The path loss parameter $G_{\mathrm{Tx}(k)\mathrm{Rx}(k)} = [20 \log_{10}(d) + 38.25]$ (in dB) [101], where d is the distance between $\mathrm{Tx}(k)$ and $\mathrm{Rx}(k)$. The number of time slots in a frame is $T = 2$. The number of frequency bands for different heterogeneous wireless networks is $|\mathcal{B}| = 2$. The worst case upper bound value for $A_l^q[t]$ is 7.3753. We set the maximum acceptable performance gap between the optimal objective values of Figure 5.8 shows the set of active links and the number of data streams per link in each time slot and each frequency band in the solution. Figure 5.9 shows the combined results for all time slots and frequency bands. Table 5.3 shows the details of DoF allocation for SM, aggregated capacity over data streams, and session's rate for all time slots and and each frequency bands for MIMO half duplex. Table 5.4 shows the details of IC for time slot $T = 2, b = 1)$. The IC within the network follows the node ordering, which is shown in the third column of Table 5.4.



(a) Time slot 1 and frequency band 1

(b) Time slot 1 and and frequency band 2

(c) Time slot 2 and and frequency band 1

(d) Time slot 2 and and frequency band 2

Figure 5.8: Scheduled links, DoFs allocation on each link, and interference pattern across different time slots an frequency bands

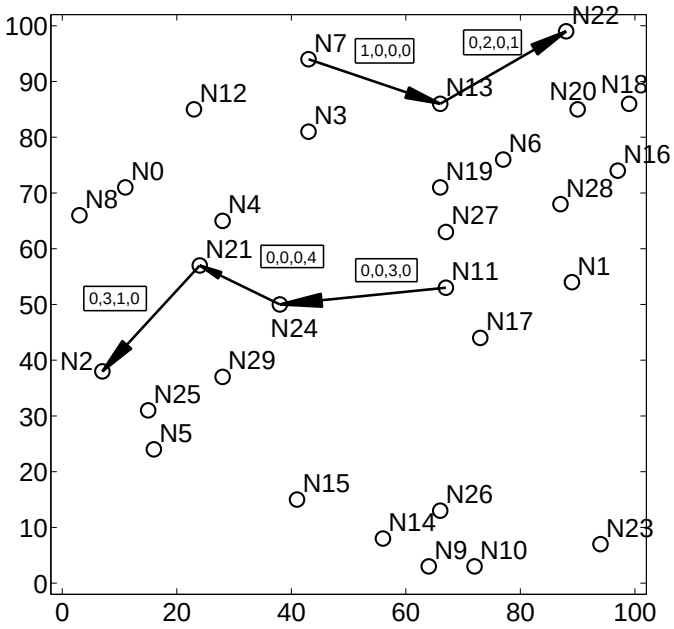


Figure 5.9: Shows the combined results for all time slots and frequency bands (with the number of data streams for each time slot and frequency band on the link shown in a box).

Table 5.4: DoF allocation for SM/IC for time slot $T = 2$ and frequency band $b = 1$.

Node i	TX/RX	$\pi_i[1]$	DoF for SM	DoF for IC to/from
N_{11}	TX	2	3	0
N_{24}	RX	1	3	0
N_{21}	TX	7	1	3 to N_{24}
N_2	RX	3	1	3 to N_1

The number of DoFs used for IC to/from other nodes is shown in the column of Table 5.4. As shown in Table 5.3, the objective value $r_{\min}$ (minimum session rate) for this network scenario is 0.483653.

5.5 Conclusions

The unified programmable control plane for wireless heterogeneous network abstracts the complexity of heterogeneous wireless network and provides a centralized control over the network resources. The abstraction of the complexity enables the interoperability between different technologies in tactical wireless networks. Moreover such a unified programmable control plane enables advanced routing, MAC/Link/Phy layer functionalities, and advanced interference management scheme. It is also providing a dynamic structure to re-configure the network to achieve various network objectives over time. We developed the necessary mathematical model to realize a unified programmable control plane for wireless heterogeneous network. Our proposed network optimization framework characterizes the interaction between physical, link, and network layer. The unified programmable control plane can dynamically solve the optimization problem to decide the optimal values for network decision variables. By applying the framework on a throughput maximization problem, we evaluate our model and show the proposed unified programmable control plane for wireless heterogeneous network can solve complex problems in tactical wireless networks.

