Some Optimization Problems in Wireless Networks

Canming Jiang

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

Doctor of Philosophy
in
Computer Engineering

Y. Thomas Hou, Chair
Yi Shi
Hanif D. Sherali
Wenjing Lou
Anil Vullikanti
Yaling Yang

June 15, 2012
Blacksburg, Virginia

Keywords: Wireless networks, optimization, algorithms, asymptotic capacity, cross-layer design

Some Optimization Problems in Wireless Networks

Canming Jiang

ABSTRACT

Recently, many new types of wireless networks have emerged for both civil and military applications, such as cognitive radio networks, MIMO networks. There is a strong interest in exploring the optimal performance of these new emerging networks, e.g., maximizing the network throughput, minimizing network energy consumption. Exploring the optimal performance objectives of these new types of wireless networks is both important and intellectual challenging. On one hand, it is important for a network researcher to understand the performance limits of these new wireless networks. Such performance limits are important not only for theoretical understanding, but also in that they can be used as benchmarks for the design of distributed algorithms and protocols. On the other hand, due to some unique characteristics associated with these networks, existing analytic techniques may not be applied directly to obtain the optimal performance. As a result, new theoretical results, along with new mathematical tools, need to be developed.

The goal of this dissertation is to make a fundamental advance on network performance optimization via exploring a series of optimization problems. Based on the scale of the underlying wireless network, the works in this dissertation are divided into two parts. In the first part, we study the asymptotic capacity scaling laws of different types of wireless networks. By "asymptotic", we mean that the number of nodes in the network goes to infinity. Such asymptotic capacity scaling laws offer fundamental understandings on the trend of maximum user throughput behavior when the network size increases. In the second part of this dissertation, we study several optimization problems of finite-sized wireless networks. Under a given network size, we accurately characterize some performance limits (e.g., throughput, energy consumption) of wireless networks and provide solutions on how to achieve the optimal objectives. The main contributions of this dissertation can be summarized as follows, where the first three problems are on asymptotic capacity scaling laws and the last three problems are optimization problems of finite-sized wireless networks.

1. **Capacity Scaling Laws of Cognitive Radio Ad Hoc Networks.** We first study the capacity scaling laws for cognitive radio ad hoc networks (CRNs), i.e., how each individual node's maximum throughput scales as the number of nodes in the network increases. This effort is critical to the fundamental understanding of the scalability of such network. However, due to the heterogeneity in available frequency bands at each node, the asymptotic capacity is much more difficult to develop than prior efforts for other types of wireless networks. To overcome this difficulty, we introduce two auxiliary networks ζ and α to analyze the capacity upper and lower bounds. We derive the capacity results under both the protocol model and the physical model. Further, we show that the seminal results developed by Gupta and Kumar for the simple single-channel single-radio (SC-SR) networks are special cases under the results for CRNs.

2. **Asymptotic Capacity of Multi-hop MIMO Ad Hoc Networks.** Multi-input multi-output (MIMO) is a key technology to increase the capacity of wireless networks. Although there has been extensive work on MIMO at the physical and link layers, there has been limited work on MIMO at the network layer (i.e., multi-hop MIMO ad hoc network), particularly results on capacity scaling laws. In this work, we investigate capacity scaling laws for MIMO ad hoc networks. Our goal is to find the achievable throughput of each node as the number of nodes in the network increases. We employ a MIMO network model that captures spatial multiplexing (SM) and interference cancellation (IC). We show that for a MIMO network with n randomly located nodes, each equipped with γ antennas and a rate of W on each data stream, the achievable throughput of each node is $\Theta(\frac{\gamma W}{\sqrt{n \ln n}})$.

3. **Toward Simple Criteria for Establishing Capacity Scaling Laws.** Capacity scaling laws offer fundamental understanding on the trend of user throughput behavior when the network size increases. Since the seminal work of Gupta and Kumar, there have been tremendous efforts developing capacity scaling laws for ad hoc networks with various advanced physical layer technologies. These efforts led to different custom-designed approaches, most of which were intellectually challenging and lacked universal properties that can be extended

to address scaling laws of ad hoc networks with a different physical layer technology. In this work, we present a set of simple yet powerful general criteria that one can apply to quickly determine the capacity scaling laws for various physical layer technologies under the protocol model. We prove the correctness of our proposed criteria and validate them through a number of case studies, such as ad hoc networks with directional antenna, MIMO, cognitive radio, multi-channel and multi-radio, and multiple packet reception. These simple criteria will serve as powerful tools to networking researchers to obtain throughput scaling laws of ad hoc networks under different physical layer technologies, particularly those to appear in the future.

4. **Exploiting SIC for Multi-hop Wireless Networks.** There is a growing interest on exploiting interference (rather than avoiding it) to increase network throughput. In particular, the so-called *successive interference cancellation* (SIC) scheme appears very promising, due to its ability to enable concurrent receptions from multiple transmitters and interference rejection. However, due to some stringent constraints and limit, SIC alone is inadequate to handle all concurrent interference. We advocate a joint interference exploitation and avoidance approach, which combines the best of interference exploitation and interference avoidance, while avoiding each's pitfalls. We discuss the new challenges of such a new approach in a multi-hop wireless network and propose a formal optimization framework, with cross-layer formulation of physical, link, and network layers. This framework offers a rather complete design space for SIC to squeeze the most out of interference. The goal of this effort is to lay a mathematical foundation for modeling and analysis of a joint interference exploitation and avoidance scheme in a multi-hop wireless network. Through modeling and analysis, we develop a tractable model that is suitable for studying a broad class of network throughput optimization problems. To demonstrate the practical utility of our model, we conduct a case study. Our numerical results affirm the validity of our model and give insights on how SIC can optimally interact with an interference avoidance scheme.

5. **Throughput Optimization with Network-wide Energy Constraint.** Conserving network-

wide energy consumption is becoming an increasingly important concern for network operators. In this work, we study network-wide energy conservation problem which we hope will offer insights to both network operators and users. Specifically, we study how to maximize network throughput under a network-wide energy constraint for a general multi-hop wireless network. We formulate this problem as a mixed-integer nonlinear program (MINLP). We propose a novel piece-wise linear approximation to transform the nonlinear constraints into linear constraints. We prove that the solution developed under this approach is near-optimal with guaranteed performance bound.

6. **Bicriteria Optimization in Multi-hop Wireless Networks.** Network throughput and energy consumption are two important performance metrics for a multi-hop wireless network. Current state-of-the-art research is limited to either maximizing throughput under some energy constraint or minimizing energy consumption while satisfying some throughput requirement. However, the important problem of how to optimize both objectives simultaneously remains open. In this work, we take a multicriteria optimization approach to offer a systematic study on the relationship between the two performance objectives. We show that the solution to the multicriteria optimization problem characterizes the envelope of the entire throughput-energy region, i.e., the so-called optimal throughput-energy curve. We prove some important properties of the optimal throughput-energy curve. For case study, we consider both linear and nonlinear throughput functions. For the linear case, we characterize the optimal throughput-energy curve precisely through parametric analysis, while for the nonlinear case, we use a piece-wise linear approximation to approximate the optimal throughput-energy curve with arbitrary accuracy. Our results offer important insights on exploiting the trade-off between the two performance metrics.

Dedication

To my parents, and my sister

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 still remember that I almost gave up my research on asymptotic capacity when I made little progress over an extended period of time. It was his encouragement and confidence that led me to continue and explore further, otherwise Chapter 4 (a significant advance in this area) would have never come out. I will always remember the numerous occasions when I sat besides him while he was helping me revise my write-up word by word. He tried to understand every sentence and equation I wrote, and helped me write more clearly and succinctly. His scholarship and dedication to research make him a role model for me as I continue along my career path.

I also want to thank Dr. Yi Shi for his help and guidance. Whenever I got stuck with a difficult problem, I would always turn to him for rescue. I enjoyed the numerous hours that we spent together solving some very hard problems, and I appreciated the lengthy hours that he spent alone proofreading my papers. Those constructive discussions and the valuable feedback from him have greatly enhanced the quality of each work in this dissertation.

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. Anil Vullikanti, and Prof. Yaling Yang. Their comments and questions have helped improve the overall quality of my dissertation.

I would like to acknowledge my current and former colleagues in my research group: Jia Liu, Sushant Sharma, Cunhao Gao, Liguang Xie, Huacheng Zeng, Xu Yuan, and Borhan Jalaeian. I not only enjoyed the time when we discussed research, but also loved the good time we had together for dinner, hiking, fishing, and basketball. They became my extended family during my five years in Blacksburg. 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 and my sister, who have supported me through many years of my academic pursuits. They have always been there to inspire and encourage me whenever I need them. No words can adequately express my deepest and utmost love to my parents and my sister, to whom I would like to dedicate this work.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and Objective

From the traditional cellular networks to the WiFi hotspots everywhere to ad hoc and sensor networks, wireless networks are becoming pervasive and are changing the way we live, work, and play. Thanks to these wireless networks, nowadays we are able to access information and communicate with others anywhere and anytime. However, with the huge demand of multimedia and video applications, existing wireless networks are not able to keep up with ever growing users' needs and demands.

To improve network performance, network operators and users are interested in finding out the performance limits of their networks. From the network operators' prospective, they want to offer as much service to users as possible with the given network resources. From the users' prospective, they want to know what is the maximum performance they can get out of their network. From either perspective, it is important to understand the performance limits of the underlying networks. Such performance limits can be obtained via network optimizations.

Network performance optimizations encompass a very broad class of problems, which intersect different techniques, mechanisms, and algorithms across the protocol stack. For wireless networks,

network optimizations are particular challenging. First and foremost, the techniques at the physical layer are changing rapidly, e.g., from the traditional single antenna system to current multiple antennas systems, from traditional fixed spectrum allocation to dynamic spectrum access. Since physical layer technique determines how information is communicated between two nodes, they directly affect network connectivity, topology, scheduling and routing decisions at the link and network layers. Each advancement at the physical layer is likely to bring new considerations in network optimization. At the link layer, sophisticated interference management algorithms will directly affect physical layer interference behavior at other nodes. Link layer scheduling also directly affects network layer routing decisions since it offers a set of active links on which a routing path can be established. At the network layer, network connectivity and throughput directly depend on the specific physical layer technique and link layer scheduling mechanism. Due to these challenges brought by the tight interaction of mechanisms and algorithms at different layers, network performance optimization remains a challenging area in wireless network research.

The goal of this dissertation is to make a concrete advance in network performance optimization via exploring a series of optimization problems. Based on the scale of the wireless networks, the work in this dissertation are divided into two parts. In the first part of this dissertation, we study the asymptotic capacity scaling laws of different types of wireless networks. By "asymptotic", we mean that the number of nodes in the network is a very large number (e.g., ∞). Such asymptotic throughput offers fundamental understanding on the trend of user throughput behavior when the network size increases. We characterize the asymptotic throughput of cognitive radio ad hoc networks and MIMO ad hoc networks. We also give two simple yet powerful criteria to determine the throughput (capacity) scaling laws for wireless networks with various physical layer technologies. In the second part of this dissertation, we study several optimization problems for finite-sized wireless networks. We first study how to exploit successive interference cancellation (SIC) for multi-hop wireless networks. Then, we study the throughput maximization problem of a multi-hop wireless network when energy constraints are presented. Finally, we study a multicriteria problem to characterize the throughput and energy envelope of a wireless ad hoc network.

1.2 Summary of Contributions

As mentioned in the last section, this dissertation studies network optimization problems for wireless networks of two scales: infinite number of nodes and finite number of nodes. Accordingly, our contribution in this dissertation will be organized in two parts.

1.2.1 Part I: Capacity Scaling Laws of Wireless Networks

For capacity scaling laws, we study how the network throughput behaviors as the total number of nodes in the network goes to infinity. In this part, we investigate the following three important problems.

- In Chapter 2, we study the capacity scaling laws of cognitive radio networks (CRN) when the number of nodes in the network goes to infinity. Specifically, we consider a CRN consisting of n nodes positioned randomly in a unit square area, where n goes to infinity. In a CRN, each node senses a set of spectrum bands that it can use, which may include the set of bands not used by primary users. Thus, the available frequency bands at one node in a CRN (which may include any unused bands by primary users) may not be identical to those at the other nodes in the network, due to its geographical location and local spectrum availability. Such heterogeneity in frequency bands at each node poses a significant challenge to analyze capacity bounds. Our goal is to overcome these challenges and obtain the maximum asymptotic per-node throughput (or per-node capacity) under both the protocol model and the physical model. Denote $\lambda(n)$ the asymptotic per-node throughput, which is the data rate that can be sent from each source to its destination.

  Under the protocol model, we show that the lower and upper bounds for $\lambda(n)$ in a random CRN are $\Omega(\frac{C_\alpha}{\sqrt{n \ln n}})$ and $O(\frac{C_\zeta}{\sqrt{n \ln n}})$, respectively (denoted as $\lambda(n) \in [\Omega(\frac{C_\alpha}{\sqrt{n \ln n}}), O(\frac{C_\zeta}{\sqrt{n \ln n}})]$), where C_α is the minimum link capacity over all links and C_ζ is the sum of capacity over all available bands in the CRN. The main novelty in our analysis is the creation of two auxiliary

networks ζ and α that help transform the original heterogeneous bands to homogenous bands on each link for asymptotic analysis.

Under the physical model, we focus on the general case where each node is allowed to perform independent power control. We show that $\lambda(n)$ in a random CRN is $\lambda(n) \in [\Omega(\frac{C_\alpha}{\sqrt{n \ln n}}), O(\frac{C_\zeta}{n^{1/\gamma}})]$. In addition, we also give the result for the special case where power level is synchronized on all nodes in the network (as done in [47]). Note that our capacity analysis are also based on the two auxiliary networks ζ and α.

- In Chapter 3, we study the asymptotic capacity of MIMO ad hoc networks when the number of nodes in the network goes to infinity. We consider a random multi-hop MIMO ad hoc network with n nodes, where each node, equipped with γ antennas, is randomly located in a unit square area. Each node acts as a source node and transmits data to a randomly chosen destination node.

  Although there are many schemes to exploit the benefits of antenna arrays at a node, we focus on the so-called *zero-forcing beamforming* (ZFBF) scheme [19, 107], which captures the two key characteristics of MIMO: *spatial multiplexing* and *interference cancellation.* For asymptotic study, we analyze both the lower bound and the upper bound. We show that although a capacity lower bound can be obtained by extending the work of Gupta and Kumar [47], a tight capacity upper bound is a much harder problem. We propose to partition the network area into small squares cleverly so that the maximum data rate that can be received by the nodes inside the small square can be computed exactly. By taking the sum of data rates from all small squares, we can obtain the maximum data rate the whole network can support. Based on this result, we develop a tight capacity upper bound for our problem. Our main result in this chapter is the following: for a MIMO network with n randomly located nodes, each equipped with γ antennas and a rate of W on each data stream, we show that the capacity upper and lower bounds have the same order, and the achievable throughput of each node is $\Theta(\frac{\gamma W}{\sqrt{n \ln n}})$.

- In Chapter 4, we provide some simple criteria for establishing asymptotic throughputs for

different types of wireless networks. Since the seminal results of Gupta and Kumar ("G&K" for short) on capacity scaling law of ad hoc networks with single omnidirectional antennas [47], there has been a flourish of research efforts on exploring capacity scaling laws for ad hoc networks under various physical layer technologies. These include directional antennas [92, 134], MIMO [56], multi-channel multi-radio (MC-MR) [68], cognitive radios [54, 55, 105, 136], and multiple packet reception (MPR) [95], among others. For each of these advanced physical layer technologies, a *custom-designed* analytical approach was developed to study its capacity scaling law. Most of these solutions are typically intellectually challenging and lack universal properties that can be extended to address scaling laws of ad hoc networks

In this chapter, we present a set of simple yet powerful general criteria that one can apply to quickly determine the asymptotic capacity upper bounds for various physical layer technologies under the protocol model. We prove the correctness of our proposed criteria and validate them through a number of case studies, such as ad hoc networks with directional antenna, MIMO, cognitive radio, and multiple packet reception. These simple criteria will serve as powerful tools to networking researchers to obtain throughput scaling laws of ad hoc networks under different physical layer technologies, particularly those to be developed in the future.

1.2.2 Part II: Optimization of Finite-Sized Wireless Networks

In the second part of this dissertation, our focus is shifted to finite-sized networks. Specifically, we investigate the following three problems.

- In Chapter 5, we exploit SIC for multi-hop wireless networks. We consider an ad hoc network with finite number of nodes, where each node is equipped with SIC capability. Under SIC, a receiver attempts to decode the concurrent signals from multiple transmitters successively, starting from the strongest signal. If the strongest signal can be decoded, it will

be subtracted from the composite signal so that the SINR (signal-to-interference-and-noise-ratio) for the remaining signals can be improved. Then the SIC receiver continues to decode the second strongest signals and so forth, until all signals are decoded, or terminates if the signal is no longer decodable. However, SIC has its own limitations and alone is inadequate to handle all concurrent interference in a multi-user wireless network. It turns out that the limitations of SIC can be alleviated precisely by a traditional interference avoidance scheme (e.g., scheduling). That is, a joint interference exploitation and avoidance approach, which combines the best of both worlds while removing each's pitfalls, is most preferable.

In this chapter, we investigate how to optimize the joint interference exploitation and avoidance scheme in a multi-hop wireless network. We address the above new challenges for such a scheme through a formal optimization framework, with cross-layer formulation of physical, link, and network layers. This framework offers a rather complete design space to squeeze the most out of interference. The goal of our effort is to lay a mathematical foundation for modeling and analysis of a joint interference exploitation and avoidance scheme in a multi-hop wireless network. Through modeling and analysis, we develop a tractable model that is suitable for studying a broad class of network throughput optimization problems when SIC is coupled with scheduling and routing. To demonstrate the practical utility of our model, we conduct a case study on maximizing the network throughput. Our numerical results affirm the validity of this model and give us insights on how SIC should optimally interact with an interference avoidance scheme.

- In Chapter 6, we study a throughput maximization problem for multi-hop wireless networks with network-wide energy constraint. With the proliferation of wireless networks, the concern of energy consumption is becoming increasingly important for network operators. Conserving network-wide energy consumption not only can help reducing CO_2 emissions [30] and protect the environment, but can also significantly reduce the operating cost for network providers [31, 93]. Consequently, in addition to pressure from government regulating agencies and citizen environmentalists, network operators now have every incentive to tackle energy conservation in their networks. As a result, energy conservation has quickly evolved

from merely an attractive but secondary feature to a key design benchmark. The so-called *Green communications* has also come to the center stage of both national and international research agenda [129, 21, 130] and consequently research efforts in this area is flourishing [71, 83].

In this chapter, we show how to maximize the network throughput under a given network-wide energy consumption budget. We show that this problem involves both network and physical layer variables and can be formulated as a mixed-integer nonlinear program (MINLP). To solve this problem efficiently, we propose a novel piece-wise linear approximation to transform the nonlinear constraints into linear constraints. We prove that the solution developed under this linear approximation is near-optimal in the sense that the performance gap between our solution and the optimal solution (despite unknown) can be set to any desired target.

- In Chapter 7, we study a bicriteria optimization problem in a multi-hop wireless network to characterize the throughput-energy envelope of the entire network. More specifically, we consider to optimize the two critical performance objectives: throughput and energy simultaneously. Current state-of-the-art research is limited to either maximizing throughput under some energy constraint or minimizing energy consumption while satisfying some throughput requirement. However, the important problem of how to optimize both objectives simultaneously remains open. In this work, we take a multicriteria optimization approach to offer a systematic study on the relationship between the two performance objectives. We show that the solution to the multicriteria optimization problem characterizes the envelope of the entire throughput-energy region, i.e., the so-called optimal throughput-energy curve.

  The significance of optimal throughput-energy curve is three-fold. First, it give an envelope of the entire throughput-energy region, which offers a global perspective on the achievable throughput-energy tradeoff. In contrast, a solution to traditional problems such as maximizing throughput under energy constraints or minimizing energy under throughput constraints only represents a point on this curve or inside this region. Second, each time when the re-

quirement on either network throughput or energy consumption changes, one can use the optimal throughput-energy curve to find a new optimal tradeoff between throughput and energy instantly, rather than resorting to solving a new optimization problem. Finally, the optimal throughput-energy curve shows us the existence of a *saturation point*, beyond which the throughput can no longer be further increased, regardless of how much additional energy is used.

Part I

Capacity Scaling Laws of Wireless Networks

Chapter 2

Capacity Scaling Laws for Cognitive Radio Networks

2.1 Introduction

In this chapter, we are interested in exploring asymptotic capacity scaling laws[1] for a cognitive radio ad hoc network (CRN) [105]. Cognitive radio (CR) [132] is a revolutionary radio technology that enjoys unparalleled advantages over traditional hardware-based radio. It is envisioned that CR will be the core radio technology for wireless networks in the future. However, due to the flexibility (and thus much larger design space) that comes with a CRN, none of the existing capacity scaling law results can be easily extended to a CRN. For instance, the available frequency bands at one node in a CRN (which may include any unused bands by primary users) may not be identical to those at the other nodes in the network, due to its geographical location and local spectrum availability. Such heterogeneity in frequency bands at each node poses a significant challenge to analyze capacity bounds.

[1]When there is no ambiguity, we use the terms "asymptotic capacity scaling law", "asymptotic throughput", "asymptotic capacity", "capacity scaling law", "scaling law" interchangeably throughout this dissertation.

2.1.1 Main Contributions

In this chapter, we study asymptotic capacity for a CRN under both the protocol and the physical models [47]. We consider a random multi-hop CRN with n nodes, with each node acting as a source node and transmitting data to a randomly chosen destination node. The per-node throughput $\lambda(n)$ is defined as the data rate that can be sent from each source to its destination. We aim to analyze the maximum asymptotic per-node throughput (or per-node capacity). Our main contributions are summarized as follows.

- Under the protocol model, we show that the lower and upper bounds for $\lambda(n)$ in a random CRN are $\Omega(\frac{C_\alpha}{\sqrt{n \ln n}})$ and $O(\frac{C_\zeta}{\sqrt{n \ln n}})$, respectively (denoted as $\lambda(n) \in [\Omega(\frac{C_\alpha}{\sqrt{n \ln n}}), O(\frac{C_\zeta}{\sqrt{n \ln n}})]$), where C_α is the minimum link capacity over all links and C_ζ is the sum of capacity over all available bands in the CRN. Note that we include the constant terms C_α and C_ζ in the scaling results to show their potential impacts (similar to that in [47] which includes a constant term W). The analysis of capacity bounds is quite challenging, due to heterogeneous available bands on each link in a random CRN. The main novelty in our analysis is the creation of two auxiliary networks ζ and α that help transform the original heterogeneous bands to homogeneous bands on each link for asymptotic analysis. The conception of auxiliary network ζ is intuitive and enables us to extend the results by Gupta and Kumar [47] to obtain an upper bound. However, creating auxiliary network α for lower bound analysis is not trivial. We show that a naive approach will yield zero for the lower bound, which is not useful (see Section 2.4.2). Instead, we offer a non-conventional approach to construct auxiliary network α, which enables us to obtain a meaningful lower bound. Note that even after proper construction of auxiliary network α, Gupta-Kumar's approach does not admit straightforward solutions and cannot be applied directly to obtain the capacity lower bound for the original CRN. This is because that although we can apply Gupta-Kumar's approach to obtain a capacity lower bound for the auxiliary network α, there does not appear to be an obvious connection between this capacity lower bound (for auxiliary network α) and that for the original heterogeneous CRN. One cannot make a straightforward claim that the capacity

lower bound for auxiliary network α is also a capacity lower bound for the original CRN without a rigorous proof, since a feasible solution to α may not be feasible to the original CRN. To obtain a meaningful capacity lower bound for the original CRN, we propose a two-step approach: (1) find a feasible solution for auxiliary network α; *and* (2) transform this solution to a feasible solution to the original CRN.

- Under the physical model, we focus on the general case where each node is allowed to perform independent power control. We show that $\lambda(n)$ in a random CRN is $\lambda(n) \in [\Omega(\frac{C_\alpha}{\sqrt{n \ln n}}), O(\frac{C_\zeta}{n^{1/\gamma}})]$. In addition, we also give the result for the special case where power level is synchronized on all nodes in the network (as done in [47]). Our capacity analysis again hinges upon the two auxiliary networks ζ and α. Again, the challenge here is the analysis for the lower bound. We exploit our earlier results for the protocol model and prove that by properly setting the interference range, a feasible solution to α under the protocol model is also a feasible solution under the physical model. Subsequently, we transform this feasible solution to α to a feasible solution to the original CRN and prove its correctness.

2.1.2 Chapter Organization

The rest of this chapter is organized as follows. Section 2.2 reviews related work on asymptotic capacity studies and offers a perspective of our work in relation to others. In Section 2.3, we describe the details of network setting for a CRN in our study. In Section 2.4, we explore capacity scaling law under the protocol model. Built upon the results in Section 2.4, we further explore capacity scaling law under the physical model in Section 2.5. Section 2.6 concludes this chapter. Table 2.1 lists the notations in this chapter.

Table 2.1: Notation in Chapter 2.

Symbol	Definition
$a(n)$	The size of a small square cell
$\mathcal{B}_{ij}$	The set of available bands on link (i,j)
$\mathcal{B}_{\zeta}$	$=\bigcup_{(i,j)\in\mathcal{E}} B_{ij}$, the union of available bands over all links in a CRN
$\mathcal{B}_{\alpha}$	The set of available bands on each link in the auxiliary network α
C_m	The capacity of band m
C_{ij}	$=\sum_{m\in\mathcal{B}_{ij}} C_m$, the aggregated link capacity on link (i,j)
C_{ζ}	Each link's capacity in the auxiliary network ζ
C_{α}	Each link's capacity in the auxiliary network α
$\mathcal{E}$	The set of all links in a CRN
$\mathcal{E}^m(t)$	The set of links that are active on band m at time t
g_{ij}	The channel gain on link (i,j)
n	The number of nodes in a CRN
$p(n)$	The common transmission power at all nodes under synchronized power control
$p_{ij}^m(t)$	The transmission power used by node i to transmit to node j on band m at time t under independent power control
$r(n)$	The common transmission range of all nodes under the protocol model
SINR_{ij}	The SINR on link (i,j)
α	An auxiliary network used for developing capacity lower bounds
β	A SINR threshold for successful transmissions in the physical model
Δ	A parameter to set the interference range in the protocol model
η	Ambient noise power
γ	Path loss index
$\lambda(n)$	Per-node throughput of a random CRN with n nodes
ψ	A feasible solution
ζ	An auxiliary network used for developing capacity upper bounds

2.2 Related Work

Since the seminal work by Gupta and Kumar [47] on capacity scaling laws for a wireless network, there has been growing interest in this important area. In [47], two types of ad hoc networks, namely, random networks and arbitrary networks, are considered. In an arbitrary network [47], each node can be arbitrarily placed and can arbitrarily choose its destination. Thus, in order to maximize the network capacity, a node is always inclined to choose a nearby node as its destination to avoid multihop routing. Given the underlying assumption for an arbitrary network, its capacity has limited utility in practice. On the other hand, in a random network [47], the topology of the network is randomly generated and cannot be changed arbitrarily afterwards. Given its generality and more realistic assumption, its capacity results have application in practice. For this reason, we focus on random networks in this chapter. Related work on random ad hoc networks can be further divided into the following two categories, namely, unicast capacity (e.g., [8, 26, 47, 68]) and multicast/broadcast capacity (e.g., [48, 81, 100, 140]). Given the focus of this chapter is on unicast capacity, we will mainly review related work in this category.

In [47], Gupta and Kumar showed that for a random network, its capacity is $\Theta(\frac{C}{\sqrt{n \ln n}})$ under the protocol model, and is $[\Omega(\frac{C}{\sqrt{n \ln n}}), O(\frac{C}{\sqrt{n}})]$ under the physical model with synchronized power control, where C is the channel capacity. In Sections 2.4 and 2.5, we will show that the result in [47] is a special case of ours.

In [8, 26, 68], unicast capacity of multi-channel ad hoc networks was investigated under the protocol model. In [8], Bhandari and Vaidya studied the capacity of multi-channel single-radio (MC-SR) networks where there are a set of channels in the network and each node can only switch to a subset of these channels. In [68], Kyasanur and Vaidya studied unicast capacity of multi-channel multi-radio (MC-MR) networks where the number of bands used at a node is limited by the number of radios at the node. In [26], Dai *et al.* extended the work of [68] on MC-MR networks with consideration of directional antennas. All these prior efforts are based on traditional hardware-based radios. There are some fundamental differences between these prior efforts and ours in this chapter. First, in these prior works, the set of channels available to each radio is the

same as those on the other radios in the network, while in our CRN, the set of bands available at each node may be different. This difference is the result of traditional hardware-based radio's inability to perform spectrum sensing and thus must be pre-assigned to a set of channels on each node, while in a CRN, each node can perform spectrum sensing and may have a different set of available frequency bands depending on spectrum availability at its location (i.e., bands unused by primary users). Second, for traditional hardware-based radio, each radio interface has limited capability, e.g., one radio can only work on one channel at any time and may not be able to switch channel on a per-packet basis. That is, at any time instance, the number of bands that can be active at a node is limited by the number of radios at the node. On the other hand, a CR that we consider in this chapter is assumed to be much more powerful and can work on many bands at the same time. That is, a CR can access to many more bands simultaneously than a hardware-based radio.