Chapter 6

Dissertation Summary and Future Work

6.1 Dissertation Summary

In this dissertation, we investigated the potential benefits of a centralized control plane for a multi-hop wireless network, identified new challenges under this new paradigm, and devised innovative solutions for optimal performance via a centralized control plane. Given that the performance of a wireless network heavily depends on its physical layer capabilities, we considered a number of advanced wireless technologies, including MIMO, full duplex, and interference cancellation and others at the physical layer. The dissertation's focus was on building tractable computational models for these wireless technologies that can be used for modeling, analysis and optimization in the centralized control plane. Problem formulation and efficient solution procedures were developed for various centralized optimization problems across multiple layers. End-to-end throughput maximization was a key objective among these optimization problems on the centralized control plane and was used to demonstrate the superior advantage of this paradigm.

In chapter 2, we proposed to integrate MIMO DoF interface cancellation (IC) and successive interference cancellation (SIC) in MIMO multi-hop network under degree-of-freedom (DoF) protocol model. We showed that DoF-based IC and SIC can help each other. We developed the

necessary mathematical models to realize the idea in a multi-hop wireless network. In chapter 3, we investigated the performance of full-duplex (FD) in MIMO multi-hop network. We showed that if IC is exploited, FD can achieve significant higher throughput gain over HD in MIMO multi-hop network, which is contrary to the recent literature suggesting an unexpected marginal gain. By incorporating IC and optimal scheduling, we manage to handle the additional network interference introduced by FD links. In chapter 4 we proposed the idea of the programmable control plane for tactical wireless network under protocol model. The programmable control plane decouples the control and data plane. In the proposed programmable control plane, the network control layer functionalities can be dynamically configured to adapt to specific physical conditions, customized applications and/or certain tactical situations. The proposed programmable control plane functionalities can be cast into a centralized optimization problem, which can be updated as needed. In chapter 5, we proposed the idea of the unified programmable control plane for tactical heterogeneous wireless networks with interference management capabilities under SINR model. The unified control plane can abstract the complexity of heterogeneous wireless networks and can provide a centralized control over the network resources. We developed the necessary mathematical model to realize the unified programmable control plane for heterogeneous wireless networks.

6.2 Future Work

In this dissertation, we laid the theoretical foundation for a centralized control plane in a multi-hop wireless network within the programmable control plane paradigm. In the next step, we would like to bridge the gap between the theory and practice.

As an initial phase, we will incorporate our centralized optimization framework within NS3 network simulator. Our purpose is to simulate the unified programmable control plane which controls the operation of a tactical wireless heterogeneous network. In section 6.2.1, we propose our centralized control plane cross-layer model, which is envisioned to periodically run in the OpenFlow centralized controller to control the operation of a multi-hop single-antenna wireless

heterogeneous network under protocol model. In section 6.2.2, we will discuss our simulation design for testing the functional fidelity of the proposed centralized control plane a multi-hop wireless network.

As a secondary phase, we are planning to create a unified programmable control plane testbed, in which we will have a heterogeneous wireless network consisting of multiple nodes operating in different technologies. We are interested in verifying the functional fidelity of the proposed ideas in this dissertation in practice.

6.2.1 Cross-layer Optimization for Programmable Control Plane

The programmable control plane provides a centralized network controller to perform resource allocation in a wireless heterogeneous network. Therefore, a centralized optimization is a natural tool to study, model, and implement various algorithms. To understand the concept of programmable control plane in a wireless heterogeneous network, we study the specific application of end-to-end session throughput maximization problem in a multi-hop wireless heterogeneous network. We will show how this particular application can be modeled and formulated in the programmable control plane.

Consider a multi-hop network consisting of a set of nodes $\mathcal{N}$ which has N elements. Each node is assumed to have a single antenna. We assume that there is a set $\mathcal{F}$ of active sessions in the network. Denote $s(f)$ and $d(f)$ as the source and destination nodes of session $f \in \mathcal{F}$, respectively.