To date, work on asymptotic capacity for CRNs remains very limited. In [121], Vu *et al.* studied the capacity of a CRN for the simple single-hop case, where n pairs of cognitive transmitter and receiver wish to communicate simultaneously in the presence of a single primary transmitter-receiver link. They showed that the sum-rate of the n cognitive links scales linearly with n as $n \to \infty$. In [135], Yin *et al.* considered a secondary CRN co-located with a primary ad hoc network under the assumption that the primary network must be sparser than the secondary network. They showed that the primary network and the secondary network both can achieve the same throughput scaling law as what Gupta and Kumar [47] established for a stand-alone wireless network. Jeon *et al.* [55] studied both the case where the primary network is an ad hoc network and the case where the the primary network is an infrastructure-supported wireless network. They also showed that the primary network and the secondary network both can simultaneously achieve the same throughput scaling law as a stand-alone ad hoc network regardless of whether the primary network is an ad hoc network or an infrastructure-supported network. However, the general case where there are heterogeneous available bands is not considered. Further, the assumption that the secondary network is denser than the primary network may be overly restrictive.

Finally, there have been some recent efforts on studying capacity scaling laws under specialized wireless communication technologies, e.g., Gaussian channel model [34] (via adaptive modula-

tion and coding scheme), ultra-wide band (UWB) [138], directional antennas [92], multiple-input multi-output (MIMO)[56, 91], network coding [75], and multiple-packet reception (MPR) [123]. These works are orthogonal to ours (on CR) and thus naturally lead to different results on capacity scaling laws.

2.3 Network Model

We consider a CRN consisting of n nodes positioned randomly in a unit square area. In a CRN, each node senses a set of spectrum bands that it can use [85], which may include the set of bands not used by primary users. Thus, a CRN has a unique property that the set of available frequency bands $\mathcal{B}_i$ at a node i depends on its location and may not be identical to that of another node in the network. A node i can transmit to a node j on a particular band only if this band is available to both nodes i and j. More formally, denote $\mathcal{B}_{ij} = \mathcal{B}_i \bigcap \mathcal{B}_j$ the set of common available bands at both nodes i and j. We have the following constraint.

$$\text{A band } m \text{ can be used on link } (i,j) \text{ only if } m \in \mathcal{B}_{ij}. \tag{2.3.1}$$

This constraint must be considered explicitly for a CRN, although it is implicitly assumed under wireless networks with traditional radios. Note that we only consider a static instance of a CRN. The temporal dynamics of available bands at each node over time is not considered and will be explored in future work.

Denote $\mathcal{E}$ the set of all links in a CRN, and $\mathcal{B} = \bigcup_{(i,j)\in\mathcal{E}} \mathcal{B}_{ij}$ the union of available bands over all links in the network. Note that $\mathcal{B}$ does not include those isolated bands that are available at a node but not at its neighbor(s). For a CRN, the bandwidth of each band may also be different, leading to different capacities on different bands. Denote C_m the capacity on band m and $C_{ij} = \sum_{m\in\mathcal{B}_{ij}} C_m$, i.e, the aggregate link capacity on link (i,j) if it uses all its available bands.

For interference modeling, we consider both the protocol model and the physical model. Under the protocol model, a node i can successfully transmit data to a node j on a band $m \in \mathcal{B}_{ij}$ at time

t if and only if

- Receiving node j is within the transmission range of node i, i.e.,

$$d_{ij} \leq r(n) \ , \tag{2.3.2}$$

  where d_{ij} is the distance between nodes i and j and $r(n)$ is the transmission range. Constraint (2.3.2) implicitly sets a constraint on routing, i.e., it defines a set of possible candidate nodes (within the transmission range) as the next-hop node.

- For any other link (k, l) that is active on band m at the same time, it is required that $d_{kj} \geq (1 + \Delta)r(n)$, where $(1 + \Delta)r(n)$ represents an interference range, which is intended to keep concurrent transmitting node k farther away from producing non-negligible interference on node j. More formally, we have

$$d_{kj} \geq (1 + \Delta)r(n) \text{ for each link } (k, l) \in \mathcal{E}^m(t) \text{ and } (k, l) \neq (i, j), \tag{2.3.3}$$

  where $\mathcal{E}^m(t)$ is the set of links in the network that are active on band m at time t. (2.3.3) implicitly sets a constraint on scheduling.

Note that under the protocol model, the transmission range $r(n)$ is assumed to be the same at all nodes. Thus, the same (synchronized) transmission power $p(n)$ is used at all nodes.

Under the physical model, each node is allowed to perform *independent* power control. Denote $p_{ij}^m(t)$ the power used by node i to transmit to node j on band $m \in \mathcal{B}_{ij}$ at time t. Such a transmission is successful if and only if the signal-to-interference-and-noise-ratio (SINR) satisfies

$$\text{SINR}_{ij} = \frac{g_{ij} \cdot p_{ij}^m(t)}{\eta + \sum_{(k,l)\in\mathcal{E}^m(t)}^{(k,l)\neq(i,j)} g_{kj} \cdot p_{kl}^m(t)} \geq \beta \ , \tag{2.3.4}$$

where $g_{ij} = d_{ij}^{-\gamma}$ is the channel gain over link (i, j), $\gamma \geq 2$ is the path loss index, η is the ambient noise power, and β is the SINR threshold.

Under the above setting for a CRN, we consider a common throughput $\lambda(n)$ for each node (source) in the network to its randomly selected destination. The goal of capacity scaling law is

to find the maximum $\lambda(n)$ that can be achieved by considering multi-hop routing and time-slotted scheduling.

2.4 Asymptotic Capacity under the Protocol Model

In this section, we analyze the capacity scaling law under the protocol model. As discussed in Section 2.1, the root of the difficulties associated with a CRN is the heterogeneity of available bands among the links. To overcome this difficulty, we create two auxiliary networks ζ and α to facilitate our analysis. In Section 2.4.1, we develop a capacity upper bound with the help of auxiliary network ζ. The lower bound analysis is given in Section 2.4.2, which is more challenging and interesting. The main difficulty here is how to create an appropriate auxiliary network α to facilitate the analysis and how to use this auxiliary network α to obtain a capacity lower bound.

The result of this section can be summarized as follows (see Theorems 2.1 and 2.2).

Under the protocol model, the capacity of a CRN with n nodes is $\lambda(n) \in [\Omega(\frac{C_\alpha}{\sqrt{n \ln n}}), O(\frac{C_\zeta}{\sqrt{n \ln n}})]$ almost surely when $n \to \infty$.

2.4.1 A Capacity Upper Bound

To address the heterogeneity problem in each node's available bands, we propose to construct an auxiliary network to facilitate our analysis. Our goal is to have an auxiliary network with the following two characteristics: (i) homogeneous bands setting, i.e., the sets of available bands on all links are the same; (ii) given a CRN, its capacity is upper bounded by its corresponding auxiliary network. If we can get such an auxiliary network, then we can focus on developing the capacity upper bound for this auxiliary network with homogeneous bands setting, which is also a capacity upper bound for the original CRN.

Such an auxiliary network can be constructed by adding some extra bands to the set of available

bands on each link so that the sets of available bands on all the links are identical. More formally, we have the following definition for this auxiliary network.

Definition 2.1. (Auxiliary network ζ) *An auxiliary network ζ contains all the nodes and links in the original CRN and the same source-destination pairs. The set of available bands on all links in auxiliary network ζ is defined as $\mathcal{B}_\zeta$, where $\mathcal{B}_\zeta = \bigcup_{(i,j)\in\mathcal{E}} B_{ij}$, i.e., the set of all available bands in the network.*

Since the frequency band resource in this auxiliary network is no less than the original CRN, we have the following lemma.

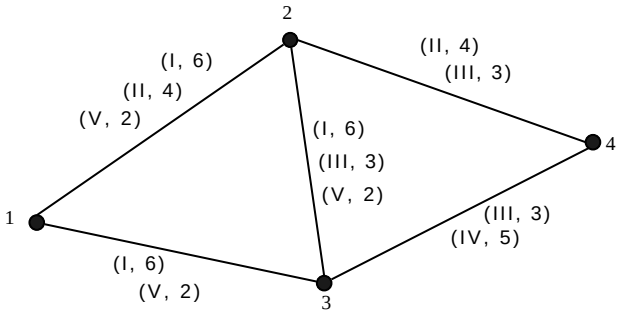
Lemma 2.1. *Under the protocol model, for any given CRN (with heterogeneous available bands on each link), its capacity is no more than the capacity of its corresponding auxiliary network ζ.*

Lemma 2.1 is easily verified since any feasible routing and scheduling solution to a CRN is also feasible to its corresponding auxiliary network ζ. Given the homogeneous auxiliary network ζ, we can now apply the result by Gupta and Kumar [47] and obtain its capacity upper bound as $O(\frac{C_\zeta}{\sqrt{n \ln n}})$, where $C_\zeta = \sum_{m\in\mathcal{B}_\zeta} C_m$. Based on Lemma 2.1 and this capacity upper bound for ζ, we have the following theorem.

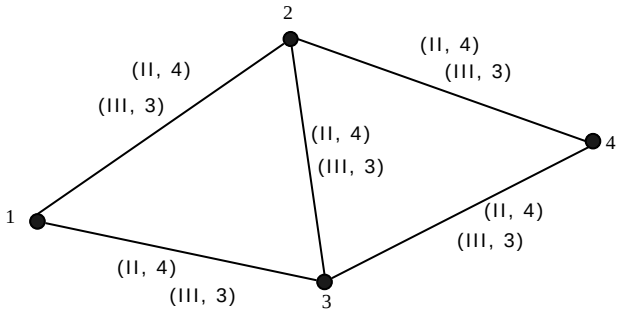
Theorem 2.1. *Under the protocol model, a capacity upper bound for a CRN with n nodes is $O(\frac{C_\zeta}{\sqrt{n \ln n}})$ almost surely when $n \to \infty$.*

2.4.2 A Constructive Lower Bound

The solution to find a lower bound is much more difficult. Note that, to develop a capacity lower bound for a CRN under the protocol model, it is sufficient to construct a feasible solution that satisfies all the constraints (see Section 2.3). This feasible solution can thus serve as a lower bound. But due to the heterogeneity of available bands on different links, band assignment becomes a very difficult problem if we want to construct a feasible solution directly.



(a) A 4-node CRN, where $\mathcal{B}_1 = \{\text{I, II, V}\}$, $\mathcal{B}_2 = \{\text{I, II, III, V}\}$, $\mathcal{B}_3 = \{\text{I, III, IV, V}\}$ and $\mathcal{B}_4 = \{\text{II, III, IV}\}$.



(b) The corresponding auxiliary network α, where the set of available bands at each node is $\{\text{II, III}\}$.

Figure 2.1: An example illustrating how to construct auxiliary network α.

Construct Auxiliary Network α

Recognizing that a direct approach to construct a feasible solution is not likely to be fruitful, we again resort to creating a homogeneous auxiliary network to facilitate our analysis. However, such a homogeneous auxiliary network is not so straightforward this time, as we illustrate in the following naive approach.

A naive approach: Recall that in the development of a capacity upper bound, we define the available bands on each link in the auxiliary network ζ to be the union of available bands on all the links in the network. To get a capacity lower bound, one might consider to define the available bands on each link in the auxiliary network α by taking the intersection of available bands among

all the links in the network. Unfortunately, such a approach may not work as it may lead to an empty set. In this case, we will have a trivial 0 capacity as lower bound, which is not useful. As an example, consider the simple 4-node CRN in Fig. 2.1(a). The available bands for each link are the common bands at two nodes. We show the available bands on each link and corresponding capacity of each band in Fig. 2.1(a). For instance, the bands on link $(1, 2)$ is the intersection of available bands on nodes 1 and 2, which are I, II, and V. The capacity of band I is 6, the capacity of band II is 4, and the capacity of band V is 2, respectively. We can see that although there are some bands on each link, there does not exist a single band that is common on all links.

Our approach: In our approach, we first identify the bottleneck link with the minimum capacity in the original CRN. We then use the set of available bands on the bottleneck link to construct an auxiliary network. We again use the CRN in Fig. 2.1(a) as an example. The bottleneck link in the network is the link between nodes 2 and 4, which has the smallest capacity (i.e., $4 + 3 = 7$) among all the links in the network. To construct the auxiliary network, we retain the same topology structure, except that the set of available bands on each link is now set to the same as that on the bottleneck link in the original CRN (see Fig. 2.1(b)). Although such construction may appear strange at first glance, its benefits will soon become clear when we analyze the lower bound (see Lemma 2.5).

More formally, we define our auxiliary network as follows.

Definition 2.2. (Auxiliary network α) *An auxiliary network α contains all the nodes and links in the original CRN and the same source-destination pairs. The set of available bands on each link in the auxiliary network is defined as $\mathcal{B}_\alpha = \mathcal{B}_{ij}$, where link (i, j) denotes the bottleneck link in the original CRN, i.e., $C_{ij} = \min\{C_{kl}, \forall(k, l) \in \mathcal{E}\}$.*

Comparing link capacities in the original CRN and in the auxiliary network, we have the following property.

Property 2.1. *For the capacity of each link (k, l) in the original CRN, we have $C_{kl} \geq C_\alpha$, where C_α is defined as $\sum_{m\in\mathcal{B}_\alpha} C_m$.*

Note that we do not consider the trivial case where a node does not have any band in common with its neighboring nodes. Such isolated node will lead to a zero capacity.

For the homogeneous auxiliary network α, we can again apply the result in [47] to obtain its capacity lower bound. A capacity lower bound (for auxiliary network α) does not offer us a straightforward solution to a capacity lower bound to the original CRN, which is our main interest. Further, there does not appear to be an obvious connection between the two since a feasible solution to the auxiliary network α is, in general, not feasible to the original CRN.

To derive a capacity lower bound for the original CRN, we propose a two-step approach: (1) find a feasible solution for auxiliary network α; *and* (2) transform this solution to a feasible solution to the original CRN. We now elaborate each step in the rest of this section.

Find a Feasible Solution for Auxiliary Network α

We now construct a feasible solution ψ_α for auxiliary network α. In this solution, we need to design both a routing scheme and a scheduling scheme. Our routing scheme follows a cell-based approach. That is, we divide the unit square area into small cells such that there is at least one node in each cell almost surely when $n \to \infty$ (see Lemma 2.3). For each source-destination pair, the straight line that connects them will pass through a number of cells. One node is chosen from each of these cells as a relay for multi-hop routing.

To avoid interference, scheduling must be performed along with cell-based routing. We can build a conflict graph to model interference relationships among the links, with each vertex representing a link in the auxiliary network and each edge between two vertices representing a conflict. We analyze the maximum number of conflict links to a link (or the maximum degree in the conflict graph) and then apply the result for graph coloring to obtain a scheduling scheme. Afterward, we obtain a complete solution with the following throughput.

Lemma 2.2. *Under the protocol model, we can construct a feasible solution ψ_α to auxiliary network α with a throughput of $\lambda_\alpha(n) = \Omega(\frac{C_\alpha}{\sqrt{n \ln n}})$.*

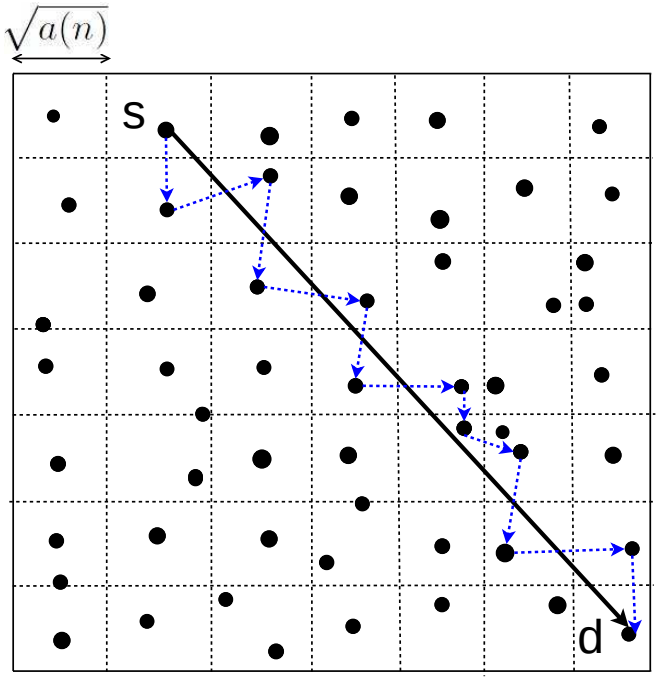


Figure 2.2: Multi-hop routing from a source node s to its destination node d.

Now we give details for routing and scheduling in the above discussion.

Routing Scheme: We develop a cell-based routing scheme. We divide the unit square into small cells of square size $a(n) = \sqrt{\frac{n}{\ln n}}$ x$\sqrt{\frac{n}{\ln n}} = \frac{\ln n}{n}$ and set the transmission range $r(n) = \sqrt{5a(n)}$ so that any node in one cell can transmit to another node in its four neighboring cells. Now, we draw a line to connect each source-destination pair, which passes through some cells. One node is chosen from each of these cells to relay the traffic from the source node to its destination. See Fig. 2.2 as an example.

Such a routing scheme requires at least one node in each cell. We call a cell without any node an "empty" cell. We have the following lemma.The proof is similar to the proof of Claim 3.1 in [66] and is omitted here.

Lemma 2.3. *For auxiliary network α with n nodes and the cell size of $a(n) = \frac{\ln n}{n}$, there is no empty cell almost surely when $n \to \infty$. That is, the routing scheme is feasible almost surely when*

$n \to \infty$.

Scheduling Scheme: In our solution, we design the same scheduling scheme for all bands, i.e., once a link is active, it treats all bands in $\mathcal{B}_\alpha$ as one aggregate band. We consider time-slot based scheduling, i.e., we divide one time frame into multiple time slots to satisfy the protocol model scheduling constraint (2.3.3).

To analyze the performance of our scheduling scheme, we need to know the number of conflicting links for each link (i, j), which directly affects the number of time slots for scheduling and capacity. Two links are in conflict if they cannot be active on the same band at the same time. Then, we have the following lemma on the number of conflicting links for any link in the network. This lemma can be proved directly by combining Lemmas 2 and 3 in [37], and is omitted here.

Lemma 2.4. *Under our routing scheme for auxiliary network α with n nodes, the number of conflicting links for any link is upper bounded by $O(n\sqrt{a(n)})$ almost surely when $n \to \infty$.*

To schedule these conflict links, we use a conflict graph to model them. Each link in the auxiliary network corresponds to a vertex in the conflict graph, and any conflict in the auxiliary network is represented by an edge connecting two corresponding vertices in the conflict graph. If we use different vertex color to represent each time slot, then the scheduling problem reduces to the well-studied vertex-color problem. Using Lemma 2.4, the degree of each vertex in the conflict graph will be at most $c_2 \cdot n\sqrt{a(n)}$ for some constant c_2 almost surely when $n \to \infty$. Then the required number of colors is at most $1+c_2 \cdot n\sqrt{a(n)}$ [126]. So we can divide one time frame into at most $1 + c_2 \cdot n\sqrt{a(n)}$ equal length time slots for scheduling. Therefore, the achievable throughput $\lambda_\alpha(n)$ is given by

$$\lambda_\alpha(n) \;\geq\; \frac{C_\alpha}{1 + c_2 \cdot n\sqrt{a(n)}} = \frac{C_\alpha}{1 + c_2\sqrt{n \ln n}} = \Omega\left(\frac{C_\alpha}{\sqrt{n \ln n}}\right) ,$$

where the second equality holds because $a(n) = \frac{\ln n}{n}$.

Obtain a Feasible Solution to the Original CRN

In the last step, we find a feasible solution ψ_α to the auxiliary network that has a throughput of $\lambda_\alpha(n) = \Omega(\frac{C_\alpha}{\sqrt{n \ln n}})$. Based on ψ_α, we now define a feasible solution ψ to the original CRN. In ψ, for routing, we follow the same routing scheme as that in ψ_α. For scheduling, if a link (i, j) is active on all bands in $\mathcal{B}_\alpha$ in a time slot under solution ψ_α (recall that in ψ_α, once a link is active, it uses all bands in $\mathcal{B}_\alpha$), then we let this link be active on all bands in $\mathcal{B}_{ij}$ in the same time slot under solution ψ. Given that each link has more diverse bands in the original CRN than in α, solution ψ constructed in this manner is also feasible. This result is stated in the following lemma.

Lemma 2.5. *The constructed ψ is a feasible solution to the original CRN under the protocol model and its throughput $\lambda(n) = \lambda_\alpha(n)$.*

Proof. To show ψ is feasible, we need to show that constraints (2.3.1), (2.3.2), and (2.3.3) hold under ψ. Based on the construction of ψ, it is clear that band assignment constraint (2.3.1) holds. Furthermore, routing constraints (2.3.2) also holds under ψ by the construction of ψ and the feasibility of ψ_α. Before we show that scheduling constraint (2.3.3) holds under ψ, we need to analyze the relationship between $\mathcal{E}^m_\psi(t)$ (the set of active links on a band m at time t under ψ) and $\mathcal{E}_{\psi_\alpha}(t)$ (the set of active links at time t under ψ_α). Note that $\mathcal{E}^m_\psi(t)$ is different for different band m. But under ψ_α, once a link is active, it uses all bands in $\mathcal{B}_\alpha$. Thus, $\mathcal{E}_{\psi_\alpha}(t)$ does not depend on bands. Based on the construction of ψ, we have that a link $(i, j) \in \mathcal{E}^m_\psi(t)$ if and only if $m \in \mathcal{B}_{ij}$ and under ψ_α, link (i, j) is active at time t under $\mathcal{B}_\alpha$. Thus, we have

$$\mathcal{E}^m_\psi(t) = \{(i,j) : m \in \mathcal{B}_{ij}, (i,j) \in \mathcal{E}_{\psi_\alpha}(t)\} \subseteq \mathcal{E}_{\psi_\alpha}(t). \tag{2.4.1}$$

Now we are able to show that (2.3.3) also holds under ψ. Consider a link $(i, j) \in \mathcal{E}^m_\psi(t)$. It is clear from (2.4.1) that any other link $(k, l) \in \mathcal{E}^m_\psi(t)$ is also in $\mathcal{E}_{\psi_\alpha}(t)$, i.e., a link that is active on band m at time t under ψ is also active at time t under ψ_α. Then, since (2.3.3) holds under ψ_α, (2.3.3) should also hold under ψ. Thus, solution ψ is feasible.

Finally, we analyze the throughput achieved by ψ. For each link (i, j), since we have $C_{ij} \geq C_\alpha$ by Property 2.1, the achieved capacity on each link (i, j) under ψ is no less than that under ψ_α.

Since link capacity constraint holds on each link under ψ_α and the same routing scheme is used in ψ, link capacity constraint also holds on each link under ψ. Thus, the same throughput $\lambda(n) = \lambda_\alpha(n)$ is achieved by ψ. □

Combining Lemmas 2.2 and 2.5, we have the following capacity lower bound for a CRN.

Theorem 2.2. *Under the protocol model, the capacity of a CRN is* $\lambda(n) = \Omega(\frac{C_\alpha}{\sqrt{n \ln n}})$ *almost surely when* $n \to \infty$.

2.5 Asymptotic Capacity under the Physical Model

In this section, we analyze the capacity scaling law under the physical model. We consider both the special case where nodes are allowed to perform synchronized power control and the general case where each node is allowed to perform independent power control. For the spacial case, the result can be summarized as follows (also see Theorems 2.3 and 2.5).

Under the physical model with synchronized power control at all nodes, the capacity of a CRN with n *nodes is* $\lambda(n) \in [\Omega(\frac{C_\alpha}{\sqrt{n \ln n}}), O(\frac{C_\zeta}{\sqrt{n}})]$ *almost surely when* $n \to \infty$.

For the general case, the result can be summarized as follows (also see Theorems 2.4 and 2.5).

Under the physical model with independent power control at each node, the capacity of a CRN with n *nodes is* $\lambda(n) \in [\Omega(\frac{C_\alpha}{\sqrt{n \ln n}}), O(\frac{C_\zeta}{n^{1/\gamma}})]$ *almost surely when* $n \to \infty$, *where* γ *is the path loss index.*

2.5.1 Finding An Upper Bound

We again employ auxiliary network ζ (see Definition 2.1) to facilitate our analysis. Similar to Lemma 2.1 for the protocol model, we have the following lemma for the physical model.

Lemma 2.6. *Under the physical model, for any given CRN, its capacity is no more than the capacity of its corresponding auxiliary network ζ, regardless of whether independent or synchronized power control at each node.*

For the homogeneous auxiliary network ζ with synchronized power control, we can apply the results in [47] and obtain its capacity upper bound as $O(\frac{C_\zeta}{\sqrt{n}})$. Then based on Lemma 2.6, we have the following capacity upper bound for the original CRN.

Theorem 2.3. *Under the physical model with synchronized power control at all nodes, a capacity upper bound for a CRN with n nodes is $O(\frac{C_\zeta}{\sqrt{n}})$ almost surely when $n \to \infty$.*

For the general case with independent power control on each node, we have the following lemma for auxiliary network ζ. The proof of this lemma is similar to the proof of Theorem 2.1 in [47], and is omitted here.

Lemma 2.7. *Under the physical model with independent power control at each node, the capacity of auxiliary network ζ is $\lambda_\zeta(n) = O(\frac{C_\zeta}{n^{1/\gamma}})$ almost surely when $n \to \infty$.*

Combining Lemmas 2.6 and 2.7, we have the following theorem.

Theorem 2.4. *Under the physical model with independent power control on each node, the capacity of a CRN is $\lambda(n) = O(\frac{C_\zeta}{n^{1/\gamma}})$ almost surely when $n \to \infty$.*

2.5.2 Construction of A Lower Bound

Similar to the case under the protocol model, we employ our auxiliary network α (see Definition 2.2) and follow the two-step approach (as we did in Section 2.4.2) to construct a lower bound, although the details are different.

Step 1: A Feasible Solution for Auxiliary Network α. We now need to construct a feasible solution for auxiliary network α. However, constructing a feasible solution for the physical model

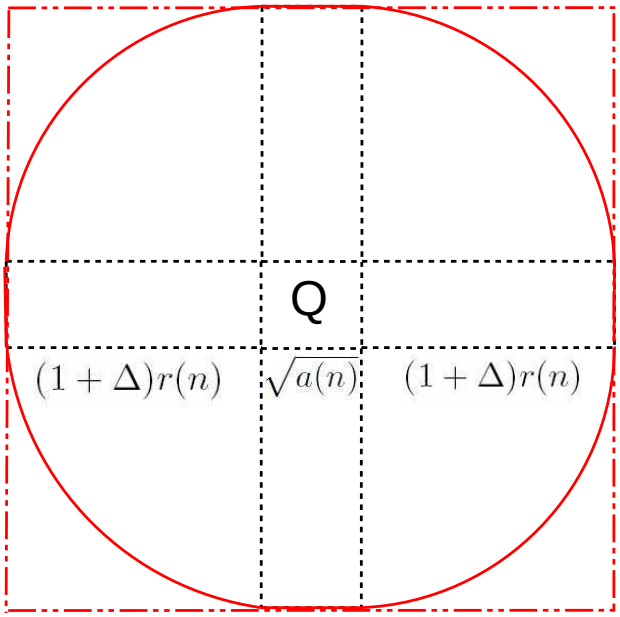


Figure 2.3: The area to cover all the transmitting nodes that may interfere with the receivers in cell Q.

is more difficult than that for the protocol model. This is because that under the protocol model, we only need to consider distance in scheduling constraint (2.3.3). But under the physical model, we also need to consider power control in the SINR constraint (2.3.4).

Given that we have developed a feasible solution ψ_α for the protocol model, we hope to use this result and develop a feasible solution for the physical model. We observe that if we set the parameter Δ in the protocol model "large enough", then ψ_α is also a feasible solution for the physical model. This is because a large Δ will impose more constraints on conflict graph, and thus reduce interference from neighboring nodes in the protocol model solution. As a result, the SINR at a receiver can be made large enough to satisfy the physical model SINR constraint (2.3.4). In this case, ψ_α is also a feasible solution for the physical model. This insight is the basis of the following lemma.

Lemma 2.8. *Under the physical model, we can construct a feasible solution ψ_α to auxiliary network α with a capacity of $\lambda_\alpha(n) = \Omega(\frac{C_\alpha}{\sqrt{n \ln n}})$.*

Proof. In this proof, we show that if we set

$$\Delta \geq \left\{ \frac{p(n)\beta}{2r(n)^2[p(n) - \beta\eta r(n)^\gamma]} \right\}^{\frac{1}{2+\gamma}} - 1 \ , \tag{2.5.1}$$

where $p(n)$ is the transmission power used in the protocol model solution ψ_α, then the solution ψ_α developed for the protocol model (in Section 2.4.2) is also feasible under the physical model.

First, we show that the physical model SINR constraint (2.3.4) is satisfied under ψ_α. Based on the construction of ψ_α in Section 2.4.2, we know that once a link (i, j) is active, other nodes within a square with side length $2(1 + \Delta)r(n) + \sqrt{a(n)}$, which is centered on the cell where node j is located, cannot transmit (see Fig. 2.3). Note that node j may be located on the boundary of the unit square. Thus, the number of links that make interference to link (i, j) is at most

$$\begin{aligned}
& \left\lfloor \frac{2(1+\Delta)r(n)+1}{2(1+\Delta)r(n)+\sqrt{a(n)}} \right\rfloor^2 - 1 \\
\leq \ & \left\lceil \frac{2(1+\Delta)r(n)+1}{2(1+\Delta)r(n)+\sqrt{a(n)}} \right\rceil^2 - 1 \\
= \ & \frac{[1-\sqrt{a(n)}][4(1+\Delta)r(n)+\sqrt{a(n)}+1]}{[2(1+\Delta)r(n)+\sqrt{a(n)}]^2} \\
< \ & \frac{2}{[2(1+\Delta)r(n)+\sqrt{a(n)}]^2} \\
< \ & \frac{2}{4[(1+\Delta)r(n)]^2} = \frac{1}{2[(1+\Delta)r(n)]^2} \ ,
\end{aligned}$$

where the third inequality holds by $1 - \sqrt{a(n)} < 1$ and $4(1 + \Delta)r(n) + \sqrt{a(n)} \to 0 < 1$ when $n \to \infty$. Based on the construction of ψ_α, the interference at receiving node j from each of these links will be at most $\frac{p(n)}{[(1+\Delta)r(n)]^\gamma}$. Therefore, we have

$$\begin{aligned}
\text{SINR}_{ij} \ &> \ \frac{\frac{p(n)}{[r(n)]^\gamma}}{\eta + \frac{1}{2[(1+\Delta)r(n)]^2}\frac{p(n)}{[(1+\Delta)r(n)]^\gamma}} \\
&= \ \frac{r(n)^{-\gamma}}{\eta p(n)^{-1} + \frac{1}{2}[(1+\Delta)r(n)]^{-(2+\gamma)}} \\
&\geq \ \frac{r(n)^{-\gamma}}{\eta p(n)^{-1} + [\beta^{-1}r(n)^{-\gamma} - \eta p(n)^{-1}]} = \beta \ ,
\end{aligned}$$

where the third inequality holds by (2.5.1). Thus, under ψ_α, for each link (i,j) that is active on a band $m \in \mathcal{B}_{ij}$ at time t, the SINR constraint (2.3.4) holds. Therefore, the power control and scheduling schemes in ψ_α are also feasible under the physical model. Furthermore, the achieved capacity on each link under the physical model is equal to that under the protocol model. As a result, link capacity constraint is still satisfied on each link under the physical model and the achieved throughput under the physical model will be the same as that under the protocol model, which is $\Omega(\frac{C_\alpha}{\sqrt{n \ln n}})$ almost surely when $n \to \infty$.

In this solution, the same transmission power $p(n)$ is used at all nodes (i.e., synchronized power control). Note that a feasible solution under synchronized power control is also a feasible solution under independent power control. Thus, under the physical model, a lower bound on the capacity of auxiliary network α is $\lambda_\alpha(n) = \Omega(\frac{C_\alpha}{\sqrt{n \ln n}})$ almost surely when $n \to \infty$ for both independent power control and synchronized power control. □

Step 2: Obtain a Feasible Solution to the Original CRN. We now need to transform the feasible solution constructed in Step 1 for auxiliary network α to a feasible solution to the original CRN. This transformation is similar to that in Section 2.4.2, except that we need to verify that the constructed solution is feasible for the physical model, i.e., the SINR constraint (2.3.4). We state our result in the following lemma.

Lemma 2.9. *Under the physical model, we can construct a feasible solution ψ to the original CRN with a throughput $\lambda(n) = \lambda_\alpha(n)$.*

Proof. We construct ψ by using the same routing scheme as that in ψ_α. The power control and scheduling schemes in ψ are constructed as follows. Let each link (i,j) be active on all bands in $\mathcal{B}_{ij}$ at time t whenever it is active on all bands in $\mathcal{B}_\alpha$ under ψ_α. Also, we set the transmission power $p(n)$ as that in ψ_α.

For a link (i,j) that is active on a band m at time t under ψ, its SINR is

$$\begin{aligned}\text{SINR}_{ij}(\psi) &= \frac{g_{ij}\cdot p(n)}{\eta+\sum_{(k,l)\in\mathcal{E}_{\psi}^{m}(t)}^{(k,l)\neq(i,j)} g_{kj}\cdot p(n)} \\ &\geq \frac{g_{ij}\cdot p(n)}{\eta+\sum_{(k,l)\in\mathcal{E}_{\psi_\alpha}(t)}^{(k,l)\neq(i,j)} g_{kj}\cdot p(n)} \\ &= \text{SINR}_{ij}(\psi_\alpha)\geq\beta\ ,\end{aligned}$$

where the second inequality holds because $\mathcal{E}_{\psi}^{m}(t)\subseteq\mathcal{E}_{\psi_\alpha}(t)$ in (2.4.1), and the last inequality holds because ψ_α is feasible. Thus, under ψ, for each link (i,j) that is active on a band $m\in\mathcal{B}_{ij}$ at time t, the SINR constraint (2.3.4) holds. Therefore, the power control and scheduling schemes in ψ are feasible.

For each link (i,j), since we have $C_{ij}\geq C_\alpha$ in Property 2.1, the total achieved link capacity under ψ is no less than that under ψ_α. Since the same routing scheme is used, this routing scheme in ψ is also feasible. Thus, ψ is a feasible solution with a throughput $\lambda_\alpha(n)$. Again, since the same transmission power $p(n)$ is used at all nodes, this solution is a feasible solution under both synchronized power control and independent power control. Thus, for the physical model, a lower bound for the capacity of a CRN is $\lambda(n)=\lambda_\alpha(n)$ under both independent power control and synchronized power control. □

Based on Lemmas 2.8 and 2.9, we have the following theorem on lower bound of a CRN capacity.

Theorem 2.5. *Under the physical model, the capacity of a CRN is* $\lambda(n)=\Omega(\frac{C_\alpha}{\sqrt{n\ln n}})$ *almost surely when* $n\to\infty$.

2.6 Chapter Summary

In this chapter, we studied capacity scaling laws for cognitive radio ad hoc networks (CRNs) under both the protocol and the physical models, respectively. The main novelties in this work are: (1)

the design of suitable auxiliary networks to analyze a network with heterogeneous bands, and (2) the analysis of relationship between solutions under the auxiliary networks and solutions to the original CRNs. Although our results for CRN share similar form to those results by Gupta-Kumar for single-channel single-radio network, our problem is much more challenging and the approach is far beyond a simple extension of that in [47]. The results in this chapter fill an important gap in the fundamental understanding of CRNs.

Chapter 3

Capacity Scaling Laws for MIMO Ad Hoc Networks

3.1 Introduction

By employing multiple antennas at both the transmitter and receiver, MIMO has brought significant benefits to wireless communications, such as increased link capacity [18, 33, 113], improved link diversity [139], and interference cancellation between conflicting links [19, 107]. Although there has been extensive work on MIMO at the physical and link layers, there is limited work on MIMO at the network layer (i.e., multi-hop MIMO network), particularly results on capacity scaling laws. The analysis of capacity scaling law studies how the achievable throughput of each node scales as the number of nodes in the network increases. Such investigation is considered critical to understand the fundamental behavior of large-sized networks. Capacity scaling law was first studied by Gupta and Kumar [47] on single-antenna ad hoc networks. Subsequently, the research community has extended this seminal work to other types of wireless networks, such as multi-channel multi-radio (MC-MR) ad hoc networks [8, 68], ultra-wide band ad hoc networks [90, 138], and cognitive radio networks [55, 135].

However, to date, there is very limited work [11, 15] on capacity scaling laws for MIMO ad hoc networks. In [11], Bolcskei *et al.* considered a MIMO source and destination pair, assisted by a set of relay nodes, and studied how the capacity between this source-destination pair scales with respect to the number of relay nodes. In [15], Chen and Gans studied the capacity of a MIMO ad hoc network with a set of simultaneous one-hop source-destination pairs. In this work, routing is not considered due to one-hop communications. Apart from these results, capacity scaling laws for multi-hop MIMO ad hoc networks remain unexplored.

In this chapter, we aim to characterize asymptotic capacity for multi-hop MIMO ad hoc networks [56]. Although there are many schemes to exploit the benefits of antenna arrays at a node, we focus on the so-called *zero-forcing beamforming* (ZFBF) scheme [19, 107], which captures the two key characteristics of MIMO: *spatial multiplexing* and *interference cancellation.* For asymptotic study, we analyze both the lower bound and the upper bound. We show that although a capacity lower bound can be obtained by extending the work of Gupta and Kumar [47], a tight capacity upper bound is a much harder problem. We propose to partition the network area into small squares cleverly so that the maximum data rate that can be received by the nodes inside the small square can be computed exactly. By taking the sum of data rates from all small squares, we can obtain the maximum data rate the whole network can support. Based on this result, we develop a tight capacity upper bound for our problem. Our main result in this chapter is the following: for a MIMO network with n randomly located nodes, each equipped with γ antennas and a rate of W on each data stream, we show that the capacity upper and lower bounds have the same order, and the achievable throughput of each node is $\Theta(\frac{\gamma W}{\sqrt{n \ln n}})$.

The remainder of this chapter is organized as follows. In Section 3.2, we present a model for MIMO network that will be used in our asymptotic capacity study. In Section 3.3, we analyze asymptotic capacity bounds. Section 3.4 presents some numerical results. Section 3.5 concludes this chapter.

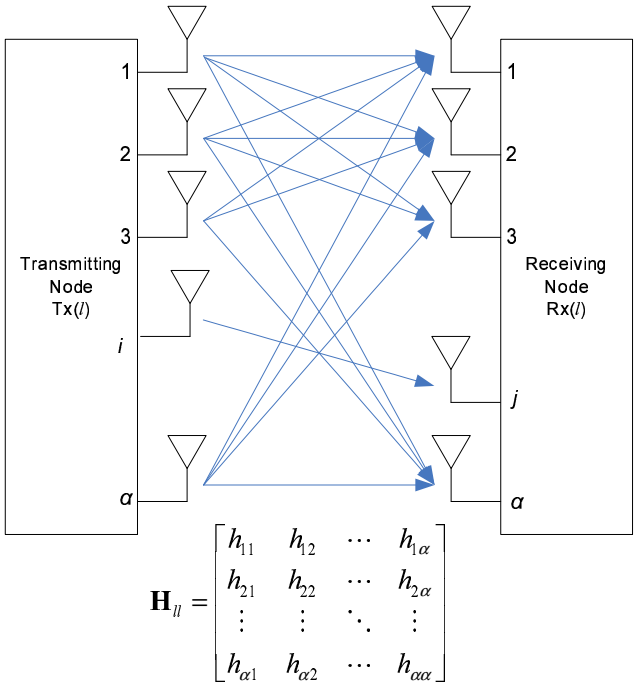


Figure 3.1: A spatial multiplexing link.

3.2 MIMO Network Modeling

In this section, we present a model for MIMO ad hoc networks which we will use in our analysis of asymptotic capacity. This model captures MIMO's spatial multiplexing and interference cancellation capabilities at the physical layer.

3.2.1 Spatial Multiplexing and Interference Cancellation

There are a number of mechanisms at the physical layer to enable spatial multiplexing and interference cancellation, such as V-Blast (Vertical-Bell labs layered space time) [33], ZFBF (zero-forcing beamforming) [19, 107], DPC (dirty chapter coding) [125], among others. Spatial multiplexing refers that a transmitter can send several independent data streams to its intended receiver simultaneously on a link. Interference cancellation refers that by properly devising the transmission and

Table 3.1: Notation in Chapter 3.

Symbol	Definition
D	The mean distance between each node and its destination
$d_{i,j}$	The distance between node i and j
$\mathbf{H}_{lk}$	The channel coefficient matrix between nodes Tx(l) and Rx(k)
n	The number of nodes in the network
Rx(l)	The receiver of link l
$r(n)$	The common transmission range of all nodes
S_{li}	The signal of data stream i on link l
$\hat{S}_{li}$	The recovered signal of data stream i on link l
Tx(l)	The transmitter of link l
$UB(n)$	Capacity upper bound
$\mathbf{u}_{li}$	The transmission vector for transmitting signal S_{li}
$\mathbf{v}_{li}$	The reception vector for receiving signal S_{li}
W	The maximum data rate that a single data stream can support
z_l	The number of active data streams on link l
Δ	The parameter to set the interference range
$\lambda(n)$	The per-node throughput of a random multi-hop MIMO ad hoc network with n nodes
γ	The number of antennas at each node
$\pi(\cdot)$	An ordering function of all nodes in the network

reception vectors, the interference between several conflicting links can be cancelled out. In this chapter, for the ease of the mathematic modeling, we employ the simple but yet powerful ZFBF for spatial multiplexing and interference cancellation. Table 3.1 lists the notations in this chapter.

Spatial multiplexing. We assume that the number of antennas at each node is γ and the network is deployed in a rich scattering environment, so that the degree of freedom (DoF) at each node is approximately equal to its number of antennas (γ). Suppose we want to activate z_l data streams on a link l (see Fig. 3.1). Denote Tx(l) and Rx(l) the transmitter and receiver of link l, respectively. Denote S_{li} the signal of data stream i ($1 \leq i \leq z_l$) at transmitter Tx(l) and $\mathbf{u}_{li}$ the $\gamma \times 1$ transmission vector of signal S_{li}, respectively. Denote $\mathbf{H}_{ll}$ the channel coefficient matrix between

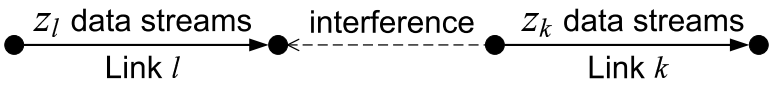


Figure 3.2: Two interfering MIMO links.

nodes Tx(l) and Rx(l). Although there are different MIMO channel models in the literature, e.g., MIMO Rician [61] and Rayleigh fading channels [19, 15], our results do not depend on the specific channel models. We assume that our channel matrix remains constant during a certain transmission period. Moreover, we assume that the channel matrix is of full rank, which is justifiable under rich scattering environment. Thus, during a transmission period, channel matrix $\mathbf{H}$ is regarded as a full rank $\gamma \times \gamma$ constant matrix. Note that although the channel matrix may change for different transmission periods, the asymptotic results (Section 3.3) will still hold under each transmission period, i.e., our results remain valid over time.

To transmit all z_l data streams, Tx(l) sends the combined signal $\sum_{j=1}^{z_l} \mathbf{u}_{lj} S_{lj}$ through its γ antennas. The signal at receiver Rx(l)'s antennas will be $(\sum_{j=1}^{z_l} \mathbf{u}_{lj} S_{lj})^T \mathbf{H}_{ll}$. Receiver Rx$(l)$ uses an $\gamma \times 1$ reception vector $\mathbf{v}_{li}$ to recover the signal of data stream i. Then the recovered signal $\hat{S}_{li}$ for data stream i is $\hat{S}_{li} = (\sum_{j=1}^{z_l} \mathbf{u}_{lj} S_{lj})^T \mathbf{H}_{ll} \mathbf{v}_{li} = \mathbf{u}_{li}^T \mathbf{H}_{ll} \mathbf{v}_{li} S_{li} + \sum_{1 \le j \le z_l}^{j \ne i} \mathbf{u}_{lj}^T \mathbf{H}_{ll} \mathbf{v}_{li} S_{lj}$. By choosing appropriate $\mathbf{u}$ and $\mathbf{v}$, we can ensure that the recovered signal $\hat{S}_{li}$ achieves a unit gain ($\mathbf{u}_{li}^T \mathbf{H}_{ll} \mathbf{v}_{li} = 1$) and zero interference ($\mathbf{u}_{lj}^T \mathbf{H}_{ll} \mathbf{v}_{li} = 0, j \ne i$) such that the data stream i can be successfully recovered. Thus, we have the following constraints to make all z_l data streams successful.

$$\mathbf{u}_{li}^T \mathbf{H}_{ll} \mathbf{v}_{li} = 1 \qquad (1 \le i \le z_l)\,, \tag{3.2.1}$$

$$\mathbf{u}_{lj}^T \mathbf{H}_{ll} \mathbf{v}_{li} = 0 \qquad (1 \le i, j \le z_l, j \ne i)\,. \tag{3.2.2}$$

Interference cancellation. Consider two links k and l and assume that the transmission on link k will interfere with the reception at link l (see Fig. 3.2). Suppose that we want to have z_k data streams on link k and z_l data streams on link l. Denote $\mathbf{H}_{kl}$ the channel coefficient matrix between nodes Tx(k) and Rx(l). To ensure these simultaneous transmission of data streams possible, we must satisfy both the spatial multiplexing constraints in (3.2.1) and (3.2.2) for each link and the

following interference cancellation constraints.

$$\mathbf{u}_{ki}^{T}\mathbf{H}_{kl}\mathbf{v}_{lj} = 0 \quad (1 \leq i \leq z_k, 1 \leq j \leq z_l) \ .$$

The above constraints guarantee that the interference coming from each data stream of node Tx(k) is cancelled out for each data stream at node Rx(l).

3.2.2 Mathematical Modeling

In this chapter, we consider a random multi-hop MIMO ad hoc network with n nodes, where each node, equipped with γ antennas, is randomly located in a unit square area. Each node acts as a source node and transmits data to a randomly chosen destination node. The per-node throughput $\lambda(n)$ is defined as the minimum data rate that can be sent from each source to its destination via multi-hop routing. Our goal is to find the maximum asymptotic per-node throughput $\lambda(n)$.

We represent the random multi-hop MIMO ad hoc network by a directed graph, denoted by $\mathcal{G} = \{\mathcal{N}, \mathcal{L}\}$, where $\mathcal{N}$ and $\mathcal{L}$ are the set of nodes and all possible MIMO links, respectively. We use the so-called protocol model [47] to deal with interference in the network. Under protocol model, each node in the network has a transmission range $r(n)$ and a node can only transmit data to the nodes within its transmission range. An ordered node pair (i, j) is said to be a link if node j is within i's transmission range, i.e., $\mathcal{L} = \{(i, j) : d_{i,j} \leq r(n), i, j \in \mathcal{N}, i \neq j\}$, where $d_{i,j}$ is the distance. Likewise, each node also has an interference range $(1 + \Delta)r(n)$ and when a node is transmitting, the other nodes (other than its intended receiver) within its interference range cannot be receiving data at the same time, where Δ is a non-negative constant. Under this protocol model, it has been shown by Gupta and Kumar in [47] that we need to set the transmission range $r(n) > \sqrt{\frac{\ln n}{\pi n}}$ to maintain network connectivity with high probability when $n \to \infty$. Due to the use of protocol model and our goal of keeping our analysis tractable, the impact of fading channel on network connectivity [27] is not considered in this study.

If one link is active and no interference cancellation scheme is used in the network, it will interfere with all its nearby links whose receiver are within the interference range of the transmitting

node. When no interference cancellation is employed, denote $\mathcal{I}_l^+$ the set of links that are interfered by link $l \in \mathcal{L}$ and $\mathcal{I}_l^-$ the set of links that interfere with link $l \in \mathcal{L}$.

MIMO physical layer. In this study, we focus on spatial multiplexing and interference cancellation to characterize MIMO physical layer behavior. Further, we employ DoF to represent MIMO resources at a node. A detailed discussion of DoF allocation for spatial multiplexing and interference cancellation is given in [104]. Simply put, when there is no interference, we need to allocate z_l DoFs at both transmitter $\text{Tx}(l)$ and receiver $\text{Rx}(l)$ to achieve z_l data streams on link l. When interference is present in the network, it is necessary to have an ordered list for all nodes and allocate DoFs sequentially to achieve interference cancellation. Denote $\pi(\cdot)$ the mapping between a node and its order in the node list. For two links l and k with z_l and z_k data streams on each link, $l \in \mathcal{L}, k \in \mathcal{I}_l^+$, we know that if $\pi(\text{Tx}(l)) > \pi(\text{Rx}(k))$ (i.e., node Tx(l) is after node $\text{Rx}(k)$ in the ordered node list), then node $\text{Tx}(l)$ will be responsible for cancelling the interference from l to k and will thus consume z_k DoFs; if $\pi(\text{Tx}(l)) < \pi(\text{Rx}(k))$ (i.e., node $\text{Tx}(l)$ is before node $\text{Rx}(k)$ in the ordered node list), then node $\text{Rx}(k)$ will be responsible for cancelling the interference from l to k and will consume z_l DoFs. Then, a link l can support z_l active data streams in the network if and only if the following two constraints are satisfied.

1. DoF constraint at transmitter $\text{Tx}(l)$: The total number of DoFs transmitter $\text{Tx}(l)$ uses for spatial multiplexing and interference cancellation cannot exceed the number of available DoFs at node $\text{Tx}(l)$, i.e.,
$$z_l + \sum_{k \in \mathcal{I}_l^+}^{\pi(\text{Tx}(l)) > \pi(\text{Rx}(k))} z_k \leq \gamma. \tag{3.2.3}$$
This constraint shows that the DoF consumption at transmitter $\text{Tx}(l)$ includes two parts: spatial multiplexing (first term on the LHS) and interference cancellation (second term on the LHS). The total DoF allocation at transmitter $\text{Tx}(l)$ cannot exceed the total DoFs.

2. DoF constraint at receiver $\text{Rx}(l)$: The total number of DoFs that transmitter $\text{Rx}(l)$ allocates for spatial multiplexing and interference cancellation cannot exceed the number of available

DoFs at node $\text{Rx}(l)$, i.e.,

$$z_l + \sum_{k \in \mathcal{I}_l^-}^{\pi(\text{Rx}(l)) > \pi(\text{Tx}(k))} z_k \leq \gamma. \tag{3.2.4}$$

Routing and scheduling. In this chapter, we assume that a node's transmitter is limited to a transmission range $r(n)$. When a source node cannot transmit data to its destination node in one hop, multi-hop routing is needed to relay the data. To avoid potential interference among active links, we employ TDMA to schedule conflict links into different time slots.

3.3 Asymptotic Capacity Bounds

In this section, we analyze the asymptotic capacity bounds for multi-hop MIMO ad hoc networks and the main result of this section is summarized as follows.

The capacity of a random multi-hop MIMO ad hoc network with n nodes is $\lambda(n) = \Theta(\frac{\gamma W}{\sqrt{n \ln n}})$ with high probability when $n \to \infty$.

The above capacity bound is determined by finding a capacity lower bound and a capacity upper bound and showing that they have the same order.

Lower bound analysis. For the capacity lower bound, it is only necessary to find a feasible routing and scheduling scheme for the underlying network. For this purpose, we can simply consider the trivial case where all DoF resource at a node is allocated for spatial multiplexing (i.e., no DoF will be used for interference cancellation). That is, when a link is active, it will use all its DoFs at the transmitter and the receiver to carry γ data streams on that link. This simple case corresponds to scaling the capacity lower bound for a single-antenna ad hoc network by a factor of γ. In [47], Gupta and Kumar showed that a capacity lower bound for a single-antenna ad hoc networks is $\Omega(\frac{W}{\sqrt{n \ln n}})$ by constructing a feasible routing and scheduling scheme. Thus, by adopting the same routing and scheduling scheme in our MIMO ad hoc networks as in [47], a capacity lower bound of $\Omega(\frac{\gamma W}{\sqrt{n \ln n}})$ can be obtained.

Upper bound analysis. The capacity upper bound analysis is more challenging and is the main contribution of this chapter. Note that here we cannot simply scale the capacity upper bound result in [47] by a factor of γ by just considering spatial multiplexing and neglecting interference cancellation. Doing so will neglect the potential increase of network capacity by joint consideration of using DoFs for spatial multiplexing and interference cancellation. Therefore, we must consider *both* spatial multiplexing and interference cancellation for upper bound analysis. As a result, the upper bound analysis in [47] cannot be extended over here.

To derive an upper bound for multi-hop MIMO ad hoc networks, we propose a novel partitioning method on the network area. We partition the unit square for the network into small squares, with the size of each small square being cleverly chosen so that the maximum data rate that can be received by the nodes inside the small square can be computed exactly. By taking the sum of data rates from all small squares, we can obtain the maximum data rate the whole network can support. Based on this result, we can further derive a capacity upper bound for the entire network.

Now the key problem in our analysis is how to set the size of each small square. This is because in a MIMO network, several receiving nodes can be active within close vicinity by using interference cancellation. If the size of each small square is set too large, then the maximum number of data streams that can be received by the nodes inside the square cannot be computed exactly. On the other hand, if the size of each small square is set too small, then the maximum number of data streams that can be received by the nodes inside the square is likely to be overestimated, leading to a loose upper bound. We show that when the length of each side of small square is set to $1/\lceil\frac{\sqrt{2}}{\Delta\cdot r(n)}\rceil$, we can precisely determine the maximum number of data streams that can be received by nodes inside the small square, regardless of the number of receiving nodes in the square. We formally state this result in the following lemma.

Lemma 3.1. *For a square with side length $1/\lceil\frac{\sqrt{2}}{\Delta\cdot r(n)}\rceil$, the maximum number of total data streams that can be received by nodes inside the square at any time slot for any routing scheme is no greater than γ regardless of the number of receiving nodes inside the square.*

Proof. Suppose there are K active links with their respective receivers being in this square. If

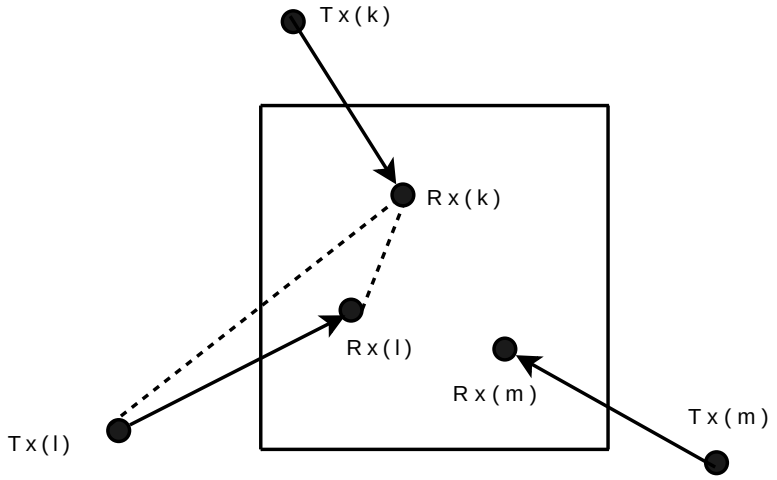


Figure 3.3: The receivers in a square with side length $1/\lceil\frac{\sqrt{2}}{\Delta r(n)}\rceil$.

$K = 1$, the theorem holds trivially, since the number of incoming data streams of a receiver cannot exceed the number of antennas at the receiver. Now we show the result also holds when $K \geq 2$.

We first show that when interference cancellation is not employed, any two links with receiving nodes inside the square will interfere with each other. Note that the distance between any two receivers inside this square is at most $\sqrt{2} \cdot \frac{\Delta r(n)}{\sqrt{2}} = \Delta \cdot r(n)$. Referring to Fig. 3.3, for two links l and k with their receivers $\text{Rx}(l)$ and $\text{Rx}(k)$ inside the square, we have $d_{\text{Rx}(l),\text{Rx}(k)} \leq \Delta \cdot r(n)$. Since $d_{\text{Tx}(l),\text{Rx}(l)} \leq r(n)$ (recall that $r(n)$ is transmission range) based on the triangle inequality, we have $d_{\text{Tx}(l),\text{Rx}(k)} \leq d_{\text{Rx}(l),\text{Rx}(k)} + d_{\text{Tx}(l),\text{Rx}(l)} \leq (1+\Delta)r(n)$. Since $(1+\Delta)r(n)$ is the interference range of $\text{Tx}(l)$, this shows that $\text{Tx}(l)$ will interfere with $\text{Rx}(k)$. Similarly, we can prove that the transmitter $\text{Tx}(k)$ of link k will interfere receiver $\text{Rx}(l)$ of link l.

Denote the set of these K links as $\mathcal{K} = \{1, \ldots, K\}$ and the number of active data streams on link l as $z_l, l \in \mathcal{K}$. We have shown that all these active links will interference with each other. Thus, in order to make them all active simultaneously, interference cancellation is necessary. Based on the MIMO model we discussed earlier, we need an ordered list for all the $2K$ nodes on these K links to determine interference cancellation. Depending on whether the last node in the ordered list is a transmitter or receiver, we have two cases.

Case i. The last node in the ordered list is a receiver. Without loss of generality, assume that receiver $\text{Rx}(m)$ of link m is the last node in this list. To have z_m data streams on link m, based on (3.2.4), we have the following constraint on $\text{Rx}(m)$.

$$z_m + \sum_{k \in \mathcal{I}_m^-}^{\pi(\text{Rx}(m)) > \pi(\text{Tx}(k))} z_k \leq \gamma \ , \tag{3.3.1}$$

where we recall $\pi(\cdot)$ is the mapping function between a node and its order in the node list. Since any two links interfere with each other in this small square, we have $\mathcal{I}_m^- = \mathcal{K} \backslash \{m\}$. Further, since $\text{Rx}(m)$ is the last node in this list, we have $\pi(\text{Rx}(m)) > \pi(\text{Tx}(k))$, for all $k \in \mathcal{K} \backslash \{m\}$. Therefore, (3.3.1) can be written as $z_m + \sum_{k \in \mathcal{K} \backslash \{m\}} z_k \leq \gamma$, which is $\sum_{k \in \mathcal{K}} z_k \leq \gamma$.

Thus, we have shown that the sum of data streams that can be received by nodes inside the small square is no greater than γ regardless of the size of the set $\mathcal{K}$.