Scheduling Constraints. We assume each heterogeneous wireless network (e.g. WiFi, LTE, WiMAX, etc.) are operating on orthogonal frequency bands $b \in \mathcal{B}$, where $\mathcal{B}$ is the set of frequency bands. We consider a time-slotted scheduling, where a time frame consists of T time slots. Depending on link scheduling, a subset of links will be active in time slot t, $1 \leq t \leq T$. For unicast communication, among all the nodes that hear the signal, only one node belongs to the unicast session and is supposed to decode the signal. Consider a transmitter and its intended receiver as a (logical) link. Denote $\mathcal{L}$ as the set of all possible links in the network and $z_l^b[t]$ as a binary variable

to indicate whether or not link $l \in \mathcal{L}$ is active in time slot t, i.e., $z_l^b[t] = 1$ if link l is active and 0 otherwise. Then we have:

$$\sum_{l \in \mathcal{L}_i^{\text{out}}} z_l^b[t] \leq 1, (i \in \mathcal{N}, b \in \mathcal{B}, 1 \leq t \leq T) \ , \tag{6.2.1}$$

$$\sum_{l \in \mathcal{L}_i^{\text{in}}} z_l^b[t] \leq 1, (i \in \mathcal{N}, b \in \mathcal{B}, 1 \leq t \leq T) \ . \tag{6.2.2}$$

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

Half Duplex Constraints. Although there has been significant advance on full duplex, it has not yet become part of any wireless standards. Therefore, we assume half duplex on a node in this paper. For half-duplex within the same band, we have the following constraint:

$$\sum_{l \in \mathcal{L}_i^{\text{out}}} z_l^b[t] + \sum_{k \in \mathcal{L}_i^{\text{in}}} z_k[t] \leq 1, \quad (i \in \mathcal{N}, b \in \mathcal{B}, 1 \leq t \leq T) \ . \tag{6.2.3}$$

Protocol Model Constraints. Denote $\text{Tx}(l)$ and $\text{Rx}(l)$ as the transmit and receive nodes of link l, $1 \leq l \leq L$. In a protocol model, the transmitter of link k is not supposed to transmit in the interference range of the receiver of link l. The above constraint can be expressed in mathematical form as following. For $(l, k \in \mathcal{L}, \text{Tx}(k) \in \mathcal{I}_{\text{Rx}(l)}, \text{Tx}(k) \neq \text{Tx}(l), b \in \mathcal{B}, 1 \leq t \leq T)$, we have:

$$z_l^b[t] + z_k^b[t] \leq 1, \tag{6.2.4}$$

Link Capacity Constraints. For each link l, denote the link capacity as C_l, e.g. $C_l = B \log_2(1 + \frac{Q_{\text{Tx}(l)} d_l^{-\alpha\lambda}}{N_0})$, where B is bandwidth, $Q_{\text{Tx}(l)}$ is the power spectral density from transmitter of link l, d_l is the distance between the transmitter and the receiver of link l, α is the path loss index, λ is the antenna related constant, and N_0 is the ambient Gaussian noise density. Denote $r_l(f)$ as the amount of data rate on link l that is attributed to session $f \in \mathcal{F}$. Since the aggregate data rate on link l cannot exceed the link's average rate (over T time slot), we have

$$\sum_{f \in \mathcal{F}} r_l(f) \leq \frac{1}{T \cdot |\mathcal{B}|} \sum_{t=1}^{T} \sum_{b \in \mathcal{B}} C_l \cdot z_l^b[t], \quad (l \in \mathcal{L}). \tag{6.2.5}$$

where the right-hand-side represents the average throughput on link l over a frame (T time slots).

Flow Routing Constraints. Denote $r(f)$ as the end-to-end throughput of session $f \in \mathcal{F}$. Then at source node, $s(f)$, $f \in \mathcal{F}$, we have the following flow balance:

$$\sum_{l \in \mathcal{L}_i^{\text{out}}} r_l(f) = r(f), \quad (i = s(f), f \in \mathcal{F}) \ . \tag{6.2.6}$$

At any intermediate relay node, we have

$$\sum_{l \in \mathcal{L}_i^{\text{in}}} r_l(f) = \sum_{l \in \mathcal{L}_i^{\text{out}}} r_l(f), (1 \leq i \leq N, i \neq s(f), i \neq d(f), f \in \mathcal{F}) \ . \tag{6.2.7}$$

At a destination node, we have **Scheduling Constraints.**

$$\sum_{l \in \mathcal{L}_i^{\text{in}}} r_l(f) = r(f), \quad (i = d(f), f \in \mathcal{F}) \ . \tag{6.2.8}$$