Case ii. The last node in the ordered list is a transmitter. In this case, based on (3.2.3) and following the same token as the above discussion, we can get the same result as *Case i*.

Combining the two cases, the proof is complete. □

Based on Lemma 3.1, we can now compute the maximum data rate that can be supported in the unit square network by taking the sum of the data rates among all small squares. Since the side length of each small square is $1/\lceil \frac{\sqrt{2}}{\Delta \cdot r(n)} \rceil$, the total number of small squares in the unit square is $\lceil \frac{\sqrt{2}}{\Delta \cdot r(n)} \rceil^2$. From Lemma 3.1, we know that the maximum number of data streams inside a small square is γ. Thus the total data rate that each square can support is at most γW. So the maximum data rate that can be supported in the network is $\lceil \frac{\sqrt{2}}{\Delta \cdot r(n)} \rceil^2 \gamma W$.

We are now ready to derive a capacity upper bound for MIMO ad hoc network, which is stated in the following theorem.

Theorem 3.1. *For a random multi-hop MIMO ad hoc network, a capacity upper bound for all possible routing and scheduling schemes is* $\lambda(n) = O(\frac{\gamma W}{\sqrt{n \ln n}})$ *with high probability when* $n \to \infty$.

Proof. Let D be the average length of source-destination lines. Since multi-hop routing is em-

ployed, we have that the average number of each source-destination pair is at least $\frac{D}{r(n)}$. Note that there are n source-destination pairs. Thus, the required transmission rate over the entire network is at least $\frac{D}{r(n)}n\lambda(n)$.

When TDMA is used to schedule conflict links into different time slots, the average rate over all time slots in the entire network is at least $\frac{D}{r(n)}n\lambda(n)$. Since the maximum data rate that can be supported in the network at any time slot is $\lceil\frac{\sqrt{2}}{\Delta\cdot r(n)}\rceil^2\gamma W$, we have $\frac{D}{r(n)}n\lambda(n) \leq \lceil\frac{\sqrt{2}}{\Delta\cdot r(n)}\rceil^2\gamma W < (\frac{\sqrt{2}}{\Delta\cdot r(n)}+1)^2\gamma W$, which can be rewritten as

$$\lambda(n) < \frac{2\gamma W}{\Delta^2 Dnr(n)} + \frac{2\sqrt{2}\gamma W}{\Delta Dn} + \frac{\gamma W r(n)}{Dn}\ . \tag{3.3.2}$$

It has been shown in [47] that to maintain the connectivity of the network, we need $r(n) > \sqrt{\frac{\ln n}{\pi n}}$. It can be verified that the right-hand-side of (3.3.2) is a non-increasing function of $r(n)$. By substituting $r(n) = \sqrt{\frac{\ln n}{\pi n}}$ into (3.3.2), we get

$$\lambda(n) < \frac{2\gamma W\sqrt{\pi}}{\Delta^2 D\sqrt{n\ln n}} + \frac{2\sqrt{2}\gamma W}{\Delta^2 Dn} + \frac{\gamma W\sqrt{\ln n}}{Dn\sqrt{\pi n}} = O(\frac{\gamma W}{\sqrt{n\ln n}}). \tag{3.3.3}$$

□

The upper bound in Theorem 3.1 is tight, because it has the same order as that of the capacity lower bound we obtained at the beginning of this section. Combining the capacity lower and upper bounds, we can see that the capacity of a random multi-hop MIMO ad hoc network with n nodes is $\Theta(\frac{\gamma W}{\sqrt{n\ln n}})$.

3.4 Numerical Results

In previous section, our theoretical results show that by using spatial multiplexing and interference cancellation, MIMO can have a constant improvement γ on asymptotic capacity compared to the results of Gupta and Kumar [47]. However, our results also show that MIMO cannot fundamentally improve the asymptotic capacity of multi-hop wireless networks, since it still has the same order as the results of Gupta and Kumar.

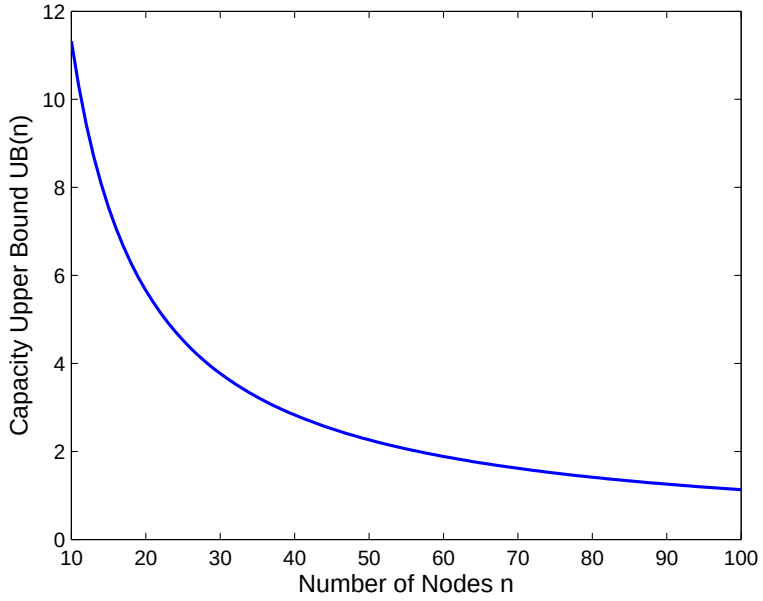


Figure 3.4: Capacity upper bound $UB(n)$.

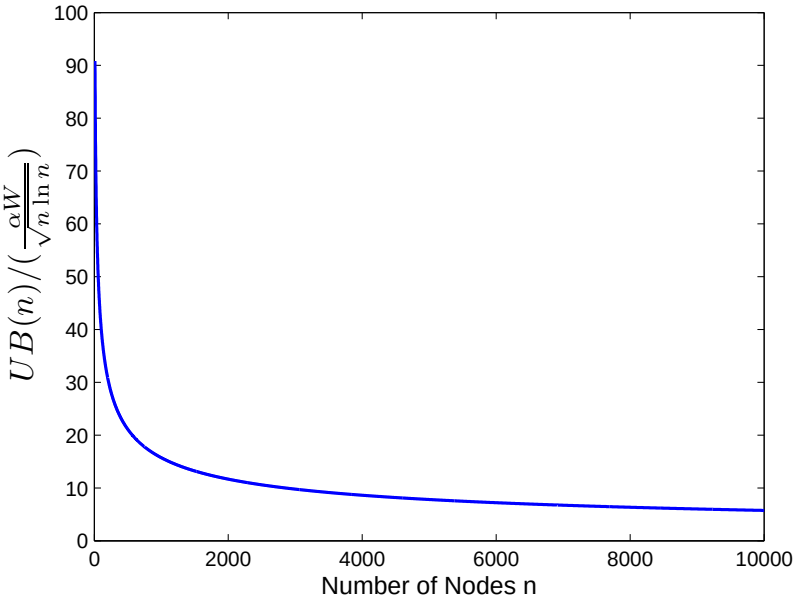


Figure 3.5: The normalized capacity upper bound $UB(n)/(\frac{\gamma W}{\sqrt{n \ln n}})$.

In this section, we will present some numerical results to validate our theoretical findings. We set $\gamma = 4$, $W = 1$, and $\Delta = 1$. By running 1000 instances, we obtain the average length of source-destination lines $D = 0.52$. Denote the capacity upper bound (in Equation (3.3.3)) as $UB(n) = \frac{2\gamma W\sqrt{\pi}}{\Delta^2 D\sqrt{n \ln n}} + \frac{2\sqrt{2}\gamma W}{\Delta^2 Dn} + \frac{\gamma W\sqrt{\ln n}}{Dn\sqrt{\pi n}}$. We show that $UB(n)$ decreases as n increases in Fig. 3.4. Then, we validate $\lambda(n) = O(\frac{\gamma W}{\sqrt{n \ln n}})$ by showing $UB(n)/(\frac{\gamma W}{\sqrt{n \ln n}})$ converges to a constant when n goes to infinity. This result is shown in Fig. 3.5.

3.5 Chapter Summary

In this chapter, we studied capacity scaling laws for MIMO ad hoc networks, i.e., the achievable throughput of each node as the number of nodes in the network increases. Our analysis was based on a MIMO network model that employs zero-forcing beamforming, a powerful physical layer technique that is capable of spatial multiplexing and interference cancellation. Based on this model, we obtained the capacity lower bound and upper bound. The main contribution of this chapter is the development of upper bound, which requires joint consideration of spatial multiplexing and interference cancellation. Our results showed that both lower bound and upper bound have the same order, thus assuring that our asymptotic capacity for MIMO ad hoc network is tight.

Chapter 4

A Unified Approach to Establish Capacity Scaling Laws

4.1 Introduction

Since the seminal results of Gupta and Kumar ("G&K" for short) on capacity scaling law of ad hoc networks with classical single omnidirectional antenna [47], there has been a flourish of research efforts on exploring capacity scaling laws for ad hoc networks with various physical layer technologies. These include directional antennas [92], MIMO [56], cognitive radios [54, 55, 105, 136], and multiple packet reception (MPR) [95], among others. For each of these advanced physical layer technologies, a *custom-designed* analytical approach was developed to study its capacity scaling law. Most of these solutions are typically intellectually challenging and lack universal properties that can be extended to address scaling laws of ad hoc networks with other physical layer technologies.

A fundamental question we ask in this chapter [58] is the following: *instead of custom-designing a sophisticated analytical approach for each physical layer technology, can we devise a set of simple yet universal rules (or general criteria) that one can be easily and quickly applied*

to determine capacity scaling laws for various physical layer technologies? Should such unified rules/criteria exist, then they will offer a set of powerful tools to networking researchers to understand throughput scaling behavior of ad hoc networks under various physical layer technologies, particularly those new technologies that will appear in the future.

The main contribution of this chapter is the development of simple criteria for establishing capacity upper bounds under the protocol model for ad hoc networks under various physical layer technologies. For capacity lower bounds, we argue that a set of simple criteria does not appear to exist, and we give rational on why this is the case in Section 4.11. The following is a summary of our contributions.

- We give an in-depth study of G&K's approach on developing asymptotic capacity upper bound for ad hoc networks with single omnidirectional antennas. We offer insight on why their approach collapses when some advanced technologies are employed at the physical layer.

- We conceive a novel "interference square" concept that divides a normalized 1×1 network area into small interference squares, each with side length $1/\lceil \frac{\sqrt{2}}{\Delta \cdot r(n)} \rceil$, where $r(n)$ is the transmission range and Δ is a parameter to set the interference range under the protocol model. For transmissions within an interference square, we show some unique interference properties.

- Based on the new interference square concept, we develop two simple yet powerful scaling order criteria to determine the asymptotic capacity upper bounds for various physical layer technologies. Either criterion is sufficient to give a capacity upper bound for a given physical layer technology, and the choice of which criterion to use is purely a matter of convenience and only depends on the underlying problem. We also prove the correctness of applying these criteria to obtain capacity upper bounds.

- To demonstrate the usage of our criteria, we study asymptotic capacity of ad hoc networks under various physical layer technologies, such as directional antenna, MIMO, MC-MR,

cognitive radio, and MPR. We show that by applying our simple criteria, one can easily obtain capacity upper bounds under these physical layer technologies, which are consistent to those results in the literature that were developed under custom-designed analytical approaches. Note that our criteria not only can validate those results already reported in literature, but can also determine the upper bounds of ad hoc networks with certain physical layer technology that has not been studied before, and ad hoc networks with new physical layer technologies that will appear in the future.

Note that the unified approach that we propose in this chapter is not without limitations. First, we have only conceived such an approach to determine upper bound. It remains an open problem whether a unified approach exists to determine lower bound. Our efforts to this question have not been fruitful. The main difficulty in deriving a capacity lower bound for a specific physical layer technology is to find a *feasible* solution, which includes resource allocation at the physical layer, scheduling at the MAC layer, and routing at the network layer. A feasible solution to variables at all these layers is tightly coupled to physical layer characteristics and there does not appear to exist a unified pattern one can employ in its development. This is in contrast to that in the development of asymptotic upper bounds, where one can exploit loose inequality relationships (rather than ensuring absolute feasibility). Second, our unified approach is only developed under the protocol model. Developing a unified approach under the physical layer is desirable but remains an open problem. Our success under the protocol model lies in the fact that accounting of interference relationship among such as model is relatively simple — the successful transmission of a link is solely dependent upon the locations of its nearby links and the determination of interference relationship is binary (based on interference range). In contrast, under the physical model, the successful transmission of a link is based on a threshold for SINR. This involves complex calculation of SINR at a receiver, which could require taking into account of interference from all the nodes in the network.

The remainder of this chapter is organized as follows. In Section 4.2, we take a closer look at G&K's classical approach (for ad hoc networks with single omnidirectional antennas) and understand why it falls short of serving as a universal approach for various physical layer technologies.

Subsequently, in Section 4.3, we propose a novel interference square concept and based on this concept, in Section 4.4, we present two simple yet powerful scaling order criteria, which can be used to easily and quickly derive capacity upper bounds for various physical layer technologies. To demonstrate the practical utility of our criteria, in Sections 4.5 to 4.9, we apply our simple criteria to ad hoc networks based on different physical layer technologies such as directional antenna, MIMO, MC-MR, cognitive radio, and MPR. We show that one can easily obtain capacity upper bounds for these networks, which are consistent to those reported in the literature under custom-designed analysis. Section 4.10 offers discussions of our work. Section 4.11 explains why a set of simple criteria for the lower bounds does not appear to exist. Section 4.12 concludes this chapter. Table 4.1 lists notations used in this chapter.

The remainder of this chapter is organized as follows. In Section 4.2, we take a closer look at G&K's approach (for ad hoc networks with classical single omnidirectional antennas) and understand why it cannot be used as a universal approach for various physical layer technologies. Subsequently, in Section 4.3, we propose a novel interference square concept and based on this concept, in Section 4.4, we present two simple yet powerful scaling order criteria, which can be used to quickly derive capacity upper bounds for various physical layer technologies. For validation, in Sections 4.5 to 4.9, we apply our simple criteria to ad hoc networks based on different physical layer technologies such as directional antenna, MIMO, cognitive radio, and MPR. We show that one can easily obtain capacity upper bounds for these networks, which are consistent to those reported in the literature under custom-designed analysis. Section 4.10 offers some discussions of our work. Section 4.11 explains why a set of simple criteria for the lower bounds does not appear to exist. Section 4.12 concludes this chapter. Table 4.1 lists notation used in this chapter.

4.2 Lesson Learned From G&K's Classical Approach

In this section, we take a close look at G&K's classical approach in analyzing capacity scaling law and try to understand why such an approach becomes a barrier in analyzing capacity scaling laws

Table 4.1: Notation in Chapter 4.

General notation	
d_{ij}	Distance between nodes i and j
D	Average distance between all source-destination pairs
$f_{\text{RX}}(n)$	An upper bound for the maximum number of successful transmissions whose receivers are in the same interference square
$f_{\text{TX}}(n)$	An upper bound for the maximum number of successful transmissions whose transmitters are in the same interference square
W	The data rate of a successful transmission in a channel
$r(n)$	The (common) transmission range of all nodes under the protocol model
$\text{Rx}(l)$	Receiver of link l
$\text{Tx}(l)$	Transmitter of link l
Δ	A parameter to set interference range in the protocol model
$\lambda(n)$	Per-node throughput of a random network with n nodes
Ad hoc network with directional antennas	
S	An interference square in the unit area
A_S	Area of S
N_S	Number of nodes in S
MIMO ad hoc network	
$\mathcal{I}_l$	The set of links that are interfered by link l
$\mathcal{Q}_l$	The set of links that are interfering link l
z_l	Number of data streams on link l
$\Pi(\cdot)$	The mapping between a node and its order in the node list
MC-MR network	
c	The number of channels in the network
m	The number of radio interfaces at each node
CR ad hoc network	
$\mathcal{B}_i$	The set of available bands at node i
$\mathcal{B}_{ij}$	The set of available bands on link (i, j)
M	$= \lvert \bigcup_{i=1}^{n} \mathcal{B}_i \rvert$, i.e., the number of distinct frequency bands in the network
Ad hoc network with MPR	
β_1	Number of simultaneous packets from intended transmitters whose transmission range covers a receiver
β_2	Number of unintended transmitters that produce interference on the same receiver
β	A constant representing the total available resource at a receiver

when advanced physical layer technologies are employed.

4.2.1 Background

In G&K's work [47], they considered an ad hoc network of n nodes that are randomly located within a unit square area. Each node in the network is a source node and transmits its data to a randomly chosen destination node. A node's transmission is limited by its transmission range. When the distance between a source node and its destination node is large, multi-hop routing is needed to relay the data. The per-node throughput $\lambda(n)$ is defined as the data rate that can be sent from each source to its destination. A capacity scaling law attempts to characterize the maximum per-node throughput $\lambda(n)$ when the number of nodes n goes to infinity.

In [47], two interference models, the protocol model and the physical model, were considered in their study. In this study, we focus on the protocol model and leave the physical model for future research. In the protocol model [47], each transmitting node is associated with a transmission range $r(n)$, and an interference range $(1+\Delta)r(n)$, where Δ is a constant. To guarantee the connectivity of the network, transmission range $r(n)$ must satisfy the following condition (regardless of the underlying physical layer technology) [46]:

$$r(n) \geq \sqrt{\frac{\ln n}{n}}\,. \tag{4.2.1}$$

When node i transmits to node j, the necessary and sufficient conditions for a successful transmission are:

- node j is within the transmission range of node i, i.e., $d_{ij} \leq r(n)$, where d_{ij} is the distance between nodes i and j, and
- node j is outside the interference range of any other transmitting node k, i.e., $d_{kj} > (1+\Delta)r(n)$, $k \neq i$.

In [47], when the transmission from a node to another node is successful, then the achieved data rate for this transmission is assumed to be a constant W.

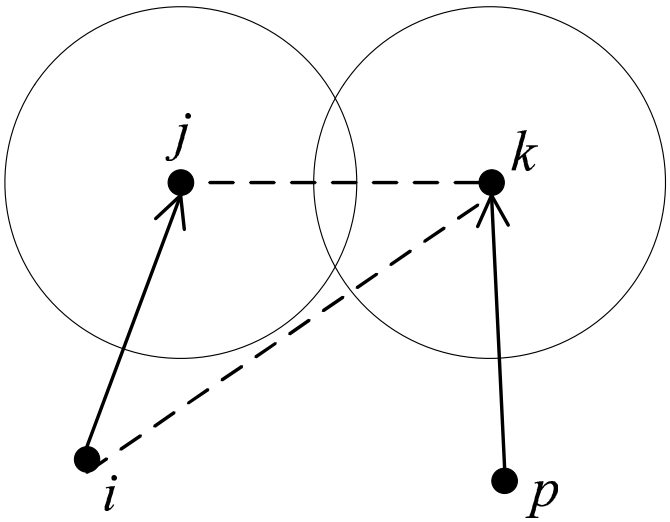


Figure 4.1: Overlapping of two circular footprints of two receiving nodes.

4.2.2 G&K's Approach and Its Limitation

A key component in G&K's approach (in deriving capacity upper bound) is to calculate how much footprint area each successful transmission occupies. Then by dividing the unit square area by this area, they were able to obtain an upper bound of the maximum number of successful transmissions at a time and subsequently to derive a capacity upper bound. Specifically, in [47], G&K showed that for a successful reception at each receiver, one can draw a circle around each receiver with radius $\frac{\Delta r(n)}{2}$ and these circles must be disjoint.[1] Under the above approach, a successful transmission will occupy a circular footprint area of at least $\pi \left[\frac{\Delta r(n)}{2}\right]^2$. Then the maximum number of successful transmissions within the unit square area is at most $1/\left[\pi(\frac{\Delta r(n)}{2})^2\right]$ at any time. Based on this result, G&K derived a capacity upper bound.

The essence of the above footprint area approach is to identify the size of the circular area that

[1]This result can be proved by contradiction. That is, suppose two circles centered at receivers j and k with radius $\frac{\Delta r(n)}{2}$ are not disjoint (see Fig. 4.1), then $d_{jk} \leq \Delta r(n)$. Suppose receiver j is receiving data from transmitter i. Then we have $d_{ij} \leq r(n)$. Based on the triangle inequality, we have $d_{ik} \leq d_{ij} + d_{jk} \leq (1+\Delta)r(n)$, which means that receiver k is within the interference range of i. But this contradicts with the fact that receiving node k must fall outside of the interference range of node i.

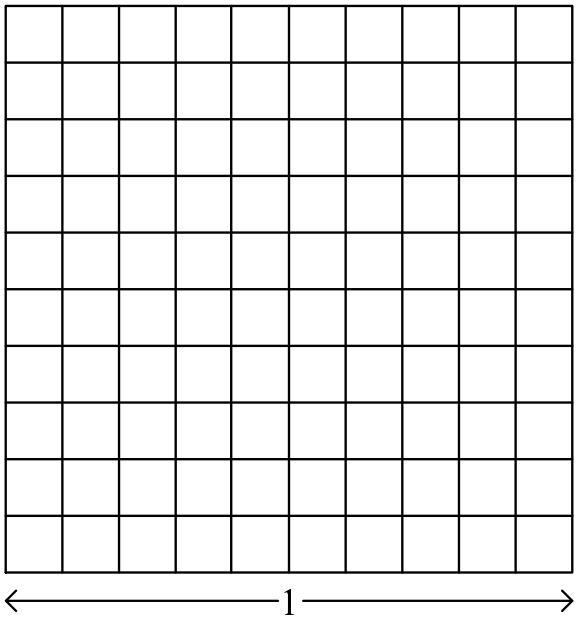


Figure 4.2: The unit square is divided into equal-sized small squares, each with a side length of $1/\lceil \frac{\sqrt{2}}{\Delta \cdot r(n)} \rceil$.

each successful transmission will occupy. But this approach poses a barrier when we encounter advanced physical layer technologies (e.g., MIMO, directional antennas) beyond single omnidirectional antenna node considered in [47]. This is because under these advanced physical layer technologies, the interference relationships among the nodes are much more complex than those under the single omnidirectional antenna scenario in [47]. In particular, the footprint area of each successful receiver does *not* have to be disjoint. For example, in a MIMO ad hoc network where each node employs multiple transmit/receive antennas, receiving node k in Fig. 4.1 may use its degree-of-freedoms (DoFs) to cancel the interference from transmitting node i [19, 107]. As a result, G&K's approach of associating disjoint footprint area with each successful transmission falls apart.

4.3 A New Approach

Given that the footprint area approach in [47] is not capable of handling more complex interference relationships (brought by advanced physical layer technologies), we propose a new approach that handles interference from a different perspective. *Instead of focusing on how much footprint area each successful transmission occupies, we will calculate how many successful transmissions that a given small area in the network can support.* Specifically, we divide the unit square into small equal-sized squares (Fig. 4.2), each with a side length of $1/\lceil \frac{\sqrt{2}}{\Delta \cdot r(n)} \rceil$. We call each small square an *interference square.* As we shall show in Section 4.4, if one can find the maximum number of successful transmissions in each interference square (under a specific physical layer technology), then we can derive the capacity upper bound for the entire network. Subsequently, in Sections 4.5 to 4.9, we show how to find the maximum number of successful transmissions in each interference square under different physical layer technologies, thus deriving capacity upper bound for each of these technologies.

Before we show how this new interference square approach can offer simple scaling law criteria, we discuss some important properties associated with a small square as follows.

Property 4.1. *For a set of successful simultaneous transmissions whose receivers fall in the same interference square, the receiver of any such transmission must be within the interference range of any other transmitter from the same set of transmissions.*

Proof. Note that the distance between any two receivers in the same interference square is at most

$$\sqrt{2} \cdot 1/\left\lceil \frac{\sqrt{2}}{\Delta \cdot r(n)} \right\rceil = \sqrt{2} \cdot \frac{\Delta r(n)}{\sqrt{2}} = \Delta \cdot r(n).$$

Denote $\mathrm{Tx}(l)$ and $\mathrm{Rx}(l)$ the transmitter and receiver of transmission l, respectively. Referring to Fig. 4.3, for any two transmissions l and k with their receivers $\mathrm{Rx}(l)$ and $\mathrm{Rx}(k)$ in the interference square, we have

$$d_{\mathrm{Rx}(l),\mathrm{Rx}(k)} \leq \Delta \cdot r(n).$$

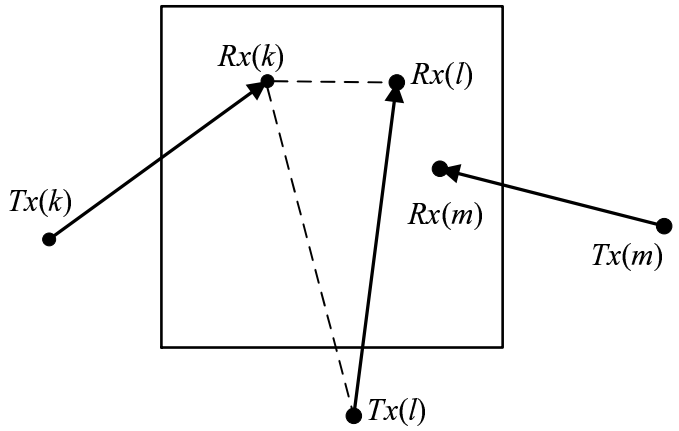


Figure 4.3: A set of transmissions whose receivers are in the same interference square.

Since $d_{\text{Tx}(l),\text{Rx}(l)} \leq r(n)$ (recall that $r(n)$ is transmission range) based on the triangle inequality, we have

$$d_{\text{Tx}(l),\text{Rx}(k)} \leq d_{\text{Rx}(l),\text{Rx}(k)} + d_{\text{Tx}(l),\text{Rx}(l)} \leq (1+\Delta)r(n) \ .$$

Similarly, we can prove that the receiver $\text{Rx}(l)$ of transmission l is also in the interference range of transmitter $\text{Tx}(k)$ of transmission k. □

Similar to Property 4.1 (which considers receivers in the same interference square), we can consider transmitters in the same interference square and have the following property.

Property 4.2. *For a set of successful simultaneous transmissions whose transmitters reside in the same interference square, the receiver of any such transmission must be within the interference range of any other transmitter from the same set of transmissions.*

The proof of Property 4.2 is similar to that of Property 4.1 and is omitted.

Properties 4.1 and 4.2 show us two complementary ways to assess interference relationship from either receiver or transmitter perspective in the same interference square. It turns out that these two properties allow us to calculate the number of successful transmissions with either their receivers or transmitters in the same interference square under various physical layer technologies.

For example, under the single omnidirectional antenna setting in Section 4.2.1, we can easily conclude that there can be at most one active receiver (or transmitter) in an interference square for a successful transmission, i.e., the maximum number of successful transmissions with either receivers or transmitters in the same interference square is one. As another example, for MIMO ad hoc network where each node is equipped with multiple transmit/receiver antennas, Properties 4.1 and 4.2 allow us to show that the maximum number of successful transmissions whose receivers (or transmitters) in the same interference square is upper bounded by the number of antennas at each node (see details in Section 4.6). As we shall show in the next section (Theorems 4.1 and 4.2), the maximum number of successful transmissions whose receivers (or transmitters) are in the same interference square will determine the capacity scaling law of an ad hoc network under various physical layer technologies.

4.4 Main Results: Simple Scaling Order Criteria

As we shall show in Sections 4.5 to 4.9, for a specific physical layer technology, the newly defined interference square and Properties 4.1 and 4.2 enable us to characterize the maximum number of successful transmissions whose receivers (or transmitters) are in the same interference square. For a specific physical layer technology, denote

- $f_{\text{RX}}(n)$ as an upper bound for the maximum number of successful transmissions whose *receivers* are in the same interference square.

Similarly, denote

- $f_{\text{TX}}(n)$ as an upper bound for the maximum number of successful transmissions whose *transmitters* are in the same interference square.

In this section, we show that once we have $f_{\text{RX}}(n)$ or $f_{\text{TX}}(n)$, we can quickly determine a capacity scaling order based on either one of two simple scaling order criteria. Figure 4.4 summarizes the idea of the above discussion.

The two criteria that we present in this section (Theorem 4.1 and 4.2) show that the capacity upper bound scales asymptotically with either $\frac{f_{\text{RX}}(n)}{nr(n)}$ or $\frac{f_{\text{TX}}(n)}{nr(n)}$. We formally state these results as follows.

Theorem 4.1 (Criterion 1)**.** *For a given $f_{RX}(n)$, the asymptotic capacity upper bound of a random ad hoc network is*

$$\lambda(n) = O\left(\frac{f_{RX}(n)}{nr(n)}\right)$$

almost surely when $n \to \infty$. In the special case when $f_{RX}(n)$ is a constant, then $\lambda(n) = O(1/\sqrt{n \ln n})$ almost surely when $n \to \infty$.

Proof. Recall that we divide the unit square into small interference squares with each having a side length of $1/\lceil \frac{\sqrt{2}}{\Delta \cdot r(n)} \rceil$ (see Fig. 4.2). Denote $f_{\text{RX}}(n)$ an upper bound of the maximum number of successful transmissions whose receivers are in the same interference square. Then, the total data rate that each interference square can support is at most $f_{\text{RX}}(n)W$. Now, we can compute the maximum data rate that can be supported by the network in the unit square by taking the sum of the data rates among all small interference squares. Since the side length of each small interference square is $1/\lceil \frac{\sqrt{2}}{\Delta \cdot r(n)} \rceil$, the total number of small interference squares in the unit area is $\lceil \frac{\sqrt{2}}{\Delta \cdot r(n)} \rceil^2$. So the maximum data rate that can be supported in the network is at most $\lceil \frac{\sqrt{2}}{\Delta \cdot r(n)} \rceil^2 f_{\text{RX}}(n)W$.