Problem Formulation. In this programmable control plane application, we are interested to maximize end-to-end session throughput while considering the fairness issue. Fairness can be defined in different ways. In this programmable control plane application, we choose the objective of maximizing the minimum throughput among all active sessions in the network ($r_{\min} = \min_{f \in \mathcal{F}} r(f)$). This objective aims to achieve both throughput maximization and fairness among the sessions. The problem can be formulated as follows:

$$\begin{aligned}
\textbf{TMP} \quad \max \quad & r_{\min} \\
\text{s.t.} \quad & r_{\min} \leq r(f) \qquad (f \in \mathcal{F}); \\
& \text{Scheduling Constraints: (6.2.1),(6.2.2);} \\
& \text{Half Duplex Constraint: (6.2.3);} \\
& \text{Protocol Model Constraints:(6.2.4);} \\
& \text{Link Capacity Constraints: (6.2.5);} \\
& \text{Flow balance constraints: (6.2.6),(6.2.7);}
\end{aligned}$$

TMP is a mixed integer linear program (MILP). Although the theoretical worst-case complexity to a general MILP problem is exponential [28, 68], there exist highly efficient heuristics (e.g.,

sequential fixing algorithm [38, Chapter 10]) to solve it. Another approach is to apply an off-the-shelf solver (CPLEX [98]), which we find can handle up to a moderate-sized network successfully.

6.2.2 Simulation Methodology

Military networks carry sensitive data and will follow a strict operational and implementation guileless in compliance with the policies. It will be difficult and challenging to test any new technologies without proving such technologies are safe to implement and bring additional capabilities. Modeling and simulation of any new technology will provide an opportunity to safely test the fidelity of new network models and implementations. We will choose the path of simulating heterogeneous wireless network topologies to apply unified control plane model and test its fidelity and optimality. We have looked into several network simulation software packages and finally decided to use NS3 for development of our unified control model.

Architecture and Design. Our envisioned programmable control plane which enables centralized optimization will be based on the modification to OpenFlow protocol. The programmable control plane will be implemented on-demand basis in tactical field along with existing infrastructure-less classical ad-hoc schemes, which serves as a back up control plane. The control channel could be accessed by all nodes, in which OpenFlow protocol messages are exchanged between the network controller and the nodes. Fig. 6.1 demonstrates our envisioned on-demand programmable control plane for wireless multi-hop network. An overview of our designs is as follows:

- **Step 1.** Nodes periodically update the network controller with their status (such as node's ID, GPS coordination, radio type, number of antenna, etc.) Fig. 6.2 shows node level data abstraction, which are transmitted periodically to the network controller.
- **Step 2.** Session initiation request are sent from the nodes to the network controller. Fig. 6.3 shows session request data abstraction sent to the network controller.
- **Step 3.** TMP described above will be solved within the network controller with respect to collected information on the network. The optimal solution for link scheduling and routing

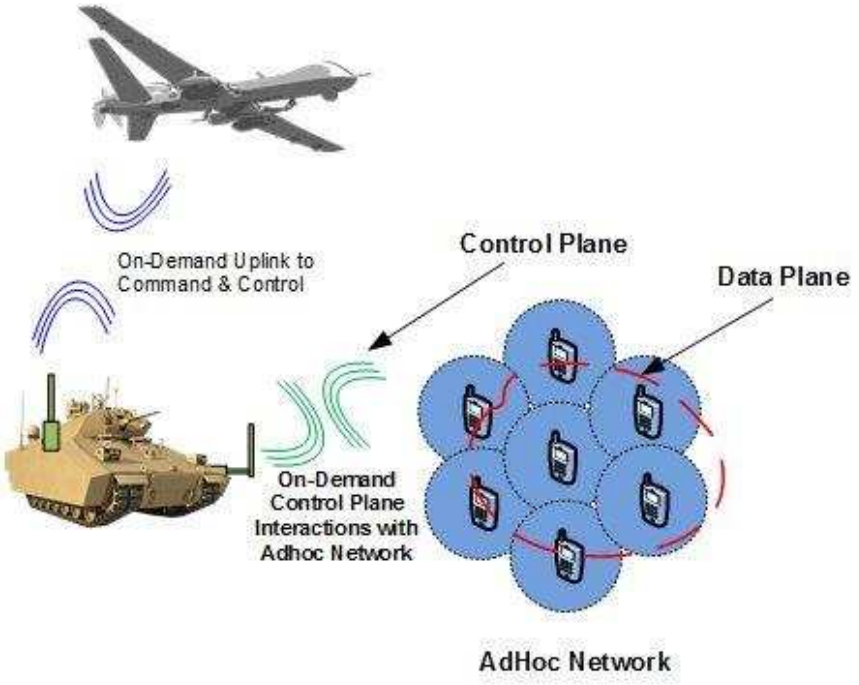


Figure 6.1: On-demand programmable control plane architecture.

```
struct ofp_node_description{
uint16_t NODE_ID
uint16_t GPS_COORDS
uint16_t TIME_STAMP
uint_16_t RADIO_TYPES
uint_16_t ANTENNA_NUMBER};
OFP_ASSERT(sizeof(struct ofp_NODE_DESCRIPTION) ==56);
```