Let D be the average distance between a source node and its destination node. Since multi-hop routing is employed, we have that the average number of hops for each source-destination pair is at least $\frac{D}{r(n)}$. Note that there are n source-destination pairs. Thus, the required transmission rate over the entire network is at least $\frac{D}{r(n)} n\lambda(n)$.

Since the maximum data transmission that can be supported in the network at a time is $\lceil \frac{\sqrt{2}}{\Delta \cdot r(n)} \rceil^2$ $f_{\text{RX}}(n)W$, we have

$$\frac{D}{r(n)} n\lambda(n) \leq \left\lceil \frac{\sqrt{2}}{\Delta \cdot r(n)} \right\rceil^2 f_{\text{RX}}(n)W < \left(\frac{\sqrt{2}}{\Delta \cdot r(n)} + 1 \right)^2 f_{\text{RX}}(n)W \ ,$$

which gives us

$$\lambda(n) < \frac{2f_{\text{RX}}(n)W}{\Delta^2 Dnr(n)} + \frac{2\sqrt{2}f_{\text{RX}}(n)W}{\Delta Dn} + \frac{f_{\text{RX}}(n)Wr(n)}{Dn} = O\left(\frac{f_{\text{RX}}(n)}{nr(n)}\right). \tag{4.4.1}$$

This proves the first part of Theorem 4.1.

Now, we show the special case when $f_{\text{RX}}(n)$ is a constant. In this case, based on (4.4.1), we have

$$\lambda(n) = O\left(\frac{1}{nr(n)}\right) . \tag{4.4.2}$$

Note that based on (4.2.1), we have $r(n) \geq \sqrt{\frac{\ln n}{n}}$. By substituting $r(n) = \sqrt{\frac{\ln n}{n}}$ into (4.4.2), we have

$$\lambda(n) = O\left(\frac{1}{n\sqrt{\frac{\ln n}{n}}}\right) = O\left(\frac{1}{\sqrt{n \ln n}}\right) .$$

□

Similarly, if we can find $f_{\text{TX}}(n)$, then the following criterion can also give an upper bound for the asymptotic capacity.

Theorem 4.2 (Criterion 2)**.** *For a given $f_{\text{TX}}(n)$, the asymptotic capacity upper bound of a random ad hoc network is*

$$\lambda(n) = O\left(\frac{f_{\text{TX}}(n)}{nr(n)}\right)$$

almost surely when $n \to \infty$. In the special case when $f_{\text{TX}}(n)$ is a constant, then $\lambda(n) = O(1/\sqrt{n \ln n})$ almost surely when $n \to \infty$.

The proof of Theorem 4.2 is similar to that of Theorem 4.1 and is omitted to conserve space.

Several remarks about the above two criteria are in order.

- First, for a specific physical layer technology, we only need to focus on the calculation of either $f_{\text{RX}}(n)$ or $f_{\text{TX}}(n)$, whichever is more convenient. An asymptotic capacity upper bound will follow once we have either $f_{\text{RX}}(n)$ or $f_{\text{TX}}(n)$, based on either Theorem 4.1 or Theorem 4.2.

- Second, when either $f_{\text{RX}}(n)$ or $f_{\text{TX}}(n)$ is a constant, then the asymptotic capacity upper bound is $O(1/\sqrt{n \ln n})$, which is precisely the same as that in [47] by G&K for the protocol model.

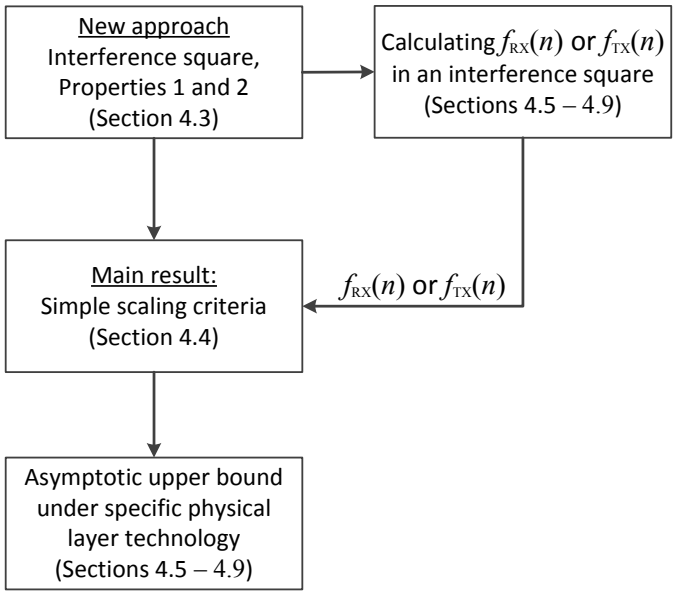


Figure 4.4: A flow chart illustrating our approach to derive asymptotic upper bound for a specific physical layer technology.

This offers a quick test on whether the underlying physical layer technology will indeed change the scaling order of capacity upper bound comparing to the classical single omnidirectional antenna based ad hoc network in [47].

- Finally, the two criteria allow us to focus on calculation ($f_{RX}(n)$ or $f_{TX}(n)$) only within a small interference square. The details associated with network-wide multi-hop end-to-end throughput have been folded in the proof of the two theorems and are no longer of concerns to users of these two theorems in deriving asymptotic capacity upper bound for a given physical layer technology.

Example 4.1. As the first application of our scaling order criterion, let's validate the single omnidirectional antenna based ad hoc network considered in [47]. As discussed in Section 4.3, we have that $f_{RX}(n) = 1$. Thus, by Theorem 1, we have $\lambda(n) = O(1/\sqrt{n \ln n})$, which is precisely the same result in [47] by G&K. □

In the remaining several sections, we will explore asymptotic capacity upper bounds for ad hoc networks under various physical layer technologies. We will present results for directional antennas, MIMO, MC-MR, cognitive radio and MPR in this chapter. Referring to Fig. 4.4, for each case, we will first calculate $f_{\text{RX}}(n)$ or $f_{\text{TX}}(n)$, whichever is more convenient, based on the new interference square and Properties 4.1 and 4.2. This is the upper righthand block in Fig. 4.4. Once we have $f_{\text{RX}}(n)$ or $f_{\text{TX}}(n)$, then we will apply one of the two criteria in this section to quickly obtain the capacity scaling law for this physical layer technology (bottom block in Fig. 4.4).

4.5 Case Study I: Ad Hoc Networks with Directional Antennas

Compared to omnidirectional antenna, directional antenna can control its beam width and concentrate its beam toward its intended destination. Since nodes outside the beam is not interfered, greater spatial reuse inside the network can be achieved. In this section, we apply our criteria in Section 4.4 to explore asymptotic capacity of a random ad hoc network with each node being equipped with a directional antenna. We follow the same model as in [92] by Peraki and Servetto.[2] The scaling law results in [92] are well known and widely cited. They showed that for the single-beam model, the asymptotic capacity scales as $O\left(r(n)\right)$ and for the multi-beam model, it scales as $O\left(nr^3(n)\right)$. The analysis in [92] was custom-designed and differed from that by G&K. In this section, we show that by applying our criteria in Section 4.4, we can quickly obtain the same results for asymptotic capacity upper bound in [92].

We organize this section as follows. First, we consider the case for the single-beam model. Then, we consider the multi-beam model.

[2]Another work on scaling law for directional antennas is [134] by Yi *et al.*, which employed a slightly different model and thus led to a different set of results. The approach in [134] followed the same token as that in [47] by G&K. It can be shown that our criteria can be easily applied there and we leave the details to readers as an exercise.

4.5.1 Scaling Law Analysis for Single Beam Model

Single Beam Model

In [92], single beam model refers that a transmitter can generate at most one directional beam to an intended receiver, although a receiver can receive multiple directed beams from different nodes.

Calculating $f_{\text{TX}}(n)$

In this case study, we choose to calculate $f_{\text{TX}}(n)$, which is more convenient than $f_{\text{RX}}(n)$. As discussed in Section 4.4, the choice of calculating $f_{\text{TX}}(n)$ or $f_{\text{RX}}(n)$ is solely based on convenience and either one is sufficient to determine asymptotic capacity.

Recall that $f_{\text{TX}}(n)$ is an upper bound for the maximum number of successful transmissions whose transmitters are in the same interference square. In the case of single-beam model, $f_{\text{TX}}(n)$ corresponds to an upper bound for the maximum number of successful beam transmissions whose transmitters are in the same interference square. To calculate $f_{\text{TX}}(n)$, we need the following lemma.

Lemma 4.1. *The number of nodes in the same interference square is* $\Theta\left(nr^2(n)\right)$ *almost surely when* $n \to \infty$.

Proof. Denote S an interference square within the unit area. Denote A_S and N_S the area and the number of nodes in S, respectively. Since nodes in S are randomly distributed, we have the average number of nodes in S is $E(N_S) = nA_S$. For the number of nodes in S, we have that the following probabilities (also known as Chernoff bounds) [86].

- For any $\delta > 0$, $P\left\{N_S > (1+\delta)nA_S\right\} < \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{nA_S}$.
- For any $0 < \delta < 1$, $P\left\{N_S < (1-\delta)nA_S\right\} < e^{-\frac{1}{2}nA_S\delta^2}$.

Combining the above two inequalities, for any $0 < \delta < 1$, we have

$$\begin{aligned} & P\{|N_S - nA_S| > \delta nA_S\} \\ =& P\{N_S > (1+\delta)nA_S\} + P\{N_S < (1-\delta)nA_S\} \\ <& \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{nA_S} + e^{-\frac{1}{2}nA_S\delta^2} \\ =& e^{-\theta_1 nA_S} + e^{-\theta_2 nA_S}, \end{aligned} \tag{4.5.1}$$

where $\theta_1 = (1+\delta)\ln(1+\delta) - \delta$ and $\theta_2 = \frac{1}{2}\delta^2$.

Note that $A_S = \frac{1}{\left\lceil \frac{\sqrt{2}}{\Delta \cdot r(n)} \right\rceil^2} = \Theta(r^2(n))$. Letting $A_S = \Theta(r^2(n))$ in (4.5.1), we have

$$P\{|N_S - nA_S| > \delta nA_S\} < e^{-\theta_1 n\Theta(r^2(n))} + e^{-\theta_2 n\Theta(r^2(n))}. \tag{4.5.2}$$

Based on (4.2.1), we have $r(n) = \Omega(\sqrt{\frac{\ln n}{n}})$. Thus, the right-hand-side of (4.5.2) goes to zero when $n \to \infty$, which shows that the probability that the deviation of the number of nodes in S from the mean by more than a constant factor of the mean is zero when $n \to \infty$. Based on the definition of $\Theta(\cdot)$, we have $N_S = \Theta(nr^2(n))$. □

Based on Lemma 4.1, we have the following lemma for $f_{\text{TX}}(n)$.

Lemma 4.2. *For a random ad hoc network under single-beam directional antenna, we have* $f_{\text{TX}}(n) = \Theta\left(nr^2(n)\right)$.

Proof. By Lemma 4.1, there are $\Theta\left(nr^2(n)\right)$ nodes in the interference square. Since each node can only generate one beam, the total number of successful beam transmissions generated by the transmitters in this interference square cannot exceed $\Theta\left(nr^2(n)\right)$, i.e., $f_{\text{TX}}(n) = \Theta\left(nr^2(n)\right)$. □

Scaling Law

Following Fig. 4.4, with $f_{\text{TX}}(n) = \Theta\left(nr^2(n)\right)$, we can now apply Theorem 4.2 and quickly obtain the following asymptotic capacity upper bound.

Proposition 4.1. *For a random ad hoc network under single-beam directional antenna, we have* $\lambda(n) = O(r(n))$ *almost surely when* $n \to \infty$.

Proof. Combining Lemma 4.2 and Theorem 4.2, we have

$$\lambda(n) = O\left(\frac{f_{\text{TX}}(n)}{nr(n)}\right) = O\left(nr^2(n) \cdot \frac{1}{nr(n)}\right) = O(r(n)).$$

□

Note that this result for single-beam case is the same as that in [92].

4.5.2 Scaling Law Analysis for the Multi-Beam Model

Multi-Beam Model

In [92], a multi-beam model refers that a transmitting node can generate multiple beams to different receiving nodes at the same time. On the other hand, a receiving node can only receive one beam from the same transmitting node but may receive multiple beams from different transmitting nodes. We follow the same model in [92] for the multi-beam case.

Calculating $f_{\text{RX}}(n)$

We will calculate $f_{\text{RX}}(n)$.[3] Recall that $f_{\text{RX}}(n)$ is an upper bound of the maximum number of successful transmissions whose receivers are in the same interference square. In the case of multi-beam model, $f_{\text{RX}}(n)$ corresponds to an upper bound of the maximum number of successful beam transmissions received by the receivers that are in the same interference square.

For receivers residing in the same interference square, it is easy to see that their transmitters cannot be outside a larger square, with the same center as the interference square, but with side

[3]The level of difficulty in calculating $f_{\text{RX}}(n)$ is the same as that for $f_{\text{TX}}(n)$ in the multi-beam model. Either choice will lead to the same result.

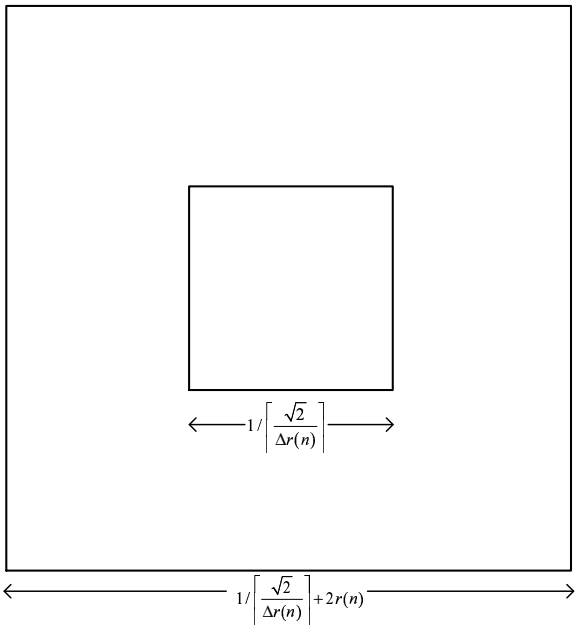


Figure 4.5: The larger square contains all the transmitters that can transmit directional beams to the receivers that are in the small interference square at the center.

length of $1/\lceil \frac{\sqrt{2}}{\Delta \cdot r(n)} \rceil + 2r(n)$ (see Fig. 4.5). Otherwise, a receiver in the interference square will be outside of a transmitter's transmission range $r(n)$. For the number of nodes inside the larger square (regardless of transmitters or receivers), we have the following lemma.

Lemma 4.3. *The number of nodes in the larger square with side length* $\frac{1}{\left\lceil \frac{\sqrt{2}}{\Delta \cdot r(n)} \right\rceil} + 2r(n)$ *is* $\Theta\left(nr^2(n)\right)$ *almost surely when* $n \to \infty$.

The proof of Lemma 4.3 is similar to the proof of Lemma 4.1 and is omitted here. Now, we are ready to calculate $f_{\text{RX}}(n)$ as follows.

Lemma 4.4. *For a random ad hoc network under multi-beam directional antenna, we have* $f_{RX}(n) = O\left(n^2 r^4(n)\right)$.

Proof. Based on Lemma 4.3, we know that the number of transmitters that can transmit beams to the same receiver in the interference square is at most $O\left(nr^2(n)\right)$. That is, a receiver in the

interference square can receive at most $O\left(nr^2(n)\right)$ beams. By Lemma 4.1, there are at most $\Theta\left(nr^2(n)\right)$ receivers in the same interference square. So we have

$$f_{\text{RX}}(n) = \Theta\left(nr^2(n)\right) \cdot O\left(nr^2(n)\right) = O\left(n^2r^4(n)\right).$$

□

Scaling Law

Following Fig. 4.4, with $f_{\text{RX}}(n) = O\left(n^2r^4(n)\right)$, we can now apply Theorem 4.1 and quickly obtain the following asymptotic capacity upper bound.

Proposition 4.2. *For a random ad hoc network under multi-beam directional antenna, we have* $\lambda(n) = O\left(nr^3(n)\right)$ *almost surely when* $n \to \infty$.

Proof. Combining Lemma 4.4 and Theorem 4.1, we have

$$\lambda(n) = O\left(\frac{f_{\text{RX}}(n)}{nr(n)}\right) = O\left(n^2r^4(n) \cdot \frac{1}{nr(n)}\right) = O\left(nr^3(n)\right) \ .$$

□

This result is the same as that in [92] for the multi-beam case.

4.6 Case Study II: MIMO Ad Hoc Networks

4.6.1 MIMO Model

By employing multiple antennas at both transmitting and receiving nodes, MIMO has brought significant benefits to wireless communications, such as increased link capacity [33, 113], improved link diversity [139], and interference cancellation between conflicting links [19, 107]. In this section, we characterize asymptotic capacity upper bound for multi-hop MIMO ad hoc networks.

Although there are many schemes to exploit the benefits of antenna arrays at a node, we focus on the two key characteristics of MIMO: *spatial multiplexing* (SM) and *interference cancellation* (IC) [19, 107, 137]. SM refers that a transmitter can send several independent data streams to its intended receiver simultaneously on a link. IC refers that by properly exploiting multiple antennas at a node, potential interference to and/or from other nodes can be cancelled.

To model SM and IC, we employ recent advance in MIMO link model in [104] by Shi *et al.* In this model, degree-of-freedom (DoF) is used to represent resource at a MIMO node. Simply put, the number of DoFs at a node is equal to the number of antennas, denoted as α, at the node. Denote z_l the number of active data streams on link l in a time slot. Denote $\text{Tx}(l)$ and $\text{Rx}(l)$ the transmitter and the receiver of link l, respectively. To spatial multiplex z_l data streams on link l, we need to allocate z_l ($z_l \leq \alpha$) DoFs at both transmitter $\text{Tx}(l)$ and receiver $\text{Rx}(l)$. To cancel interference from and/or to other nodes in the network, it is necessary to have an ordered list for all nodes and allocate DoFs at each node following this order [104]. Denote $\Pi(\cdot)$ the mapping between a node and its order in the node list. Suppose that link l is carrying z_l data streams. Denote $\mathcal{I}_l$ and $\mathcal{Q}_l$ the set of links that are interfered by link l and the set of links that are interfering link l, respectively. Transmitter $\text{Tx}(l)$ is responsible for cancelling the interference from itself to all receivers $\text{Rx}(k)$, $k \in \mathcal{I}_l$, that are before node $\text{Tx}(l)$ in the order list. Similarly, receiver $\text{Rx}(l)$ of link l is responsible for cancelling the interference from all transmitters $\text{Tx}(k)$, $k \in \mathcal{Q}_l$, that are before node $\text{Rx}(l)$ in the order list. Since the total number of DoFs for SM and IC cannot exceed α, we have the following two constraints on each active link l in the network.

1. DoF constraint at $\text{Tx}(l)$: The number of DoFs that $\text{Tx}(l)$ can use for SM (for transmission) and IC cannot exceed the total number of DoFs at node $\text{Tx}(l)$, i.e.,

$$z_l + \sum_{k \in \mathcal{I}_l}^{\Pi(\text{Tx}(l)) > \Pi(\text{Rx}(k))} z_k \leq \alpha \ . \tag{4.6.1}$$

2. DoF constraint at $\text{Rx}(l)$: The number of DoFs that receiver $\text{Rx}(l)$ can use for SM (for recep-

tion) and IC cannot exceed the total number of DoFs at node $\mathrm{Rx}(l)$, i.e.,

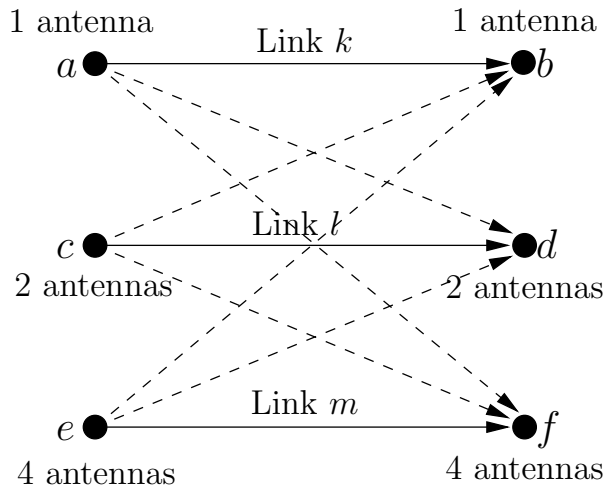
$$z_l + \sum_{k \in \mathcal{Q}_l}^{\Pi(\mathrm{Rx}(l)) > \Pi(\mathrm{Tx}(k))} z_k \leq \alpha \ . \tag{4.6.2}$$

We use the following simple example to illustrate DoF allocation in a MIMO network.

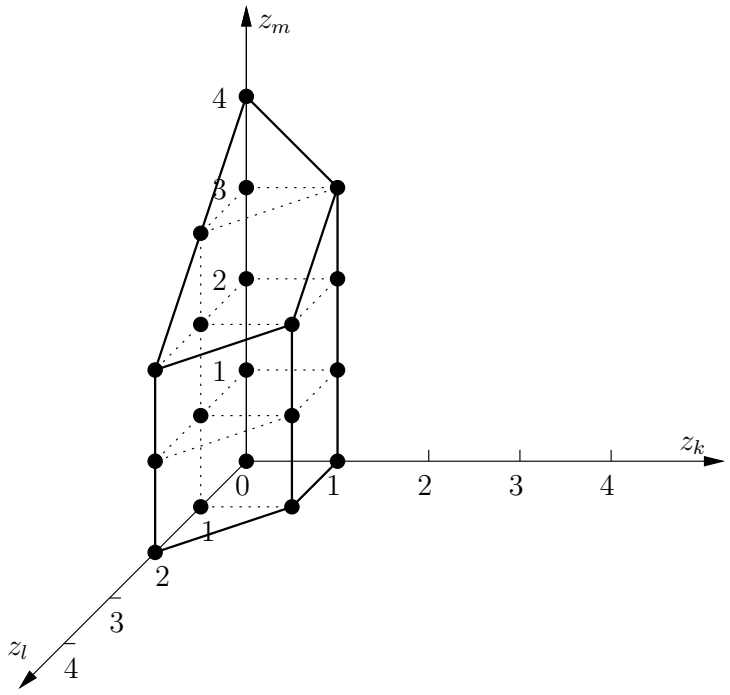
Example 4.2. Consider the three-link (k, l, and m) example in Fig. 4.6(a). The number of antennas at each node is also shown in the figure. Under the above MIMO model, we need an order to determine the DoF resource usage at each node. Suppose we are following an order list, say $a \to d \to b \to c \to e \to f$ among the nodes. Then, the DoF allocation in this MIMO network works as follows.

We start with node a, which is the first node in the list. Given it is the first in the list, node a does not have any interference with which it needs to be concerned. Since node a has only 1 antenna, it can transmit at most 1 data stream to its intended receiver b. The second node on the ordered list is node d. Since it appears in the order list after node a, node d needs to suppress the interference from a. This implies that node d needs to expend 1 DoF to cancel the interference from a. Since d has 2 antennas, we have that d can receive at most $2 - 1 = 1$ stream, i.e., $z_l \leq 1$. The DoF consumption on nodes b and c follows exactly the same token, and it can be verified that b and c can each receive and transmit 1 stream, respectively. Since node e's transmission should not interfere with the reception at b and d that had appeared in the order list earlier, e needs to expend 2 DoFs for this purpose. At this point, e can transmit up to $4 - 1 - 1 = 2$ streams, i.e., $z_m \leq 2$. Finally, along the same line, node f can receive at most $4 - 1 - 1 = 2$ streams, i.e., $z_m \leq 2$. Therefore, after the above steps, we can see that the stream combination $(z_k = 1, z_l = 1, z_m = 2)$ can be scheduled feasibly on links k, l, and m. It can be shown that the entire DoF region (the set of all feasible stream combinations) for the three-link example in Fig. 4.6(a) can be found by enumerating all possible choices of the node order list. Each stream combination offers a feasible point (e.g., $(1, 1, 2)$), the union of which constitutes the DoF region, which we plot in Fig. 4.6(b).

□



(a) Inter-nodal interference relationship for three links.



(b) Achievable DoF region of the three MIMO links.

Figure 4.6: A three-link MIMO network example.

4.6.2 Calculating $f_{\text{RX}}(n)$

Based on the MIMO network model, we now calculate $f_{\text{RX}}(n)$.[4] Recall that $f_{\text{RX}}(n)$ is an upper bound of the maximum number of successful transmissions whose receivers are in the same interference square. In the case of MIMO, this corresponds to the maximum number of successful data streams on all active links whose receivers are in the same interference square.

Lemma 4.5. *For a random MIMO ad hoc network, we have* $f_{RX}(n) = \alpha$.

Proof. Denote $\mathcal{L}$ the set of active links with their receivers being in the same interference square. Denote $|\mathcal{L}|$ the number of links in $\mathcal{L}$, and let $\mathcal{L} = \{1, \ldots, |\mathcal{L}|\}$. Our goal is to find an upper bound for the sum of data streams on these links, i.e., $\sum_{k\in\mathcal{L}} z_k$.

If $|\mathcal{L}| = 1$, i.e., only one active link with its receiver in the interference square, then $z_1 \leq \alpha$ (since the number of data streams on this link cannot exceed the number DoFs of a node). We can set $f_{\text{RX}}(n) = \alpha$ and the lemma holds trivially.

For the general case of $|\mathcal{L}| \geq 2$, Property 4.1 says that these $|\mathcal{L}|$ links interfere with each other and IC is necessary. Based on the MIMO model we discussed earlier, we need to follow an ordered list for the nodes (both transmitters and receivers) on these $|\mathcal{L}|$ links for DoF allocation at each node. We have two cases, depending on whether the last node in the list is a transmitter or a receiver.

Case (i). The last node in the ordered list is a receiver. Without loss of generality, denote m as the link of which this node is the receiver. To have z_m data streams on link m, based on (4.6.2), we have the following constraint on receiver $\text{Rx}(m)$.

$$z_m + \sum_{k\in\mathcal{Q}_m}^{\Pi(\text{Rx}(m))>\Pi(\text{Tx}(k))} z_k \leq \alpha \ , \tag{4.6.3}$$

where the sum for z_k is taken over all interfering links whose transmitters are before receiver $\text{Rx}(m)$ in the node list. Since link m is being interfered by all other links in $\mathcal{L}$ in the same inter-

[4]For MIMO, the level of difficulty in calculating $f_{\text{RX}}(n)$ is the same as $f_{\text{TX}}(n)$ and either approach will yield the same result.

ference square, we have $\mathcal{Q}_m = \mathcal{L}\backslash\{m\}$. Further, since $\text{Rx}(m)$ is the last node in this list, we have $\Pi(\text{Rx}(m)) > \Pi(\text{Tx}(k))$, for all $k \in \mathcal{L}\backslash\{m\}$. Therefore, (4.6.3) can be re-written as

$$z_m + \sum_{k \in \mathcal{L}\backslash\{m\}} z_k \leq \alpha \ ,$$

which is

$$\sum_{k \in \mathcal{L}} z_k \leq \alpha \ .$$

Thus, we have shown that the sum of data streams that can be received by nodes in the interference square over all links is upper bounded by α, i.e., $f_{\text{RX}}(n) = \alpha$.

Case (ii). The last node in the ordered list is a transmitter. In this case, we employ (4.6.1) and follow the same token as the above discussion. We again have $f_{\text{RX}}(n) = \alpha$.

Combining the two cases, we have $f_{\text{RX}}(n) = \alpha$. □

4.6.3 Scaling Law

Following Fig. 4.4, with $f_{\text{RX}}(n) = \alpha$, we can now apply Theorem 4.1 and obtain asymptotic capacity upper bound of a random MIMO ad hoc network as follows.

Proposition 4.3. *For a random MIMO ad hoc network, we have* $\lambda(n) = O(1/\sqrt{n \ln n})$ *almost surely when* $n \to \infty$.

This result is the same as that in [56]. It is also interesting to see that, despite MIMO's ability to increase capacity in a *finite-sized* network, the scaling order for its asymptotic capacity remains the same as that for a single omnidirectional antenna network as in [47].

4.7 Case Study III: Multi-Channel and Multi-Radio

4.7.1 Multi-Channel Multi-Radio Model

Multi-channel multi-radio (MC-MR) refers that there are multiple channels in the network and there are multiple radio interfaces at each node in the network [68, 69]. By equipping each node with multiple radio interfaces, each node has more flexibility in channel access in the network. Following [68], we assume that there are c channels in the network and each node in the network is equipped with m radio interfaces, where c and m are constants, and $1 \leq m \leq c$. A radio interface is capable of transmitting or receiving data on only one channel at any given time, i.e., half-duplex.

4.7.2 Calculating $f_{\text{RX}}(n)$

Based on the MC-MR model, we now calculate $f_{\text{RX}}(n)$.[5] Assuming each band has the same bandwidth in the MC-MR network, then $f_{\text{RX}}(n)$ corresponds to the maximum number of successful transmissions over all available channels on all radio interfaces whose receivers are in the same interference square. We have the following lemma.

Lemma 4.6. *For a random MC-MR network, we have $f_{\text{RX}}(n) = c$.*

Proof. Let's focus on one channel at a time. Since the links with receivers in the interference square interfere with each other (Property 4.1), there can be at most one radio at a node receiving on this channel. Summing up all such radios (or successful transmissions) over c channels, we have $f_{\text{RX}}(n) = c$. □

[5] For an MC-MR network, the level of difficulty in calculating $f_{\text{RX}}(n)$ is the same as $f_{\text{TX}}(n)$ and either approach will yield the same result.

4.7.3 Scaling Law

Following Fig. 4.4, with $f_{\text{RX}}(n) = c$, we can now apply Theorem 4.1 and obtain asymptotic capacity upper bound of an MC-MR ad hoc network as follows.

Proposition 4.4. *For a random MC-MR ad hoc network, we have* $\lambda(n) = O\left(\frac{1}{\sqrt{n \ln n}}\right)$ *almost surely when* $n \to \infty$.

Note that this result is the same as the result in [68] for the case when $\frac{c}{m} = O(\ln n)$.

4.8 Case Study IV: Cognitive Radio Ad Hoc Networks

4.8.1 Cognitive Radio Network Model

Cognitive radio (CR) is another new physical layer technology that enables more efficient utilization of radio spectrum [132]. A CR is able to constantly sense the radio spectrum and explore any available spectrum bands for data communication. Consider a random ad hoc network where each node is equipped with a CR. Consider a specific time instance where each node i senses a set of available frequency bands $\mathcal{B}_i$ that it can use.[6] Note that due to differences in locations, the set of available frequency bands $\mathcal{B}_i$ at a node i may be different from that at another node in the network. Denote $\mathcal{B}_{ij} = \mathcal{B}_i \bigcap \mathcal{B}_j$ the set of common bands that are available at both nodes i and j. Then node i can communicate to node j on band m only if $m \in \mathcal{B}_{ij}$.

4.8.2 Calculating $f_{\text{RX}}(n)$

Based on the CR network model, we now calculate $f_{\text{RX}}(n)$.[7] Assuming that each band has the same bandwidth in the CR network, then $f_{\text{RX}}(n)$ corresponds to the maximum number of successful

[6] These bands may be those that are currently unused by the primary users.

[7] For a CR network, the level of difficulty in calculating $f_{\text{RX}}(n)$ is the same as $f_{\text{TX}}(n)$ and either approach will yield the same result.

transmissions over all available bands whose receivers are in the same interference square. Denote $M = |\bigcup_{i=1}^{n} \mathcal{B}_i|$, i.e., M is the number of distinct frequency bands in the network. Then we have the following lemma.

Lemma 4.7. *For a random CR ad hoc network, we have* $f_{RX}(n) = M$.

Proof. Consider one band at a time. Within each band, by Property 4.1, the links with receivers in the interference square interfere with each other. So the maximum number of active links (or successful transmissions) is at most one. Summing up all active links (or successful transmissions) over M bands, we have $f_{\text{RX}}(n) = M$. □

4.8.3 Scaling Law

Following Fig. 4.4, with $f_{\text{RX}}(n) = M$, we can now apply Theorem 4.1 and obtain asymptotic capacity upper bound for a random CR ad hoc network as follows.

Proposition 4.5. *For a random CR ad hoc network, we have* $\lambda(n) = O\left(\frac{1}{\sqrt{n \ln n}}\right)$ *almost surely when* $n \to \infty$.

This result is consistent to those found in [54, 105]. It is interesting to see that, despite that CR can utilize spectrum bands more efficiently (and thus higher capacity for a finite-sized network), the scaling order of its asymptotic capacity remains the same as that for the classical single omnidirectional antenna network in [47].

4.9 Case Study V: Ad Hoc Networks with Multi-Packet Reception

Multi-packet reception (MPR) is a conceptual abstraction of a physical layer capability that a receiver can correctly decode multiple packets from different transmitters simultaneously [114]. As

described in [95], such capability may be implemented by a variety of advanced physical layer technologies, such as multiuser detection [119], directional antennas [92, 134], and MIMO. In other words, MPR refers to a reception capability of a node at the physical layer, rather than referring to a specific physical layer technology. In this section, we employ our criteria in Section 4.4 to explore capacity scaling law of MPR-based ad hoc networks.

4.9.1 An MPR Model

Under MPR, a transmitter can transmit packet to only one receiver at a time, but a receiver is capable of receiving multiple packets simultaneously from multiple transmitters within its transmission range. For unintended transmissions whose interference range covers a receiver, the receiver will consider them as interference. Such interference may be cancelled by the receiver. Specifically, in the MPR model, we assume a receiver has finite resource available for MPR and interference cancellation. Denote β_1 the number of simultaneous packets from intended transmitters whose transmission range covers the receiver and β_2 the number of unintended transmitters that produce interference on the same receiver. We have

$$\beta_1 + \beta_2 \leq \beta \ ,$$

where β is a constant and represents the total available resource at a receiver. For example, if MIMO is employed to implement MPR, then the number of DoFs at a MIMO node may correspond to β.

Note that this MPR model is a generalization of the idealized MPR model in [95] which assumes $\beta_1 \leq \beta = \infty$ and $\beta_2 = 0$, i.e., a receiver can successfully decode arbitrary number of simultaneous packet transmissions and no interference is allowed on the receiver.

4.9.2 Calculating $f_{\text{RX}}(n)$

We choose to calculate $f_{\text{RX}}(n)$, which is more convenient than calculating $f_{\text{TX}}(n)$. In the case of MPR ad hoc networks, $f_{\text{RX}}(n)$ corresponds to an upper bound of the maximum number of packets that are successfully received simultaneously by all the receivers in the same interference square. We have the following lemma for $f_{\text{RX}}(n)$.

Lemma 4.8. *For a random MPR ad hoc network, we have* $f_{\text{RX}}(n) = \beta$.

Proof. Denote $\mathcal{L}$ the set of successful links with their receivers residing in the same interference square. By a "successful" link, we mean the receiver of this link can successfully decode the packet on this link. Denote $|\mathcal{L}|$ the number of links in $\mathcal{L}$, and let $\mathcal{L} = \{1, \ldots, |\mathcal{L}|\}$. Then $f_{\text{RX}}(n)$ is an upper bound of $|\mathcal{L}|$.

Note that for two successful links, their transmitters are different but their receivers may be the same. Consider one receiver j in the interference square. From receiver j's perspective, we divide $\mathcal{L}$ into two subsets: $\mathcal{L}_1$ — the set of links whose receivers are j, and $\mathcal{L}_2$ — the set of links whose receivers are not j. Based on Property 4.1, we know that the transmitters of the links in subset $\mathcal{L}_2$ are all in the interference range of receiver j. Since packets on $\mathcal{L}_1$ are successfully received by j, then based on the MPR model, we have

$$|\mathcal{L}| = |\mathcal{L}_1| + |\mathcal{L}_2| = \beta_1 + \beta_2 \leq \beta \ .$$

Therefore, we have $f_{\text{RX}}(n) = \beta$. □

4.9.3 Scaling Law

Following Fig. 4.4, with $f_{\text{RX}}(n) = \beta$, we can now apply Theorem 4.1 and directly obtain the following asymptotic capacity upper bound for an MPR-based ad hoc network.

Proposition 4.6. *For a random MPR ad hoc network, we have* $\lambda(n) = O(1/\sqrt{n \ln n})$ *almost surely when* $n \to \infty$.

Remark 4.1. *For the idealized MPR model described in [95], where $\beta_1 \leq \beta = \infty$ and $\beta_2 = 0$, one can still apply our simple scaling order criteria. In particular, it can be shown that for this idealized MPR model, we have $f_{RX}(n) = \Theta\left(nr^2(n)\right)$ in Lemma 4.9.* □

Lemma 4.9. *For a random ad hoc network under the idealized MPR model, we have $f_{RX}(n) = \Theta(nr^2(n))$.*

Proof. First, we show that there can be only one receiver (say j) in the interference square receiving packets. This can be shown by contradiction. Suppose there is another receiver i, $i \neq j$, that receives packets in the same interference square. Then, based on Property 4.1, one of receiver i's transmitters must be within the interference range of node j. This transmitter of receiver i will interfere node j, which contradicts with $\beta_2 = 0$ under the idealized MPR model.

Although there is only one receiver j receiving packets, it may receive packets from multiple transmitters. Note that all nodes that can transmit to receiver j must fall within the larger square with a side length of $1/\lceil \frac{\sqrt{2}}{\Delta \cdot r(n)} \rceil + 2r(n)$ (see Fig. 4.5). Based on Lemma 4.3, we know that the number of all nodes inside the larger square is $\Theta(nr^2(n))$. Since each transmitter transmits one packet to receiver j at a time, the number of simultaneous packets received by receiver j cannot exceed the number of nodes in the larger square, i.e., $\Theta(nr^2(n))$. Therefore, we have $f_{\text{RX}}(n) = \Theta(nr^2(n))$. □

Combining Lemma 4.9 and Theorem 4.1, we have

$$\lambda(n) = O\left(\frac{f_{\text{RX}}(n)}{nr(n)}\right) = O\left(nr^2(n) \cdot \frac{1}{nr(n)}\right) = O\left(r(n)\right) \ .$$

This is exactly the result developed in [95].

Table 4.2: A summary of asymptotic capacity upper bounds obtained via our simple criteria. "—" sign indicates new result not available in literature.

Physical layer technology		$f_{\text{RX}}(n)$ or $f_{\text{TX}}(n)$	Upper bound	Reference
Directional antenna	Single beam	$f_{\text{TX}}(n) = \Theta\left(nr^2(n)\right)$	$O\left(r(n)\right)$	[92]
	Multi-beam	$f_{\text{RX}}(n) = O\left(n^2r^4(n)\right)$	$O\left(nr^3(n)\right)$	[92]
MIMO		$f_{\text{RX}}(n) = \alpha$	$O\left(\frac{1}{\sqrt{n\ln n}}\right)$	[56]
MC-MR		$f_{\text{RX}}(n) = c$	$O\left(\frac{1}{\sqrt{n\ln n}}\right)$	[68]
CR		$f_{\text{RX}}(n) = M$	$O\left(\frac{1}{\sqrt{n\ln n}}\right)$	[54, 105]
MPR	Idealized	$f_{\text{RX}}(n) = \Theta\left(nr^2(n)\right)$	$O\left(r(n)\right)$	[95]
	General	$f_{\text{RX}}(n) = \beta$	$O\left(\frac{1}{\sqrt{n\ln n}}\right)$	—

4.10 Discussions

4.10.1 Summary of Results

Table 4.2 summarizes asymptotic capacity upper bounds that we derived in the last five sections by applying our unified approach. For the MPR general model, the result we developed is new and there is no prior result available in the literature. The upper bounds under the other physical layer technologies are consistent to those reported in the literature (references given in the last column of Table 4.2). Recall those results were developed by various custom-designed approaches.

It is important to realize that the goal of this chapter is not to obtain a capacity bound under a particular physical layer technology, as those in [54, 56, 68, 92, 95, 105], but rather to offer a *unified* tool to determine asymptotic capacity upper bounds for wireless networks with different physical layer technologies.

4.10.2 Limitation

Although Table 4.2 demonstrates the potential capability of our simple scaling order criteria, we caution that the success of our simple criteria hinges upon a successful calculation of $f_{\text{RX}}(n)$ or $f_{\text{TX}}(n)$. For other physical layer technologies, there is no guarantee that one can always calculate $f_{\text{RX}}(n)$ or $f_{\text{TX}}(n)$ as we have done in this chapter. Further, one should calculate $f_{\text{RX}}(n)$ or $f_{\text{TX}}(n)$ as tight as possible since loose $f_{\text{RX}}(n)$ or $f_{\text{TX}}(n)$ (e.g., infinity) will yield trivial upper bounds. But one thing that we can guarantee is that should one be able to find $f_{\text{RX}}(n)$ or $f_{\text{TX}}(n)$ for the underlying physical layer technology, then she can easily apply our simple scaling order criteria to quickly obtain asymptotic upper bound.

4.10.3 Asymptotic Order Change

We observe that for advanced physical layer techniques such as MIMO, MC-MR, cognitive radio, and general MPR, the asymptotic capacity upper bounds are $O\left(\frac{1}{\sqrt{n \ln n}}\right)$, which is the same as that under single omnidirectional antenna [47]. Given that $O\left(\frac{1}{\sqrt{n \ln n}}\right)$ is a tight upper bound, we conclude MIMO, MC-MR, cognitive radio, and general MPR cannot make fundamental change in asymptotic capacity upper bound.[8] This is an interesting result. On the other hand, under directional antenna and idealized MPR, the asymptotic capacity upper bounds are on a higher order than $O\left(\frac{1}{\sqrt{n \ln n}}\right)$. This indicates that there could be some potential to improve the network capacity.

[8]It is important to realize that capacity scaling law only shows a general trend on how capacity changes when $n \to \infty$. Therefore, no improvement in asymptotic capacity does not mean there is no improvement in capacity when network size is finite. It is well known that most of these advanced physical layer technologies can significantly improve network capacity in finite-sized networks.

4.11 Asymptotic Capacity Lower Bounds

Note that so far the simple scaling order criteria that we developed in Section 4.4 can only offer asymptotic capacity upper bounds for different physical layer technologies. A natural question to ask is whether one can develop a set of simple criteria to quickly obtain asymptotic capacity lower bounds for any physical layer technology. Our efforts to this question have not been fruitful. The main difficulty in deriving a capacity lower bound for a specific physical layer technology is to find a *feasible* solution, which includes resource allocation at the physical layer, scheduling at the MAC layer, and routing at the network layer. A feasible solution to variables at all these layers is much harder to obtain than just developing inequality relationships that are needed to derive asymptotic upper bounds. Given such feasible solution is hard to obtain, it remains an open problem whether or not it is possible to develop a unified approach that yields a set of simple criteria for asymptotic capacity lower bounds.

Despite the absence of a simple criteria for the lower bounds, we may use $\Omega(1/\sqrt{n \ln n})$ (capacity lower bound for single omnidirectional antenna ad hoc networks by G&K [47]) as a lower bound in many cases. This is because single omnidirectional antenna can usually be considered as a special case of these advanced physical layer technologies. In particular, for MIMO, MC-MR, CR, MPR general model in Table 4.2, we have lower bounds of $\Omega(1/\sqrt{n \ln n})$ and upper bounds of $O(1/\sqrt{n \ln n})$. In these cases, since the upper bound and lower bound have the same scaling order, we conclude that $\lambda(n) = \Theta(1/\sqrt{n \ln n})$ for these advanced physical layer technologies. In other cases where $\Omega(1/\sqrt{n \ln n})$ may appear loose (e.g., single beam and multi-beam directional antenna, idealized MPR), one would need to develop a tighter lower bound by exploiting the unique properties of the underlying physical layer technology.

4.12 Chapter Summary

In this chapter, we presented a set of simple yet powerful general criteria that one can apply to quickly determine the asymptotic capacity upper bounds for ad hoc networks under the protocol model for various physical layer technologies. This new approach offers a unifying methodology to determine capacity scaling law, which is in contrast to many custom-designed approaches. We proved the correctness of our proposed criteria and demonstrated their application through a number of case studies, such as ad hoc networks with directional antenna, MIMO, MC-MR, cognitive radio, and MPR. These simple criteria offer a set of powerful tools to networking researchers to understand throughput scaling behavior of ad hoc networks under different physical layer technologies, particularly new technologies that will appear in the future.

Part II

Optimizations of Finite-Sized Wireless Networks

Chapter 5

Exploiting SIC for Multi-Hop Wireless Networks

5.1 Introduction

Interference is widely regarded as the fundamental impediment to throughput performance in wireless networks. In the wireless networking community, the classical and main stream approach to handle interference is to employ certain *interference avoidance* scheme, which can be done either through deterministic resource allocation (e.g., TDMA, FDMA, or CDMA) or random access based schemes (e.g., CSMA, CSMA/CA). The essence of an interference avoidance scheme is to avoid any potential overlap among the transmitting signals (the root of interference). Although natural and easy to implement, an interference avoidance scheme, in general, cannot offer a performance close to network information theoretical limit [115].

Recently, there is a growing interest in exploiting interference (rather than avoiding it) to increase network throughput (see Section 5.2 for related work). In essence, such an *interference exploitation* approach allows overlap among transmitting signals and relies on some advanced physical layer schemes to remove or cancel interference. In particular, the so-called *successive*

interference cancellation (SIC) scheme appears very promising [3, 10, 40, 49, 78, 124] and has already attracted development efforts from industry (e.g., QUALCOMM's CSM6850 chipset for cellular base station [96]). Under SIC, a receiver attempts to decode concurrent signals from multiple transmitters successively, starting from the strongest signal. If the strongest signal can be decoded, it will be subtracted from the aggregate signal so that the SINR (signal-to-interference-and-noise-ratio) for the remaining signals can be improved. Then the SIC receiver continues to decode the second strongest signal and so forth, until all signals are decoded, or terminates if the remaining signals are no longer decodable (see Section 5.3 for more details).

Although SIC has been extensively studied as a physical layer technology, its limitation and optimal application in the context of *multi-hop* wireless networks remain limited. In this chapter [59], we will try to answer the following fundamental questions.

- What are the limitations of SIC? How to overcome such limitations?
- How to develop an optimization framework for optimal interaction between SIC and interference avoidance? How to incorporate variables from multiple layers (physical, link, and network) into such an optimization framework?

We take a formal optimization approach to address these fundamental questions. We find that the limitations of SIC come from its stringent constraints when decoding multiple signals. Specifically, in order to decode aggregate signals successively, an SIC receiver must meet a series of SINR constraints for its received signal powers. Further, due to these constraints, there exists a decoding limit for SIC in its abilities for concurrent receptions or interference rejection. Due to this limit, We find that SIC alone is inadequate to handle all concurrent interference in a multi-user wireless network.

Interestingly, the limitations of SIC can be compensated precisely by the classical *interference avoidance* scheme. Therefore, *we advocate a joint interference exploitation and interference avoidance approach, which combines the best of both worlds while avoid each other's pitfalls.* We believe such a joint approach is most efficient to handle interference in a multi-hop wireless

network.

Although the need of such a joint approach is easy to understand, a number of new technical challenges arise in the context of a multi-hop network. This is particularly true when our optimization space encompasses physical layer SIC, link layer scheduling, and network layer flow routing. We address these new challenges by developing a formal optimization framework, with cross-layer formulation of physical, link, and network layers. This new optimization framework offers a holistic design space to squeeze the most out of interference and lay a mathematical foundation for the modeling and analysis of a joint interference exploitation and avoidance scheme in a multi-hop wireless network. To the best of our knowledge, this is the first effort toward this direction. To demonstrate the practical utility of our optimization framework, we conduct a case study for the network throughput maximization problem. Our numerical results affirm the efficacy of this framework and give us insights on how SIC should optimally interact with an interference avoidance scheme.

The rest of this chapter is organized as follows. Section 5.2 presents related work on interference exploitation. Section 5.3 offers a primer on SIC and illustrates its benefits. In Section 5.4, we discuss some inherent limitations of SIC. In Section 5.5, we advocate a joint interference exploitation and avoidance approach to overcome these limitations. In Section 5.6, we develop mathematical models for constraints under such a scheme. In Section 5.7, we develop a formal optimization framework for joint interference exploitation-avoidance scheme. In Section 5.8, we apply our optimization framework on a case study and present some numerical results. Section 5.9 concludes this chapter. Table 5.1 lists notations used in this chapter.

5.2 Related Work

At the physical layer, a major reference on interference exploitation (cancellation) is the book by Verdu [119] and references therein. For more details and new advances of some important interference cancellation techniques, we refer readers to study SIC [35, 120], parallel interference

Table 5.1: Notation in Chapter 5.

Symbol	Definition
A_j	The maximum number of signals an SIC receiver j can decode
B	Channel bandwidth
C_{ij}	The maximum achievable rate on link $i \to j$
d_{ij}	Distance between nodes i and j
$d(f)$	Destination node of session $f \in \mathcal{F}$
$\mathcal{F}$	The set of user sessions in the network
g_{ij}	Channel gain from node i to node j
$\mathcal{I}_i$	The set of neighboring nodes of node i
$\mathcal{L}$	The set of links in the network
$\mathcal{N}$	The set of nodes in the network
$\mathcal{N}_j$	The set of transmitting nodes when j is receiving
P	The transmission power of each node
$P_j^{\max}$	$= \max_{i \in \mathcal{N}_j} P_{ij}$, the maximum power of all signals received at node j
P_{ij}	The received power at node j from node i
$r(f)$	Data rate of session $f \in \mathcal{F}$
$r_{ij}(f)$	Data rate that is attributed to session f on link $i \to j$
$s(f)$	Source node of session f
T	The number of time slots in a time frame
$x_{ij}[t]$	A binary indicator of weather the transmission on link $i \to j$ is successful or not in time slot t
$w(f)$	A weight associated with session f
R	The data rate of a successful transmission
β	The SINR threshold for successful decoding
γ	Path loss index
$\lambda_i[t]$	A binary indicator of weather node i is transmitting or not in time slot t
σ^2	The power level of ambient noise

cancellation [42, 118], iterative interference cancellation (turbo multiuser user detection) [60, 122], which all aim to enable a receiver to decode multiple signals at the same time, and reject interference from other unintended transmitters. A recent review on how to apply interference cancellation for cellular systems was given in [3], which positioned SIC as one of the most promising techniques to mitigate interference due to its simplicity and effectiveness.

Note that the SIC considered in this chapter differs from some new interference cancellation schemes such as analog network coding [62] and ZigZag decoding [43]. Both were proposed to resolve packet collisions, and they require knowledge of some bits in one of the colliding packets. SIC also differs from smart antenna-based interference cancellation schemes, such as Zero-Forcing Beam Forming (ZFBF) [5, 104, 137] in MIMO[1] and directional antennas [76, 97, 108].

Very recently, there is a growing interest to exploit SIC at the physical layer to improve performance at upper layers in a wireless network [10, 40, 49, 78, 79, 80, 124]. In [49], Halperin *et al.* built a ZigBee prototype of SIC based on [119, Ch. 7] using software radios and used experimental results to validate that SIC is an effective way to improve system throughput. In [78], Lv *et al.* studied a scheduling problem in an ad hoc network with SIC. To simplify network-layer problem, the authors considered fixed routes in the network (e.g., based on shortest path), and subsequently developed a greedy heuristic scheduling algorithm based on conflict set graph. Link scheduling problem for wireless networks with SIC was also studied in [79, 80], but the aggregate interference effect of the practical SINR model was not considered. In [40], Gelal *et al.* proposed a topology control framework to exploit SIC. They studied how to divide a network topology into a minimum number of sub-topologies where the set of links in each sub-topology can be active at the same time. In [124], Weber *et al.* studied asymptotic transmission capacity of one-hop ad hoc networks with SIC under a simplified model where all signals from transmitters within a specific radius can all be successfully decoded. More realistic SIC model for asymptotic transmission capacity was later explored by Blomer and Jindal in [10]. We also notice a recent paper [99] claiming that the potential gain by SIC is marginal. This is in contrast to the state-of-the-art [10, 40, 49, 78, 79, 124]

[1]Note that MIMO requires multiple antennas for interference cancellation, while SIC does not have such requirement. This chapter considers SIC with a single antenna on each node.

as well as our findings in this chapter. A closer look at [99] shows that their SIC scheme did not fully exploit the benefits of SIC. They only considered a simple network with two links. They compared the completion time required to transmit one packet on both links with and without SIC. When without SIC, the two links transmits data sequentially and the completion time is the sum of the time used on both links. With SIC, the two links can transmit data simultaneously and the completion time was defined as the maximum time used by these two links. We argue that such a comparison is not fair, as the link that finishes its transmission first can start to transmit other packets instead of being idle.

To date, results on how to apply SIC in a *multi-hop* network remain very limited, particularly those results that explores the limitations of SIC and how such limitations can be compensated by interference avoidance. This is the focus of this chapter.

5.3 SIC: A Primer

Under the classical information reception model in a wireless network, a receiver j treats all interfering signals from other concurrent (non-intended) transmissions as noise. For the signal coming from the intended transmitting node i, if its SINR at node j is greater than or equal to a threshold β, then the transmission is said to be successful (i.e., the signal from node i to j can be decoded successfully). Denote P_{ij} the power level of the signal from node i that is received by node j. Denote $\mathcal{N}_j$ the set of concurrent transmitting nodes that can be heard by node j. Then, under the classical model, a successful transmission from node i to node j occurs if

$$\frac{P_{ij}}{\sum_{k\in\mathcal{N}_j}^{k\neq i} P_{kj} + \sigma^2} \geq \beta \ ,$$

where constant σ^2 is power level of the ambient noise.

In contrast to the above classical paradigm, a receiver with SIC capability can decode a number of concurrent signals (including some interfering signals) rather than treating them blindly as noise [49], [119, Ch. 7], [124]. This is done by decoding concurrent signals in a *sequential* or-

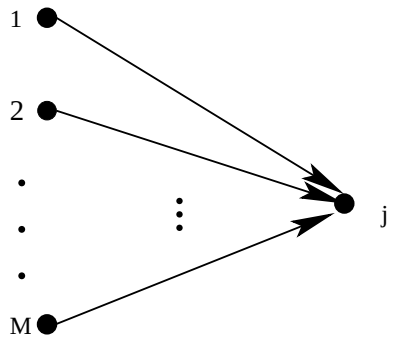


Figure 5.1: A receiver with M concurrent transmitters.

der and subtracting each successfully decoded signal before proceeding to decode the next signal. Figure 5.1 illustrates a communication scenario where a node j is receiving from M concurrent transmitters. Under SIC, receiver j first attempts to decode the strongest signal. If the strongest signal can be decoded successfully (i.e., the SINR of this signal is no less than the threshold β), then this signal will be subtracted from the aggregate signal (see Fig. 5.2). Then the receiving node j tries to decode the second strongest signal and so forth. The process continues until all the signals are successfully decoded or at some stage the SINR criterion for the underlying signal is no longer satisfied.

Without loss of generality, referring to Fig. 5.1, suppose that the power levels of the signals from the M transmitters received at node j are in nondecreasing order as $P_{1j} \leq P_{2j} \leq \cdots \leq P_{Mj}$. Receiving node j tries to decode the signals from transmitting nodes in the order of $M, M-1, \cdots 1$. Then, the signal with received power P_{ij} can be decoded successfully if and only if

$$\begin{aligned} \text{Step 1} \qquad & \frac{P_{Mj}}{\sum_{k=1}^{M-1} P_{kj} + \sigma^2} \geq \beta \ , \\ \text{Step 2} \qquad & \frac{P_{M-1,j}}{\sum_{k=1}^{M-2} P_{kj} + \sigma^2} \geq \beta \ , \\ \vdots \qquad & \qquad \vdots \\ \text{Step } (M-i+1) \qquad & \frac{P_{ij}}{\sum_{k=1}^{i-1} P_{kj} + \sigma^2} \geq \beta \ . \end{aligned} \tag{5.3.1}$$

As shown in (5.3.1), in order to decode the signal with received power P_{ij}, it is necessary to

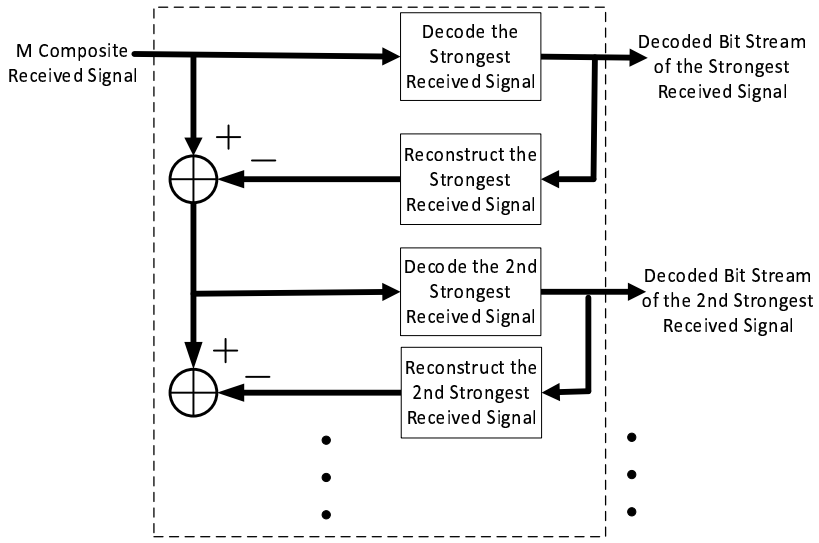


Figure 5.2: A schematic of the SIC process.

decode all the preceding stronger signals first. Note that we assume perfect cancellation of a successfully decoded signal in the iterative process. Similar to [40, 78, 79], we do not consider link rate adaptation in our model and assume that the data rate for each successful transmission is $R = B\log_2(1+\beta)$, where B is the channel bandwidth. We leave the more complex case with link rate adaption as future work.

There are two key benefits associated with SIC, namely, *concurrent receptions from multiple transmitters* and *interference rejection*. In the rest of this section, we elaborate these two benefits.