Figure 6.2: Node's data collected periodically by OpenFlow network controller.

will be sent to the network by the network controller. Fig. 6.4 shows the data abstraction for optimal network solution sent from the controller to the network.

- **Step 4.** The network will operate according to scheduling data received from the network controller until further update.

```
struct ofp_session_vars{
uint16_t SESSION_ID
uint16_t SRC_NODE
uint16_t DST_NODE
uint16_t TIME_STAMP};
OFP_ASSERT(sizeof(struct ofp_SESSION) ==56);
```

Figure 6.3: Session initiation request data.

```
struct oft_opt_push{
uint16_t NODE_ID
uint16_t OPT_FREQUENCY
uint16_t TIME_SLOT
uint16_t NODE_SELF_ROLE
uint16_t DEST_NODE_ID
uint16_t NODE_SELF_ROLE
uint16_t OPT_DEST_ROLE
uint16_t TIME_STAMP};
ofp_OPT_PUSH==56);
```

Figure 6.4: Optimal scheduling and resource allocation network decision variables sent to nodes by OpenFlow controller.

6.2.3 Challenges and Limitations

It could be debated whether the centralized global optimization over the programmable control plane introduces any additional delay for the resource management and control in a wireless multi-hop network. We argue that the delay for control exists in both classical and programmable control networks. Therefore, the proposed idea in this dissertation will only improve the current existing challenges in the optimization and control of the wireless multi-hop network and will not introduce any additional delay to the existing classical network. On the other hand, the delay originates from the control data overhead and the control channel bandwidth. The delay can be tuned depending on what amount of computation (abstraction level) is performed locally and what computation is done at the network controller. In other words, the node data abstraction level determines the amount of overhead. For example, under our proposed scheme the advanced physical layer signal processing locally at the access point or at the network controller depending on the additional delay introduced and how frequently these computations needs to be done.

In conclusion, by utilizing the network function abstraction and the network virtualization approaches, only necessary limited meta-data needs to be transfered over the control channel, which significantly reduces the overhead. This essential meta-data is sufficient input for the global network optimization algorithm to perform network-wide resource allocation. Lastly, the programmability of the control plane allows an adaptive network control framework, i.e., the network control framework can be modified adaptively based on the network topology, node's computation power, control channel bandwidth, high performance computing power for the network controller, wireless channel conditions, and use-case scenarios.

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] E. Aryafar, N. Anand, T. Salonidis, and E. Knightly, "Design and experimental evaluation of multi-user beamforming in wireless LANs," in *Proc. ACM MobiCom,* pp. 197–208, Chicago, IL, Sept. 20–24, 2010.

[3] 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.

[4] J.G. Andrews, "Interference cancellation for cellular systems: A contemporary overview," *IEEE Wireless Commun. Magazine,* vol. 12, no. 2, pp. 19–29, Apr. 2005.

[5] 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.

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

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

[8] D. Bharadia and S. Katti, “Full Duplex MIMO Radios,” in *Proc. USENIX NSDI,* pp. 359–372, Seattle, WA, Apr. 2014.

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

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

[11] J. Blomer and N. Jindal, “Transmission capacity of wireless ad hoc networks: Successive interference cancellation vs. joint detection,” in *Proc. IEEE ICC,* Dresden, Germany, June 14–18, 2009.

[12] 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. 404–412, Salt Lake City, UT, June 2011.

[13] T. Chen, M. Matinmikko, X. Chen, X. Zhou, and P. Ahokangas, “Software defined mobile networks: concept, survey, and research directions,” *IEEE Communications Magazine,* vol. 53, no. 11, pp. 123–133, November 2015.

[14] 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.

[15] 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.

[16] W.-J. Choi, R. Negi, and J.M. Cioffi, “Combined ML and DFE decoding for the V-BLAST system,” in *Proc. ICC ,* vol. 3, pp. 1243–1248, 2000.