Concurrent Receptions from Multiple Transmitters. Note that under the classical reception model, only one intended transmitter is allowed to transmit; concurrent transmissions to the same receiver will lead to a collision and are considered wasteful of resource. In contrast, an SIC receiver is capable of receiving from multiple transmitters at the same time (if the criteria in (5.3.1) are met) and thus can substantially increase throughput in the network. As a simple example, consider Fig. 5.3, where both nodes 1 and 2 wish to transmit to node 3. Assume $P_{13} = 1$, $P_{23} = 2$, $\sigma^2 = 1$, and $\beta = 1$, where all units are normalized with appropriate dimensions. Under the classical interference avoidance model, nodes 1 and 2 cannot transmit to node 3 at the same time due to

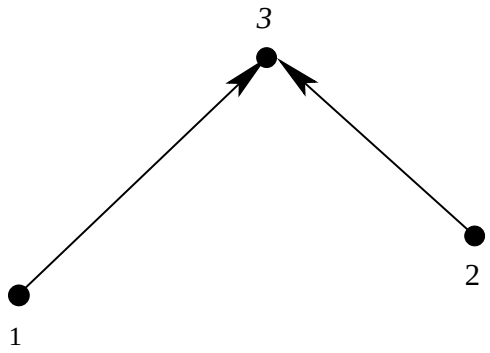


Figure 5.3: An example of concurrent receptions from multiple transmitters.

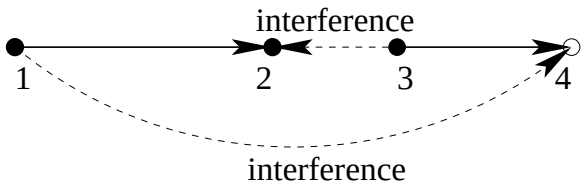


Figure 5.4: An example for interference rejection.

interference. Under SIC, receiver 3 can first decode the stronger signal from node 2 by treating the signal from 1 as interference. We have $\frac{2}{1+1} = 1 \geq \beta$. Next, receiver 3 subtracts the decoded signal from the aggregate signal. The SINR from node 1 is $\frac{P_{13}}{\sigma^2} = \frac{1}{1} = 1 \geq \beta$, which shows transmission from node 1 is also successful.

Interference Rejection. The ability to decode multiple received signals can also help the receiving node to reject interference from unintended transmitters. As a simple example, consider the two-transmitter two-receiver case in Fig. 5.4. Node 1 wishes to send data to node 2 while node 3 wishes to send data to node 4. Due to the broadcast nature of a wireless channel, the signal from node 3 will interfere with the reception at node 2 and likewise the signal from node 1 will interfere with the reception at node 4. Assume $P_{12} = 1$, $P_{14} = 0.5$, $P_{32} = 3$, $P_{34} = 1.6$, $\sigma^2 = 1$, and $\beta = 1$. Under the classical interference avoidance model, links $1 \rightarrow 2$ and $3 \rightarrow 4$ cannot be active at the same time. Under SIC, receiver 2 can first try to decode the stronger received signal, which is the signal from node 3. Since $\frac{P_{32}}{P_{12}+\sigma^2} = \frac{3}{1+1} = 1.5 \geq \beta$, such decoding is successful. Then, node

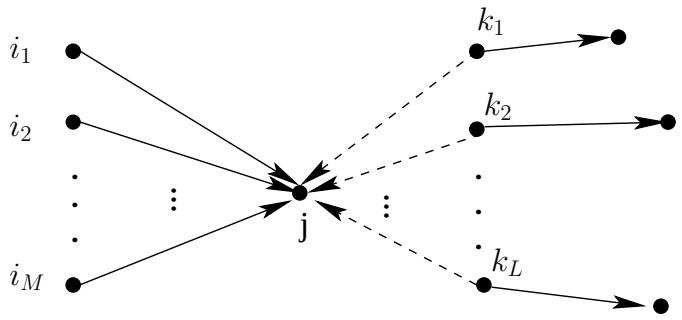


Figure 5.5: The general case of concurrent reception and interference rejection at a receiving node j. A solid arrow represents intended transmission and a dashed arrow represents interference.

2 subtracts this decoded signal from the aggregate signal, and tries to decode the second stronger signal, which is from node 1. We have $\frac{P_{12}}{\sigma^2} = \frac{1}{1} = 1 \geq \beta$. So this decoding is again successful. Likewise, on node 4, it tries to decode the stronger received signal first, which is from node 3. Since $\frac{P_{34}}{P_{14}+\sigma^2} = \frac{1.6}{0.5+1} = 1.07 \geq \beta$, this decoding is successful.

Summary. Our discussion of the above two benefits can be generalized by Fig. 5.5. In this figure, a receiving node j tries to decode all the signals it receives, among which it tries to retain the desired bit streams from the M intended transmitters and reject the interfering bit streams from the L unintended transmitters.

5.4 Understanding the Limitations of SIC

The potential benefits of SIC are easy to recognize. But we caution the readers that this is one part of the story. In this section, we show some stringent constraints and hard limits that SIC must satisfy before yielding any potential benefit.

Sequential SINR Constraint. As we have shown in Section 5.3, at any stage when a receiver tries to decode the desired signal from the aggregate signal, the SINR must satisfy (5.3.1). Otherwise, the current signal cannot be decoded successfully, neither will all the remaining weaker

signals.

Again let's use the two-transmitter one-receiver example in Fig. 5.3. Assuming $P_{13} < P_{23}$, to decode both the signals from nodes 1 and 2 successfully, P_{13} and P_{23} must satisfy

$$\frac{P_{23}}{P_{13}+\sigma^2} \geq \beta \text{ and } \frac{P_{13}}{\sigma^2} \geq \beta \ .$$

But suppose $P_{23} = 1.2, P_{13} = 1$, $\sigma^2 = 0.5$ and $\beta = 1$. Then we have $\frac{P_{23}}{P_{13}+\sigma^2} = \frac{1.2}{1+0.5} = 0.8 < \beta$. This means that even the stronger signal from node 2 cannot be successfully decoded. Therefore, the weaker signal from node 1 cannot be decoded either. In this case, SIC will not work.

Sequential Decoding Limit. Another limitation of SIC is that it can only decode a limited number of signals (either intended or unintended). Such limit is determined by (5.3.1) and sets up a cap on the number of decodable signals. Before we calculate this limit, we present the following property.

Property 5.1. (Geometric Power Property) Denote P_{1j}, P_{2j}, $\cdots$, P_{Mj} the received powers of the signals that can be successfully decoded at node j via SIC. Without loss of generality, suppose $P_{1j} \leq P_{2j} \leq \ldots \leq P_{Mj}$. Then, we have

$$P_{ij} \geq \beta(1+\beta)^{i-1}\sigma^2, \text{ for } i = 1, \ldots, M \ .$$

Proof. Our proof is based on induction. First consider $i = 1$. Since all previous stronger interference are removed from the composite interference when decoding the weakest signal, the SINR for P_{1j} is $\frac{P_{1j}}{\sigma^2}$, which must be no less than β. Then, we have $\frac{P_{1j}}{\sigma^2} \geq \beta$, which is $P_{1j} \geq \beta\sigma^2$.

Next, suppose that

$$P_{ij} \geq \beta(1+\beta)^{i-1}\sigma^2, \ i = 1, \ldots, l \ . \tag{5.4.1}$$

We will prove that $P_{l+1,j} \geq \beta(1+\beta)^l\sigma^2$. We know that we still have all the interference from the weaker signals when we decode the signal from $P_{l+1,j}$. Then, we have $P_{l+1,j}/(\sum_{i=1}^{l} P_{ij}+\sigma^2) \geq \beta$,

which gives us

$$\begin{aligned} P_{l+1,j} &\geq \beta \left(\sum_{i=1}^{l} P_{ij} + \sigma^2 \right) \\ &\geq \beta \left[\sum_{i=1}^{l} \beta(1+\beta)^{i-1}\sigma^2 + \sigma^2 \right] \\ &= \beta \left[1 + \beta \frac{(1+\beta)^l - 1}{(1+\beta) - 1} \right] \sigma^2 \\ &= \beta(1+\beta)^l \sigma^2 \ , \end{aligned}$$

where the second inequality holds due to (5.4.1). □

Now we are ready to calculate the limit on the number of signals that can be decoded. More formally, denote A_j an upper bound of the number of signals that receiver j can decode. Then we have the following lemma.

Lemma 5.1. *Denote $P_j^{\max}$ the strongest possible received power at receiver j, i.e., $P_j^{\max} = \max_{i \in \mathcal{N}_j} P_{ij}$, where $\mathcal{N}_j$ is the set of all active concurrent transmitters. Then the number of successfully decoded signals at receiver j is no more than $A_j = 1 + \log_{\beta+1}(\frac{P_j^{\max}}{\beta\sigma^2})$.*

Proof. Let $P_{1j} \leq P_{2j} \leq \ldots \leq P_{Mj}$ be any set of powers of the signals successfully decoded at receiver j, we have $P_j^{\max} \geq P_{Mj}$. Combining $P_j^{\max} \geq P_{Mj}$ with Property 5.1 gives us $P_j^{\max} \geq P_{Mj} \geq \beta(1+\beta)^{M-1}\sigma^2$, which gives us

$$M \leq 1 + \log_{\beta+1} \left(\frac{P_j^{\max}}{\beta\sigma^2} \right) \ .$$

The above inequality says that the number of successfully decoded signals at receiver j is upper bounded by $A_j = 1 + \log_{\beta+1}(\frac{P_j^{\max}}{\beta\sigma^2})$. □

As an example of the sequential decoding limit, we assume that $P_j^{\max} = 10$, $\sigma^2 = 1$, and $\beta = 1$. Based on Lemma 5.1, we have $A_j = 1 + \log_{1+1}(\frac{10}{1 \cdot 1}) = 4.32$. That is, only up to four signals can be successfully decoded at receiver j.

Remark 5.1. *Note that A_j given in Lemma 5.1 is only an upper bound. The actual number of decodable signals may be much lower than this bound. This is because that the powers of decodable signals must also satisfy the sequential SINR constraints in (5.3.1).* □

5.5 An Optimal Approach for Interference Exploitation

5.5.1 A New Approach

Based on the discussion in Section 5.3, an interference exploitation scheme such as SIC has clear advantage over a pure interference avoidance scheme. On the other hand, due to the intrinsic limitations associated with SIC, it is evident that SIC alone is inadequate in a multi-hop network. As a result, it is necessary to incorporate interference avoidance (e.g., scheduling) to mitigate such limitations. This is true for both the sequential SINR constraints and sequential decoding limit. In particular, when the sequential SINR constraints are no longer satisfied at certain stage, one has to resort to scheduling (e.g., time slot assignment) to avoid interference so that different transmissions can be carried out successfully. Likewise, whenever the number of interfering transmissions exceeds the sequential decoding limit, one again has to employ scheduling to allocate these transmissions into different time slots such that the number of interfering transmissions in each time slot is within the decoding limit. In other words, *we should take the best of both worlds (interference exploitation and interference avoidance) while avoid each other's pitfalls.* This is the approach that we advocate in this chapter. Given its clear advantage over a pure interference exploitation paradigm and a pure (classical) interference avoidance paradigm, it has the potential to evolve into the main stream paradigm for wireless medium access in the coming years.

5.5.2 New Challenges

There are several new challenges when developing a joint interference exploitation and interference avoidance scheme, particularly in a multi-hop network.

- At the physical layer, under the classical SINR model, a receiving node treats all the other concurrent (unintended) interfering transmissions as noise when deciding whether or not the underlying intended transmission is successful. This itself is not a trivial problem as the set of interfering transmissions is usually coupled with upper layer scheduling and routing algorithms. In the context of SIC, not only one needs to deal with such coupling with upper layer algorithms, one also has to deal with multiple transmissions, in the sense that one has to decode those stronger signals before decoding its own signal (in a sequential order). This sequential decoding imposes significant difficulty in developing a tractable model for mathematical programming.

- At the link layer, a scheduling algorithm (i.e., interference avoidance scheme) is needed to address the limitations of SIC at the physical layer. Note that such scheduling algorithm is also coupled with routing in a multi-hop network environment. How to design an optimal scheduling algorithm to fulfill certain network performance objective in this context is a new and non-trivial problem.

- As discussed in Section 5.3, SIC allows more concurrent transmissions in the network than traditional interference avoidance model. This offers many more available links for choosing a path at the network layer. Consequently, the design space at the network layer is much larger, leading to a more complex optimization problem.

To address these new challenges, it is necessary to develop a framework of joint interference exploitation and avoidance for multi-hop wireless networks.

5.6 Modeling of Cross-layer Constraints

As a first step toward a formal optimization framework, we examine constraints across the three lower layers for a multi-hop network. Consider a single antenna multi-hop wireless network, with a set of SIC-capable nodes $\mathcal{N}$ operating within the same channel. For interference avoidance, we

consider TDMA in the time domain.[2] Under TDMA, we assume a frame is divided into T time slots, each of equal length. For simplicity, we do not consider power control of individual node and assume each node transmits at the same power P. Denote g_{ij} the channel gain from node i to node j. Then, when node i is transmitting, the received power at node j is $P_{ij} = P \cdot g_{ij}$.

Scheduling Constraints. We first define a binary scheduling variable $x_{ij}[t]$ for link $i \to j$ in time slot t ($1 \le t \le T$).

$$x_{ij}[t] = \begin{cases} 1 & \text{if node } i \text{ transmits data to node } j \text{ successfully in time slot } t; \\ 0 & \text{otherwise.} \end{cases}$$

By "successfully," we mean that the intended transmission from node i can be decoded at node j via SIC, i.e., the sequential SINR constraints in (5.3.1) are satisfied for this signal. In the case of an "unsuccessful" transmission (i.e., the sequential SINR constraints in (5.3.1) are not satisfied for this signal), it is desirable to turn off the transmitter rather than having it transmit undecodable signals. Therefore, when $x_{ij}[t] = 0$, we will not have any transmission from node i to node j.

Denote $\mathcal{I}_i$ the set of all neighboring nodes of node $i \in \mathcal{N}$. For unicast communication in the network, a node transmits data to only one node in a time slot, i.e.,

$$\sum_{j \in \mathcal{I}_i} x_{ij}[t] \le 1 \quad (i \in \mathcal{N}, 1 \le t \le T) \, . \tag{5.6.1}$$

For reception at a node, it becomes more interesting, as a node can receive data from multiple nodes in a time slot. That is, for a receiver j, we may have $\sum_{i \in I_j} x_{ij}[t] > 1$.

Based on the state-of-the-art in the literature, there is no evidence that SIC can achieve full-duplex with a single antenna. Therefore, half-duplex will still be necessary at each node. To model half-duplex at a node i, we have

$$x_{ki}[t] + x_{ij}[t] \le 1 \quad (i \in \mathcal{N}, k, j \in \mathcal{I}_i, 1 \le t \le T) \, . \tag{5.6.2}$$

That is, node i cannot transmit and receive at the same time.

[2] Interference avoidance in the frequency domain via FDMA can also be developed in the same manner.

Denote C_{ij} the achievable link rate on link $i \to j$. Then, we have $C_{ij} = \frac{1}{T}\sum_{t=1}^{T} R \cdot x_{ij}[t]$.

Joint PHY-Link Constraints. We first give a definition for *residual SINR*, which characterize the SINR value in a sequential fashion under SIC. For a signal from node i to node j in time slot t (from either intended or unintended transmission), we define the residual SINR (or r-SINR) of this signal, r-SINR$_{(i,j)}[t]$, as

$$\begin{aligned}
&\text{r-SINR}_{(i,j)}[t] \\
=&\frac{P_{ij}}{\sum_{k\neq i}\sum_{l\in\mathcal{I}_k} P_{kj}x_{kl}[t] - \sum_{k\neq i}^{P_{kj}>P_{ij}}\sum_{l\in\mathcal{I}_k} P_{kj}x_{kl}[t] + \sigma^2} \\
=&\frac{P_{ij}}{\sum_{k\neq i}^{P_{kj}\leq P_{ij}}\sum_{l\in\mathcal{I}_k} P_{kj}\cdot x_{kl}[t] + \sigma^2}\,.
\end{aligned} \tag{5.6.3}$$

Note that $\sum_{k\neq i}^{P_{kj}\leq P_{ij}}\sum_{l\in\mathcal{I}_k} P_{kj}\cdot x_{kl}[t]$ is the residual interference when node j attempts to decode the signal from node i after subtracting all the stronger received signals from other concurrent transmissions.

To see the coupling of r-SINR with scheduling, note that when $x_{ij}[t] = 1$, we have a successful decoding for the signal from node i to node j under SIC. This implies that

- The r-SINR's of all stronger received signals at node j from other concurrent transmissions are no less than the SINR threshold β.
- The r-SINR of the signal from node i to node j is no less than the SINR threshold β.

More formally, we have following coupling constraints for PHY-Link layers.

$$\begin{aligned}
&\text{If } x_{ij}[t] = 1, \text{ then r-SINR}_{(m,j)}[t] \geq \beta \ (j \in \mathcal{N}, i \in \mathcal{I}_j, m \neq i, n \in \mathcal{I}_m, \\
&\qquad\qquad P_{mj} > P_{ij}, x_{mn}[t] = 1, 1 \leq t \leq T)
\end{aligned} \tag{5.6.4}$$

$$\text{If } x_{ij}[t] = 1, \text{ then r-SINR}_{(i,j)}[t] \geq \beta \ (j \in \mathcal{N}, i \in \mathcal{I}_j, 1 \leq t \leq T). \tag{5.6.5}$$

Flow Routing Constraints. For a set of unicast communication sessions $\mathcal{F}$, denote $r(f)$ the data rate of session $f \in \mathcal{F}$, $s(f)$ and $d(f)$ the source and the destination nodes of session $f \in \mathcal{F}$,

respectively. Denote $r_{ij}(f)$ the amount of rate on link $i \to j$ that is attributed to session $f \in \mathcal{F}$. Then we have the following flow balance. If node i is the source node of session f, i.e., $i = s(f)$, then

$$\sum_{j \in \mathcal{I}_i} r_{ij}(f) = r(f) \quad (f \in \mathcal{F}, i = s(f)) \,. \tag{5.6.6}$$

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

$$\sum_{j \in \mathcal{I}_i}^{j \neq s(f)} r_{ij}(f) = \sum_{k \in \mathcal{I}_i}^{k \neq d(f)} r_{ki}(f) \quad (f \in \mathcal{F}, i \neq s(f), d(f)) \,. \tag{5.6.7}$$

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

$$\sum_{k \in \mathcal{I}_i} r_{ki}(f) = r(f) \quad (f \in \mathcal{F}, i = d(f)) \,. \tag{5.6.8}$$

Note that in the above flow balance equations, we allow flow splitting/merging inside the network, which is more general than single-path flow routing. Further, it can be easily verified that if (5.6.6) and (5.6.7) are satisfied, then (5.6.8) is also satisfied. As a result, it is sufficient to list only (5.6.6) and (5.6.7) in the optimization framework.

Since the aggregate flow rate on any link $i \to j$ cannot exceed the achievable link rate C_{ij}, we have

$$\sum_{f \in \mathcal{F}}^{s(f) \neq j, d(f) \neq i} r_{ij}(f) \leq C_{ij} = \sum_{t=1}^{T} \frac{R}{T} \cdot x_{ij}[t] \quad (j \in \mathcal{N}, i \in \mathcal{I}_j). \tag{5.6.9}$$

5.7 A Formal Optimization Framework

5.7.1 Motivation

Note that the two sets of constraints in (5.6.4) and (5.6.5) are stated in the form of sufficient conditions rather than in the form of mathematical programming that is suitable for problem solving.[3] Therefore, a reformulation of (5.6.4) and (5.6.5) is needed.

[3]By "the form of mathematical programming," we mean that a constraint should be written in the form: $h(\mathbf{x}) \leq 0$ or $h(\mathbf{x}) = 0$, where $\mathbf{x}$ is the set of variables in the constraint and h is a function mapping $\mathbf{x}$ into real space.

As the first step to reformulate (5.6.4), we move $x_{mn}[t] = 1$ out of the range in (5.6.4). By treating $x_{mn}[t] = 1$ as part of the sufficient condition, (5.6.4) can be re-stated as follows:

$$\text{If } (x_{ij}[t] = 1 \text{ and } x_{mn}[t] = 1), \text{ then r-SINR}_{(m,j)}[t] \geq \beta$$
$$(j \in \mathcal{N}, i \in \mathcal{I}_j, m \neq i, n \in \mathcal{I}_m, P_{mj} > P_{ij}, 1 \leq t \leq T)\ . \tag{5.7.1}$$

To combine $x_{ij}[t] = 1$ and $x_{mn}[t] = 1$ into one condition, we can introduce a binary variable, $y_{(i,j)(m,n)}[t]$, as follows.

$$y_{(i,j)(m,n)}[t] = 1 \text{ if and only if } (x_{ij}[t] = 1 \text{ and } x_{mn}[t] = 1)$$
$$(j \in \mathcal{N}, i \in \mathcal{I}_j, m \neq i, n \in \mathcal{I}_m, P_{mj} > P_{ij}, 1 \leq t \leq T).$$

For time slot t, we note that binary variable y has subscripts for four node dimensions, i, j, m, n, which means the number of such y variables could be a very large number. Fortunately, we find that we can remove the last node dimension n and reduce the number of y variables based on the following lemma.

Lemma 5.2. *Statement (5.7.1) is equivalent to the following statement:*

$$\textit{If } \big(x_{ij}[t] = 1 \textit{ and } \textstyle\sum_{n\in\mathcal{I}_m} x_{mn}[t] = 1\big), \textit{ then r-SINR}_{(m,j)}[t]$$
$$\geq \beta \quad (j \in \mathcal{N}, i \in \mathcal{I}_j, m \neq i, P_{mj} > P_{ij}, 1 \leq t \leq T). \tag{5.7.2}$$

Note that the differences between (5.7.1) and (5.7.2) are that $x_{mn}[t] = 1$ in (5.7.1) is replaced by $\sum_{n\in\mathcal{I}_m} x_{mn}[t] = 1$ in (5.7.2) and that $n \in \mathcal{I}_m$ in the range of (5.7.1) disappears in that of (5.7.2).

Proof. We first show that if (5.7.1) holds, then (5.7.2) also holds. If $x_{ij}[t] = 1$ and $\sum_{n\in\mathcal{I}_m} x_{mn}[t] = 1$, then there must exists one node $\hat{n} \in \mathcal{I}_m$ such that

$$x_{m\hat{n}}[t] = 1\ .$$

Combining $x_{ij}[t] = 1$ and $x_{m\hat{n}}[t] = 1$, based on (5.7.1), we have r-SINR$_{(m,j)}[t] \geq \beta$.

Next, we show that if (5.7.2) holds, then (5.7.1) also holds. If $x_{ij}[t] = 1$ and $x_{mn}[t] = 1$, we have

$$\sum_{n \in \mathcal{I}_m} x_{mn}[t] = 1$$

based on (5.6.1). Combining $x_{ij}[t] = 1$ and $\sum_{n \in \mathcal{I}_m} x_{mn}[t] = 1$, based on (5.7.2), we have r-SINR$_{(m,j)}[t] \geq \beta$. □

To simplify (5.7.2), we introduce a new binary variable $\lambda_m[t]$ and define it as follows:

$$\lambda_m[t] = \sum_{n \in \mathcal{I}_m} x_{mn} \quad (m \in \mathcal{N}, 1 \leq t \leq T) \, . \tag{5.7.3}$$

Intuitively, $\lambda_m[t]$ can be regarded as a variable representing whether or not node m is transmitting in time slot t, regardless of to whom it is transmitting. Then, (5.7.2) becomes

$$\begin{aligned} &\text{If } (x_{ij}[t] = 1 \text{ and } \lambda_m[t] = 1), \text{ then r-SINR}_{(m,j)}[t] \geq \beta \\ &(j \in \mathcal{N}, i \in \mathcal{I}_j, m \neq i, P_{mj} > P_{ij}, 1 \leq t \leq T). \end{aligned} \tag{5.7.4}$$

To combine both conditions $x_{ij}[t] = 1$ and $\lambda_m[t] = 1$ into just one condition, we introduce a binary variable $y_{(i,j)(m)}[t]$ as follows:

$$\begin{aligned} &y_{(i,j)(m)}[t] = 1 \text{ if and only if } (x_{ij}[t] = 1 \text{ and } \lambda_m[t] = 1) \\ &(j \in \mathcal{N}, i \in \mathcal{I}_j, m \neq i, n \in \mathcal{I}_m, P_{mj} > P_{ij}, 1 \leq t \leq T) \, . \end{aligned} \tag{5.7.5}$$

Note that variable y only has three node dimensions, i, j, m, which shows that the number of variables in the optimization framework has been decreased. Combining (5.7.5) and (5.7.4), we have

$$\text{If } y_{(i,j)(m)}[t] = 1, \text{ then r-SINR}_{(m,j)}[t] \geq \beta \ (j \in \mathcal{N}, i \in \mathcal{I}_j, m \neq i, P_{mj} > P_{ij}, 1 \leq t \leq T). \tag{5.7.6}$$

Now, (5.6.4) is replaced by (5.7.3), (5.7.5) and (5.7.6). Although (5.7.5) and (5.7.6) are still not yet in the form of mathematical programming, they are ready to be reformulated into such form. In the rest of this section, we show how to reformulate (5.7.5), (5.7.6) and (5.6.5).

5.7.2 Revised PHY-Link Constraints

Based on the definition of new variable $\lambda_m[t]$, we can refine the earlier definition of residual SINR in (5.6.3) as follows.

Definition 5.1. (r-SINR). *For a signal from node i to node j in time slot t (from either intended or unintended transmission), the residual SINR (or r-SINR) of this signal is*

$$\textit{r-SINR}_{(i,j)}[t] = \frac{P_{ij}}{\sum_{k\neq i}^{P_{kj}\leq P_{ij}} P_{kj}\cdot\lambda_k[t]+\sigma^2}\ . \tag{5.7.7}$$

(i) Reformulation of (5.7.5)

Statement (5.7.5) is equivalent to the following two statements:

$$\text{If } (x_{ij}[t]=1 \text{ and } \lambda_m[t]=1), \text{ then } y_{(i,j)(m)}[t]=1\ (j\in\mathcal{N}, i\in\mathcal{I}_j, m\neq i, P_{mj}>P_{ij}, 1\leq t\leq T). \tag{5.7.8}$$

$$\text{If } y_{(i,j)(m)}[t]=1, \text{ then } (x_{ij}[t]=1 \text{ and } \lambda_m[t]=1)\ (j\in\mathcal{N}, i\in\mathcal{I}_j, m\neq i, P_{mj}>P_{ij}, 1\leq t\leq T). \tag{5.7.9}$$

Statement (5.7.8) can be written as

$$y_{(i,j)(m)}[t]\geq x_{ij}[t]+\lambda_m[t]-1\ (j\in\mathcal{N}, i\in\mathcal{I}_j, m\neq i, P_{mj}>P_{ij}, 1\leq t\leq T)\ , \tag{5.7.10}$$

which means that when $x_{ij}[t]=1$ and $\lambda_m[t]=1$, we have $y_{(i,j)(m)}[t]=1$. Statement (5.7.9) can be written as

$$x_{ij}[t]\geq y_{(i,j)(m)}[t] \qquad (j\in\mathcal{N}, i\in\mathcal{I}_j, m\neq i, P_{mj}>P_{ij}, 1\leq t\leq T) \tag{5.7.11}$$

$$\lambda_m[t]\geq y_{(i,j)(m)}[t] \qquad (j\in\mathcal{N}, i\in\mathcal{I}_j, m\neq i, P_{mj}>P_{ij}, 1\leq t\leq T)\ . \tag{5.7.12}$$

Inequalities (5.7.11) and (5.7.12) ensure that when $y_{(i,j)(m)}[t]=1$, we have $x_{ij}[t]=1$ and $\lambda_m[t]=1$.

Now statement (5.7.5) is reformulated as (5.7.10), (5.7.11), and (5.7.12), which are in the form of mathematical programming.

(ii) Reformulation of (5.7.6)

By substituting (5.7.7) to (5.7.6), (5.7.6) becomes

$$\text{If } y_{(i,j)(m)}[t]=1, \text{ then } \frac{P_{mj}}{\sum_{k\neq m}^{P_{kj}\leq P_{mj}} P_{kj}\cdot\lambda_k[t]+\sigma^2}\geq\beta$$
$$(j\in\mathcal{N}, i\in\mathcal{I}_j, m\neq i, P_{mj}>P_{ij}, 1\leq t\leq T)\ ,$$

which is equivalent to

$$P_{mj}-\sum_{k\neq m}^{P_{kj}\leq P_{mj}}\beta P_{kj}\lambda_k[t]-\beta\sigma^2\geq\left(1-y_{(i,j)(m)}[t]\right)D_{ijm}$$
$$(j\in\mathcal{N}, i\in\mathcal{I}_j, m\neq i, P_{mj}>P_{ij}, 1\leq t\leq T), \tag{5.7.13}$$

where D_{ijm} is a lower bound of $P_{mj}-\sum_{k\neq m}^{P_{kj}\leq P_{mj}}\beta P_{kj}\ \lambda_k[t]-\beta\sigma^2$ (e.g., we can set $D_{ijm}=P_{mj}-\sum_{k\neq m}^{P_{kj}\leq P_{mj}}\beta P_{kj}-\beta\sigma^2$). We can verify that when $y_{(i,j)(m)}[t]=1$, (5.7.13) becomes $P_{mj}-\sum_{k\neq m}^{P_{kj}\leq P_{mj}}\beta P_{kj}\cdot\lambda_k[t]-\beta\sigma^2\geq 0$, which is r-SINR$_{(m,j)}[t]\geq\beta$; when $y_{(i,j)(m)}[t]=0$, (5.7.13) becomes $P_{mj}-\sum_{k\neq m}^{P_{kj}\leq P_{mj}}\beta P_{kj}\cdot\lambda_k[t]-\beta\sigma^2\geq D_{ijm}$, which holds by the definition of D_{ijm}.