[17] M. Duarte and A. Sabharwal, "Full-duplex Wireless Communications Using Off-the-shelf Radios: Feasibility and First Results," in *Proc. Asilomar Conf. Conference on Signals, Systems, and Computers,*, pp. 1558–1562, Pacific Grove, CA, USA, 2010.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

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

[23] R. Fletcher and S. Leyffer, "Solving mixed integer programs by outer approximation," *Mathematical Programming,* vol. 66, no. 1–3, pp. 327–349, 1994.

[24] G.J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas", *Bell Laboratories Technical Journal,* vol. 1, no. 2, pp. 41–59, 1996.

[25] G.J. Foschini, G.D. Golden, R.A. Valenzuela, and P.W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE Journal on Selected Areas in Commun.,* vol. 17, no. 11, pp. 1841–1852, No. 1999.

[26] P. Frenger, P. Orten, and T. Ottosson, "Code-spread CDMA with interference cancellation", *IEEE Journal on Selected Areas in Commun.,* vol. 17, no. 12, pp. 2090–2095, Dec. 1999.

[27] 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.

[28] M.R. Garey and D.S. Johnson, *Computers and Interactability: A Guide to the Theory of NP-Completeness,* W. H. Freeman and Company, New York, 1979.

[29] E. Gelal, N. Jianxia, K. Pelechrinis, K. Tae-Suk, I. Broustis, S.V. Krishnamurthy, and B.D. Rao, "Topology control for effective interference cancellation in multiuser MIMO networks," *IEEE/ACM Trans. on Networking,* vol. 21, no. 2, pp. 455–468, Apr. 2013.

[30] A.M. Geoffrion, "A generalized Benders decomposition," *J. Optimization Theory and Applications,* vol. 10, no. 4, pp. 237–260, 1972.

[31] T.R. Giallorenzi and S.G. Wilson, "Suboptimum multiuser receivers for convolutionally coded asynchronous DS-CDMA systems," *IEEE Trans. on Commun.,* vol. 44, no. 9, pp. 1183–1196, Sept. 1996.

[32] 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.

[33] O. Goussevskaia and R. Wattenhofer, "Scheduling with interference decoding: Complexity and algorithms," *Ad Hoc Networks,* vol. 11, no. 6, pp. 1732–1745, Aug. 2013.

[34] O. K. Gupta and A. Ravindran, "Branch and bound experiments in convex nonlinear integer programming," *Management Science,* vol. 31, no. 12, pp. 1533–1546, 1985.

[35] D. Halperin, T. Anderson, and D. Wetherall, "Taking the sting out of carrier sense: Interference cancellation for wireless LANs," in *Proc. ACM MobiCom.,* pp. 339–350, San Francisco, CA, Sept., 2008 .

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

[38] Y.T. Hou, Y. Shi, and H.D. Sheralli, *Applied Optimization Methods for Wireless Networks,* Cambridge University Press, 2014, ISBN-13: 978-1107018808.

[39] 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.

[40] B. Jalaeian, Y. Shi, X. Yuan, Y.T. Hou, W. Lou, and S.F. Midkiff, "Harmonizing SIC and MIMO DoF Interference Cancellation for Efficient Network-wide Resource Allocation," in *Proc. IEEE International Conference on Mobile Ad hoc and Sensor Systems (IEEE MASS 2015),* pp. 316–323, Dallas, Texas, October, 2015.

[41] C. Jiang, Y. Shi, Y.T. Hou, W. Lou, S. Kompella, and S. F. Midkiff, "Squeezing the most out of interference: An optimization framework for joint interference exploitation and avoidance," in *Proc. IEEE INFOCOM,* pp. 424–432, Orlando, Florida, March, 2012.

[42] C. Jiang, Y. Shi, Y.T. Hou, W. Lou, and H.D. Sherali, "Throughput Maximization for Multi-Hop Wireless Networks with Network-Wide Energy Constraints," in *IEEE Trans. on Wireless Communication,* vol. 2, no. 3, pp. 1255–1267, 2013.

[43] P. Jung and M. Nasshan, "Results on Turbo-codes for speech transmission in a joint detection CDMA mobile radio system with coherent receiver antenna diversity," *IEEE Trans. on Vehicular Technology,* vol. 46, no. 4, pp. 862–870, Apr. 1997.

[44] 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.

[45] 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.

[46] 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

[47] 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.

[48] 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.

[49] Y. Shi, J. Liu, C. Jiang, C. Gao, and Y.T. Hou, "A DoF-based Link Layer Model for Multihop MIMO Networks," *IEEE Trans. on Mobile Computing,* no. 99, 2014.

[50] P. Liu and I.-M. Kim, "Exact and closed-form error performance analysis for hard MMSE-SIC detection in MIMO systems," *IEEE Trans. on Communications,* vol. 59, no. 9, pp. 2463–2477, Sept. 2011.

[51] X. Liu, A. Sheth, M. Kaminsky, K. Papagiannaki, S. Seshan, and P. Steenkiste, "Pushing the envelope of indoor wireless spatial reuse using directional access points and clients," in *Proc. ACM MobiCom,* pp. 209–220, Chicago, IL, Sept., 2010.

[52] S. Lv, X. Wang, and X. Zhou, "Scheduling under SINR model in adhoc networks with successive interference cancellation," in *Proc. IEEE GLOBECOM,* Miami, FL, Dec. 6–10, 2010.

[53] S. Lv, W. Zhuang, X. Wang, C. Liu, and X. Zhou, "Maximizing Capacity in the SINR model in Wireless Networks with Successive Interference Cancellation," in *Proc. IEEE ICC,* pp. 1–6, June 2011.

[54] S. Lv, W. Zhuang, X. Wang, and X. Zhou, "Scheduling in wireless ad hoc networks with successive interference cancellation," in *Proc. IEEE INFOCOM,* pp. 1282–1290, Shanghai, China, Apr. 10–15, 2011.

[55] S. Lv, W. Zhuang, X. Wang, and X. Zhou, "Link scheduling in wireless networks with successive interference cancellation," *Elsevier Computer Networks,* vol. 55, no. 13, pp. 2929–2941, Sept. 2011.

[56] S. Lv, W. Zhuang, M. Xu, X. Wang, C. Liu, and X. Zhou, "Understanding the scheduling performance in wireless networks with successive interference cancellation," *IEEE Trans. on Mobile Computing,* vol. 12, no. 8, pp. 1625–1639, Aug. 2013.

[57] 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.

[58] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "OpenFlow: Enabling Innovation in Campus Networks," in *Proc. SIGCOMM Comput. Commun. Rev.* vol. 38, no. 2, pp. 69–74, March 2008.

[59] D.N. Nguyen and M. Krunz, "Be Responsible: A Novel Communications Scheme for Full-Duplex MIMO Radios," in *Proc. IEEE INFOCOM,* Hongkong, May 2015.

[60] H.X. Nguyen, C. Jinho, and T. Le-Ngoc, "High-rate groupwise STBC using low-complexity SIC based receiver," *IEEE Trans. Wireless Commun.,* vol. 8, no. 9, pp. 4677–4687, Sept. 2009.

[61] B.A.A. Nunes, M. Mendonca, X.N. Nguyen, K. Obraczka, and T. Turletti, "A Survey of Software-Defined Networking: Past, Present, and Future of Programmable Networks," *IEEE Communications Surveys Tutorials,*vol. 16,no. 3, pp. 1617–1634, March 2014.

[62] 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 2005.

[63] 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.

[64] 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.

[65] 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.

[66] 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.

[67] A.A. Sani, L. Zhong, and A. Sabharwal, "Directional antenna diversity for mobile devices: Characterizations and solutions," in *Proc. ACM MobiCom,* pp. 221–232, Chicago, IL, Sept., 2010.

[68] A. Schrijver, "Theory of Linear and Integer Programming," Cambridge University Press, New York, NY 1986.

[69] S. Sfar, R.D. Murch, and K.B. Letaief, "Layered space-time multiuser detection over wireless uplink systems," *IEEE Trans. Wireless Commun.,* vol. 2, no. 4, pp. 653–668, July 2003.

[70] H.D. Sherali and W.P. Adams, *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems,* Chapter 8, Kluwer Academic Publishers, 1999.

[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 Trans. on Mobile Computing,* vol. 13, no. 99, pp. 1395–1408, July 2014.

[72] A.P. Subramanian, H. Lundgren, and T. Salonidis, "Experimental characterization of sectorized antennas in dense 802.11 wireless mesh networks," in *Proc. ACM MobiHoc,* pp. 259268, New Orleans, LA, May 1821, 2009.

[73] S. Sun, L. Gong, B. Rong, and K. Lu, "An intelligent SDN framework for 5G heterogeneous networks," *IEEE Communications Magazine,* vol. 53, no. 11, pp. 142–147, Nov. 2015.

[74] 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 Trans. on Mobile Computing,* vol. 3, no. 4, pp. 350–365, Oct. 2004.

[75] 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.

[76] D.N.C. Tse and P. Viswanath. *Fundamentals of Wireless Communication,* Cambridge University Press, May 2005.

[77] A. Tolli, M. Codreanu, and M. Juntti, "Cooperative MIMO-OFDM cellular system with soft handover between distributed base station antennas," *IEEE Trans. on Wireless Commun,* vol. 7, no. 4, pp. 1428–1440, April, 2008.

[78] M.K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code-division multiple access communications," *IEEE Trans. on Commun.,* vol. 38, no. 4, pp. 509519, Apr. 1990.

[79] M. K. Varanasi and T. Guess, "Optimum decision feedback multiuser equalization and successive decoding achieves the total capacity of the Gaussian multiple-access channel," in *Proc. Asilomar Conference on Signals, Systems and Computers,* vol. 2, pp. 1405–1409, Pacific Grove, CA, Nov., 1997.

[80] A.J. Viterbi, "Very low rate convolutional codes for maximum theoretical performance of spread-spectrum multiple-access channel," *IEEE Journal on Selected Areas in Commun.,* vol. 8, no. 4, pp. 641–649, May 1990.

[81] S. Verdu, *Multiuser Detection,* Cambridge University Press, 1998.

[82] X. Wang and H.V. Poor, "Iterative (Turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. on Commun.,* vol. 47, no. 7, pp. 1046–1061, July 1999.

[83] S. Weber, J.G. Andrews, X. Yang, and G. de Veciana, "Transmission capacity of wireless ad hoc networks with successive interference cancellation," *IEEE Trans. on Information Theory,* vol. 53, no. 8, pp. 2799–2814, Aug. 2007.

[84] T. Westerlund and F. Pettersson, "An extended cutting plane method for solving convex MINLP problems," *Computers Chem. Eng.,* vol. 19, supplement 1, pp. 131–136, 1995.

[85] P.W. Wolniansky, G.J. Foschini, G.D. Golden, and R. Valenzuela,, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," *International Symposium on Signals, Systems, and Electronics,* pp. 295–300, Pisa, Sep., 1998.

[86] K. Wong and R. D. Murch, and K.B Letaief, "Performance enhancement of multiuser MIMO wireless communication systems," *IEEE Trans. on Communications,* vol. 50, no. 12, pp. 1960–1970, Dec, 2002.

[87] 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.

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

[89] Y. Yang, B. Chen, K. Srinivasan, and N.B. Shroff, "Characterizing the achievable throughput in wireless networks with two active RF chains," in *Proc. IEEE INFOCOM,* pp. 262–270, Toronto, Canada, April 2014.

[90] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Commun.,* vol. 24, no. 3, pp. 528–541, March 2006.

[91] D. Yuan, V. Angelakis, L. Chen, E. Karipidis, and E.G. Larsson, "On Optimal Link Activation with Interference Cancellation in Wireless Networking," *IEEE Trans. on Vehicular Technology,* vol. 62, no. 2, pp. 939–945, Feb. 2013.

[92] A. Zanella, M. Chiani, and M.Z. Win, "MMSE reception and successive interference cancellation for MIMO systems with high spectral efficiency," *IEEE Trans. on Wireless Communications,* vol. 4, no. 3, pp. 1244-1253, May 2005.

[93] H. Zeng, Y. Shi, Y.T. Hou, R. Zhu, and W. Lou, "A novel MIMO DoF model for multi-hop networks," *IEEE Network Magazine,* vol. 28, pp. 81–85, October 2014.

[94] L. Zheng and D.N.C. 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.

[95] 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.

[96] Q. Zhou, C.X. Wang, S. McLaughlin, and X. Zhou, "Network virtualization and resource description in software-defined wireless networks," *IEEE Communications Magazine,* vol. 53, no. 11, pp. 110–117, 2015.

[97] X. Zhu and R.D. Murch, "Performance analysis of maximum likelihood detection in a MIMO antenna system," *IEEE Trans. on Communications,* vol. 50,no. 2,pp. 187–191, Feb. 2002.

[98] IBM ILOG CPLEX Optimizer, http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/.

[99] Open Network Foundation, https://www.opennetworking.org.

[100] Open Networking Research Center (ONRC), http://onrc.net/.

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