(iii) Reformulation of (5.6.5)

Following the same token as that in reformulating (5.7.6) into (5.7.13), we can rewrite (5.6.5) as

$$P_{ij}-\sum_{k\neq i}^{P_{kj}\leq P_{ij}}\beta P_{kj}\cdot\lambda_k[t]-\beta\sigma^2\geq(1-x_{ij}[t])H_{ij}\quad(j\in\mathcal{N}, i\in\mathcal{I}_j, 1\leq t\leq T), \tag{5.7.14}$$

where H_{ij} is a lower bound of $P_{ij}-\sum_{k\neq i}^{P_{kj}\leq P_{ij}}\beta P_{kj}\cdot\lambda_k[t]-\beta\sigma^2$ (e.g., we can set $H_{ij}=P_{ij}-\sum_{k\neq i}^{P_{kj}\leq P_{ij}}\beta P_{kj}-\beta\sigma^2$).

5.7.3 Revised Scheduling Constraints

Inspired by the λ-variable's ability to reduce the dimension of y-variable from four to three, we would like to use λ-variable to formulate constraints for half-duplex. We have

$$\frac{1}{\min\{A_j,|\mathcal{I}_j|\}}\sum_{i\in\mathcal{I}_j}x_{ij}[t]+\lambda_j[t]\leq 1\quad(j\in\mathcal{N}, 1\leq t\leq T)\ , \tag{5.7.15}$$

Scheduling:

$\lambda_m[t] = \sum_{n \in \mathcal{I}_m} x_{mn}$ $\quad (m \in \mathcal{N}, 1 \le t \le T)$

$\frac{1}{\min\{A_j, |\mathcal{I}_j|\}} \sum_{i \in \mathcal{I}_j} x_{ij}[t] + \lambda_j[t] \le 1$ $\quad (j \in \mathcal{N}, 1 \le t \le T)$

PHY-Link:

$y_{(i,j)(m)}[t] \ge x_{ij}[t] + \lambda_m[t] - 1$ $\quad (j \in \mathcal{N}, i \in \mathcal{I}_j, m \ne i, P_{mj} > P_{ij}, 1 \le t \le T)$

$x_{ij}[t] \ge y_{(i,j)(m)}[t]$ $\quad (j \in \mathcal{N}, i \in \mathcal{I}_j, m \ne i, P_{mj} > P_{ij}, 1 \le t \le T)$

$\lambda_m[t] \ge y_{(i,j)(m)}[t]$ $\quad (j \in \mathcal{N}, i \in \mathcal{I}_j, m \ne i, P_{mj} > P_{ij}, 1 \le t \le T)$

$P_{mj} - \sum_{k \ne m}^{P_{kj} \le P_{mj}} \beta P_{kj} \lambda_k[t] - \beta\sigma^2 \ge \left(1 - y_{(i,j)(m)}[t]\right) D_{ijm}$ $\quad (j \in \mathcal{N}, i \in \mathcal{I}_j, m \ne i, P_{mj} > P_{ij}, 1 \le t \le T)$

$P_{ij} - \sum_{k \ne i}^{P_{kj} \le P_{ij}} \beta P_{kj} \cdot \lambda_k[t] - \beta\sigma^2 \ge (1 - x_{ij}[t]) H_{ij}$ $\quad (j \in \mathcal{N}, i \in \mathcal{I}_j, 1 \le t \le T)$

Flow routing:

$\sum_{j \in \mathcal{I}_i} r_{ij}(f) = r(f)$ $\quad (f \in \mathcal{F}, i = s(f))$

$\sum_{j \in \mathcal{I}_i}^{j \ne s(f)} r_{ij}(f) = \sum_{k \in \mathcal{I}_i}^{k \ne d(f)} r_{ki}(f)$ $\quad (f \in \mathcal{F}, i \ne s(f), d(f))$

$\sum_{f \in \mathcal{F}}^{s(f) \ne j, d(f) \ne i} r_{ij}(f) \le \sum_{t=1}^{T} \frac{R}{T} \cdot x_{ij}[t]$ $\quad (j \in \mathcal{N}, i \in \mathcal{I}_j)$

Figure 5.6: An optimization framework for joint SIC and interference avoidance for a multi-hop wireless network.

where A_j is an upper bound of the number of signals that node j can decode (see Lemma 5.1) and $|\mathcal{I}_j|$ is the number of neighboring nodes of node j. If node j is receiving from some node, the first term of the Left-Hand-Side in (5.7.15) is greater than 0. Then, $\lambda_j[t]$ must be 0. If node j is transmitting to some node (i.e., $\lambda_j[t] = 1$), then we must have $\frac{1}{|\mathcal{I}_j|}\sum_{i\in\mathcal{I}_j} x_{ij}[t] = 0$, which means that node j is not receiving from any node. Comparing the new half-duplex constraints (5.7.15) (formulated by using λ-variable) to the previously formulated half-duplex constraints (5.6.2), we find that the number of constraints in (5.7.15) is much smaller.

Moreover, due to the definition of variable λ in (5.7.3) and the fact that λ is binary, constraints (5.6.1) are redundant and can be removed from the framework.

5.7.4 Summary

Now we have all the constraints needed in an optimization framework for joint SIC and interference avoidance for a multi-hop wireless network. This framework includes scheduling constraints (5.7.3), (5.7.15), joint PHY-Link constraints (5.7.10), (5.7.11), (5.7.12), (5.7.13), (5.7.14), and flow routing constraints (5.6.6), (5.6.7), (5.6.9). We summarize them in Fig. 5.6.

5.8 A Case Study

The goal of this effort is twofold. First, we want to validate the application of our framework of joint SIC and interference avoidance in solving a practical network optimization problem. Second, we would like to have a closer look at how an interference exploitation scheme such as SIC can optimally interact with an interference avoidance scheme in a multi-hop wireless network.

5.8.1 A Throughput Maximization Problem

Consider a typical throughput maximization problem where we want to maximize the sum of weighted rates of active user sessions in a multi-hop wireless network.[4] We assume each session $f \in \mathcal{F}$ is associated with a weight $w(f)$. Then, our objective is to maximize $\sum_{f\in\mathcal{F}} w(f) \cdot r(f)$. Listing all the constraints summarized in Fig. 5.6, we have the following network throughput maximization problem (TMP).

$$\textbf{TMP:} \quad \max \sum_{f\in\mathcal{F}} w(f) \cdot r(f)$$

$$\text{s.t. All constraints in Fig. 5.6.}$$

TMP is a mixed integer linear program (MILP). Although the theoretical worst-case complexity to a general MILP problem is exponential [39, 98], there exist highly efficient optimality/approximation algorithms (e.g., branch-and-bound with cutting planes [101]) and heuristics (e.g., sequential fixing algorithm [38, 52, 53]) to solve it. Another approach is to apply an off-the-shelf solver (CPLEX [23]), which can successfully handle a moderate-sized network. Since the main goal of this chapter is to advocate joint interference exploitation and avoidance approach for a multi-hop wireless network, we will use CPLEX for this purpose. A more efficient solution to the TMP problem that exploits its specific mathematical structure and physical interpretations can be developed separately and is beyond the scope of this chapter.

5.8.2 A 20-node Example

In this section, we will use a 20-node 3-session network as an example to explain the details of our solution. Another set of results for a 50-node 5-session network will be provided in the next section.

[4] Note that problems with objectives such as maximizing the minimum session rate among all sessions or maximizing a scaling factor of all session rates belong to the same class of problems and can be solved similarly.

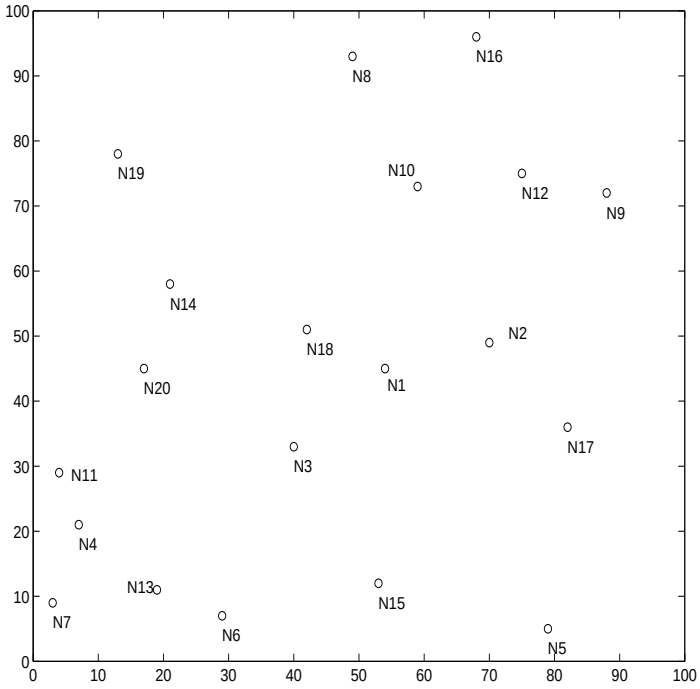


Figure 5.7: The topology of a 20-node network.

Simulation Setting

We consider a randomly generated multi-hop wireless network with 20 nodes, which are distributed in a 100×100 area. For generality, we normalize all units for distance, data rate, bandwidth, and power with appropriate dimensions. The topology of the network is shown in Fig. 5.7. There are three active sessions in the network, with each session's source node, destination node, and weight given in Table 5.2.

The transmission power of each node is set to $P = 1$. For simplicity, we assume that channel gain g_{ij} only includes the path loss between nodes i and j and is given by $g_{ij} = d_{ij}^{-\gamma}$, where d_{ij} is the distance between nodes i and j, and $\gamma = 3$ is the path loss index. The power of ambient noise is $\sigma^2 = 10^{-6}$. There are $T = 10$ time slots in each time frame. The SINR threshold for a successful transmission is $\beta = 1$. When a node i transmits to node j successfully in time slot t (i.e. $x_{ij}[t] = 1$), the achieved data rate is $R = 1$.

Table 5.2: Source node, destination node, and weight of each session in the 20-node network.

Session f	Source Node $s(f)$	Dest. Node $d(f)$	Weight $w(f)$
1	2	11	5.0
2	8	3	6.0
3	19	9	7.0

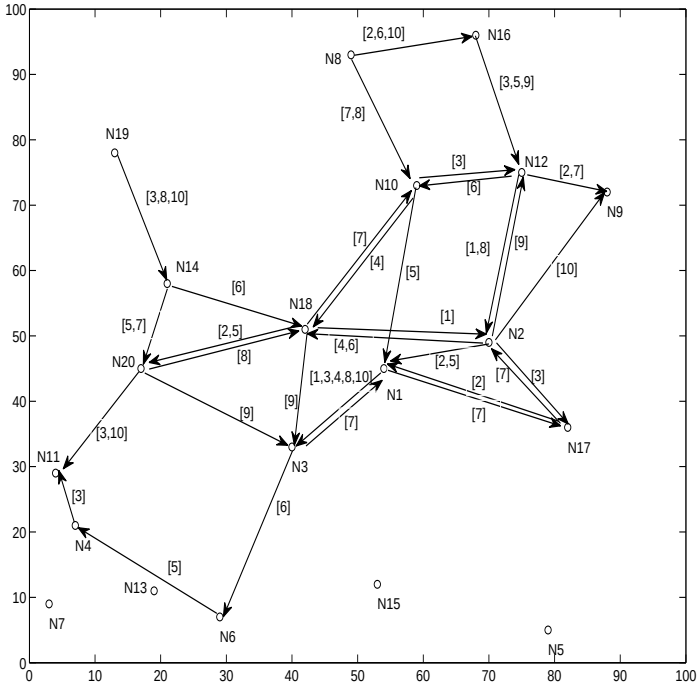


Figure 5.8: Optimal routing and scheduling solution to TMP problem for the 20-node network.

Table 5.3: Active links in each time slot in the optimal solution for the 20-node network.

Time slot	Active links
1	$1 \rightarrow 3$, $12 \rightarrow 2$, $18 \rightarrow 2$
2	$2 \rightarrow 1$, $8 \rightarrow 16$, $12 \rightarrow 9$, $17 \rightarrow 1$, $18 \rightarrow 20$
3	$1 \rightarrow 3$, $2 \rightarrow 17$, $4 \rightarrow 11$, $10 \rightarrow 12$, $16 \rightarrow 12$,$19 \rightarrow 14$, $20 \rightarrow 11$
4	$1 \rightarrow 3$, $2 \rightarrow 18$, $10 \rightarrow 18$
5	$2 \rightarrow 1$, $6 \rightarrow 4$, $10 \rightarrow 1$, $14 \rightarrow 20$, $16 \rightarrow 12$, $18 \rightarrow 20$
6	$1 \rightarrow 17$, $2 \rightarrow 18$, $3 \rightarrow 6$, $8 \rightarrow 16$, $12 \rightarrow 10$, $14 \rightarrow 18$
7	$3 \rightarrow 1$, $8 \rightarrow 10$, $12 \rightarrow 9$, $14 \rightarrow 20$, $17 \rightarrow 2$, $18 \rightarrow 10$
8	$1 \rightarrow 3$, $8 \rightarrow 10$, $12 \rightarrow 2$, $19 \rightarrow 14$, $20 \rightarrow 18$
9	$2 \rightarrow 12$, $10 \rightarrow 1$, $16 \rightarrow 12$, $18 \rightarrow 3$, $20 \rightarrow 3$
10	$1 \rightarrow 3$, $2 \rightarrow 9$, $8 \rightarrow 16$, $19 \rightarrow 14$, $20 \rightarrow 11$

Joint Interference Exploitation and Avoidance

For the 20-node network, we apply CPLEX solver for the TMP formulation. The optimal objective value (maximum weighted sum throughput) is 6.6, with respective data rates for sessions 1, 2 and 3 being 0.3, 0.5 and 0.3. Fig. 5.8 shows the optimal routing and scheduling in the solution, where the numbers in the brackets next to a link show the time slots in a frame when the link is active. For example, $[3, 8, 10]$ next to link $19 \rightarrow 14$ means that this link is active in time slots 3, 8, and 10.

Table 5.3 shows the set of active links in each time slot. Our solution divides different links which are used to support the end-to-end sessions into different time slots so that the set of links in each time slot can successfully coexist (i.e., all links in each time slot satisfy the sequential SINR constraints in (5.3.1)). We use interference avoidance (scheduling) to overcome the limitations of SIC (clearly, the links in Table 5.3 cannot be active in one single time slot). By exploiting the interference through SIC, we are able to activate as many links as possible in a time slot to maximize the network throughput. For example, in time slot 2, both nodes 2 and 17 transmit to node 1 simultaneously.

Some Details

From Table 5.3, we can validate the behavior of SIC quantitatively as follows. SIC allows a node to receive signals from multiple transmitters and reject the interference from other nodes in the same time slot. As an example, we look at the active links in time slot 1 in Table 5.3. In this time slot, links $1 \rightarrow 3$, $12 \rightarrow 2$ and $18 \rightarrow 2$ are active simultaneously. For receiver 2, the signal from node 1 (transmitting to node 3) is an interference to receiver 2, while the signals from nodes 12 and 18 are intended signals. In this example, we will show that receiver 2 rejects the interference from node 1 and receives concurrent transmissions from nodes 12 and 18.

The received signal powers from nodes 1, 12 and 18 at node 2 are $P_{1,2} = 22.29 \times 10^{-5}$, $P_{12,2} = 5.39 \times 10^{-5}$ and $P_{18,2} = 4.52 \times 10^{-5}$, respectively. Receiver 2 first tries to decode the strongest signal, which is from node 1. Note that this is an interference signal. The r-SINR for decoding this signal is

$$\begin{aligned} \frac{P_{1,2}}{P_{12,2} + P_{18,2} + \sigma^2} &= \frac{22.29 \times 10^{-5}}{(5.39 + 4.52 + 0.1) \times 10^{-5}} \\ &= 2.23 > \beta = 1 \ , \end{aligned}$$

which shows that the interference signal from node 1 can be successfully decoded at receiver 2. After subtracting the interference from node 1 from the composite signal (i.e., interference rejection), receiver 2 moves on to decode the second strongest signal, which is from node 12. For this intended signal, its r-SINR is

$$\begin{aligned} \frac{P_{12,2}}{P_{1,2} + P_{18,2} - P_{1,2} + \sigma^2} &= \frac{5.39 \times 10^{-5}}{(4.52 + 0.1) \times 10^{-5}} \\ &= 1.17 > \beta = 1 \ . \end{aligned}$$

Thus, the signal from node 12 can be decoded successfully at receiver 2. Receiver 2 subtracts this signal from node 12 from the remaining composite signal and continues to decode the intended signal from node 18. The r-SINR for for decoding this signal is

$$\frac{P_{18,2}}{\sigma^2} = \frac{4.52 \times 10^{-5}}{10^{-6}} = 45.2 > \beta = 1 \ ,$$

which shows a successful decoding and reception.

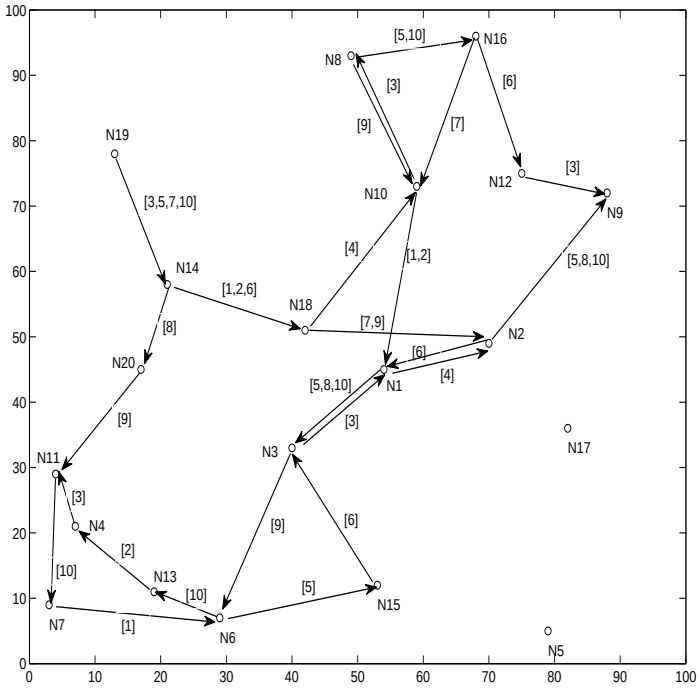


Figure 5.9: The routing and scheduling results under pure interference avoidance model for the 20-node network.

Comparison to Pure Interference Avoidance Model

As a final part of our numerical results for this 20-node network, we compare our optimal result to the TMP problem to the optimal result under pure interference avoidance model (i.e., SIC is not employed and all interference in the network is handled by scheduling). The problem formulation under pure interference avoidance model (called TMP-Pure) is given as follows, which is also a MILP problem.

Under the pure interference avoidance model, only scheduling is employed (no SIC). The joint PHY-Link constraints will change. When decoding a signal from i to node j, we treat all the signals from other transmitting nodes as noise. Then, for a successful transmission from node i to node j in time slot t (i.e., $x_{ij}[t] = 1$), we need the following statement:

$$\text{If } x_{ij}[t] = 1, \text{ then } \frac{P_{ij}}{\sum_{k \neq i} P_{kj} \lambda_k[t] + \sigma^2} \geq \beta \ (j \in \mathcal{N}, i \in \mathcal{I}_j, 1 \leq t \leq T).$$

The above statement can be written as

$$P_{ij} - \sum_{k \neq i} \beta P_{kj} \lambda_k[t] - \beta\sigma^2 \geq (1 - x_{ij}[t]) M_{ij} \quad (j \in \mathcal{N}, i \in \mathcal{I}_j, 1 \leq t \leq T), \tag{5.8.1}$$

where M_{ij} is a lower bound of $P_{ij} - \beta \sum_{k \neq i} P_{kj} \lambda_k[t] - \beta\sigma^2$. Under this model, TMP-Pure has the same scheduling and flow routing constraints as that of problem TMP. Then, the formulation of TMP-Pure is as follows.

$$\begin{aligned}
&\max \sum_{f \in \mathcal{F}} w(f) \cdot r(f) \\
&\text{s.t. Constraints } (5.6.6), (5.6.7), (5.6.9), (5.7.3), (5.7.15), (5.8.1) \\
&\quad x_{ij}[t], \lambda_i[t] \in \{0, 1\} \ (i \in \mathcal{N}, j \in \mathcal{I}_i, 1 \leq t \leq T) \\
&\quad r(f), r_{ij}(f) \geq 0 \quad (f \in \mathcal{F}, i \in \mathcal{N}, j \in \mathcal{I}_i)
\end{aligned}$$

The formulated problem TMP-Pure is also a mixed integer linear program (MILP). Again, we use CPLEX to solve TMP-Pure for the same 20-node network.

The optimal objective value (maximum weighted sum of throughput) is now 4.5 (vs. 6.6 for TMP), with the data rates for the three sessions being 0.1, 0.2 and 0.4, respectively. In other words, comparing to the pure interference avoidance model, joint interference exploitation-avoidance can increase throughput by $\frac{6.6-4.5}{4.5} = 47\%$.

The optimal routing and scheduling results are shown in Fig. 5.9. The active links in each time slot are given in Table 5.4. We now compare Fig. 5.9 and Table 5.4 to Fig. 5.8 and Table 5.3, respectively. It is clear that without SIC, fewer number of links are active in a pure interference avoidance solution.

5.8.3 A 50-node Example

In this section, we consider a 50-node 5-session network which is distributed in a 150×150 area. The topology of the network is shown in Fig. 5.10. For the five active sessions, the source node, destination node, and weight of each session are given in Table 5.5.

Table 5.4: Active links in each time slot under pure interference avoidance model for the 20-node network.

Time slot	Active links
1	$7 \rightarrow 6$, $10 \rightarrow 1$, $14 \rightarrow 18$
2	$10 \rightarrow 1$,$13 \rightarrow 4$, $14 \rightarrow 18$
3	$3 \rightarrow 1$,$4 \rightarrow 11$,$10 \rightarrow 8$,$12 \rightarrow 9$,$19 \rightarrow 14$
4	$1 \rightarrow 2$, $18 \rightarrow 10$
5	$1 \rightarrow 3$, $2 \rightarrow 9$, $6 \rightarrow 15$, $8 \rightarrow 16$, $19 \rightarrow 14$
6	$2 \rightarrow 1$, $14 \rightarrow 18$, $15 \rightarrow 3$, $16 \rightarrow 12$
7	$16 \rightarrow 10$, $18 \rightarrow 2$, $19 \rightarrow 14$
8	$1 \rightarrow 3$, $2 \rightarrow 9$, $14 \rightarrow 20$
9	$3 \rightarrow 6$, $8 \rightarrow 10$, $18 \rightarrow 2$,$20 \rightarrow 11$
10	$1 \rightarrow 3$,$2 \rightarrow 9$, $6 \rightarrow 13$,$8 \rightarrow 16$

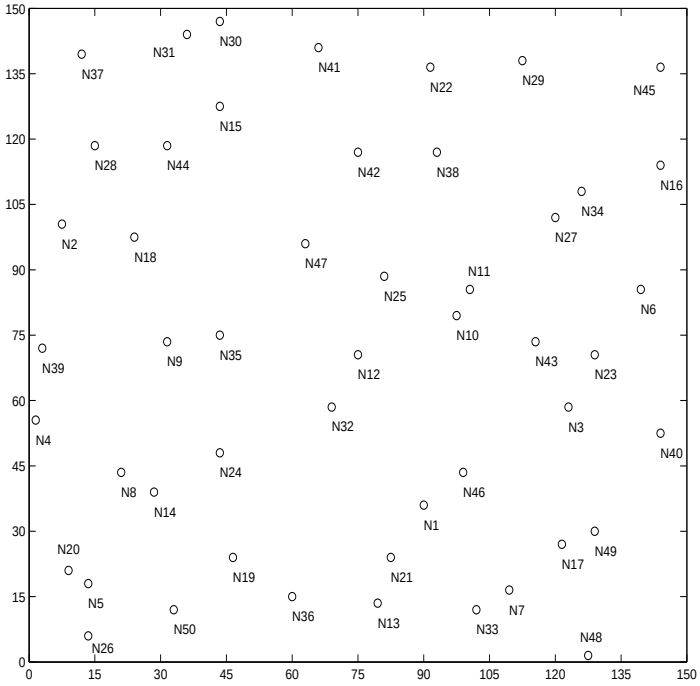


Figure 5.10: The topology of a 50-node network.

Table 5.5: Source node, destination node, and weight of each session in the 50-node network.

Session f	Source Node $s(f)$	Dest. Node $d(f)$	Weight $w(f)$
1	15	29	7.0
2	40	10	6.0
3	38	35	10.0
4	4	19	8.0
5	9	7	9.0

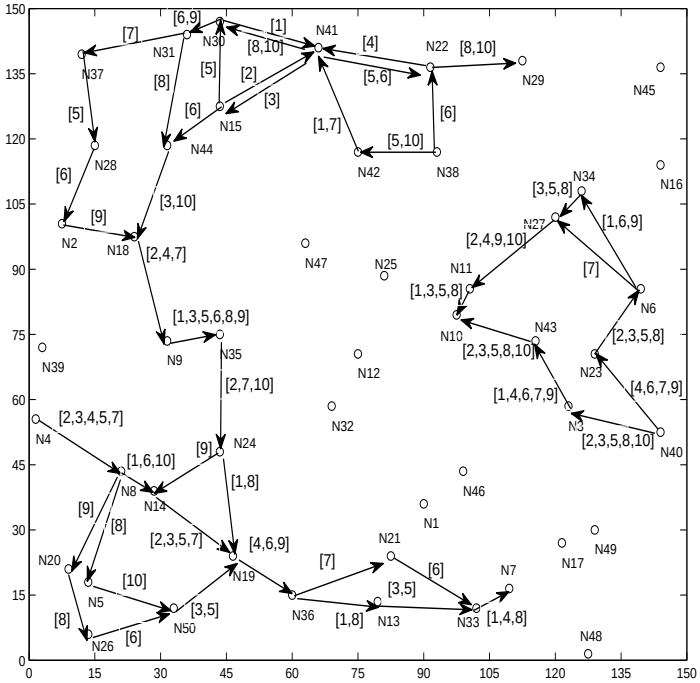


Figure 5.11: Optimal routing and scheduling solution to TMP problem for the 50-node network.

Table 5.6: Active links in each time slot in the optimal solution for the 50-node network.

Time slot	Active links
1	$3 \to 43$, $6 \to 34$, $8 \to 14$, $9 \to 35$, $11 \to 10$, $24 \to 19$, $30 \to 41$, $33 \to 7$, $36 \to 13$, $42 \to 41$
2	$4 \to 8$, $14 \to 19$, $15 \to 41$, $18 \to 9$, $23 \to 6$, $27 \to 11$, $35 \to 24$, $40 \to 3$, $43 \to 10$
3	$4 \to 8$, $9 \to 35$, $11 \to 10$, $13 \to 33$, $14 \to 19$, $23 \to 6$, $34 \to 27$, $40 \to 3$, $41 \to 15$, $43 \to 10$ $44 \to 18$, $50 \to 19$
4	$3 \to 43$, $4 \to 8$, $18 \to 9$, $19 \to 36$, $22 \to 41$, $27 \to 11$, $33 \to 7$, $40 \to 23$
5	$4 \to 8$, $9 \to 35$, $11 \to 10$, $13 \to 33$, $14 \to 19$, $15 \to 30$, $23 \to 6$, $34 \to 27$, $37 \to 28$, $38 \to 42$ $40 \to 3$, $41 \to 22$, $43 \to 10$, $50 \to 19$
6	$3 \to 43$, $6 \to 34$, $8 \to 14$, $9 \to 35$, $15 \to 44$, $19 \to 36$, $21 \to 33$, $26 \to 50$, $28 \to 2$, $30 \to 31$ $38 \to 22$, $40 \to 23$, $41 \to 22$
7	$3 \to 43$, $4 \to 8$, $6 \to 27$, $14 \to 19$, $18 \to 9$, $31 \to 37$, $35 \to 24$, $36 \to 21$, $40 \to 23$, $42 \to 41$
8	$8 \to 5$, $9 \to 35$, $11 \to 10$, $20 \to 26$, $22 \to 29$, $23 \to 6$, $24 \to 19$, $31 \to 44$, $33 \to 7$, $34 \to 27$ $36 \to 13$, $40 \to 3$, $41 \to 30$, $43 \to 10$
9	$2 \to 18$, $3 \to 43$, $6 \to 34$, $8 \to 20$, $9 \to 35$, $19 \to 36$, $24 \to 14$, $27 \to 11$, $30 \to 31$, $40 \to 23$
10	$5 \to 50$, $8 \to 14$, $22 \to 29$, $27 \to 11$, $35 \to 24$, $38 \to 42$, $40 \to 3$, $41 \to 30$, $43 \to 10$, $44 \to 18$

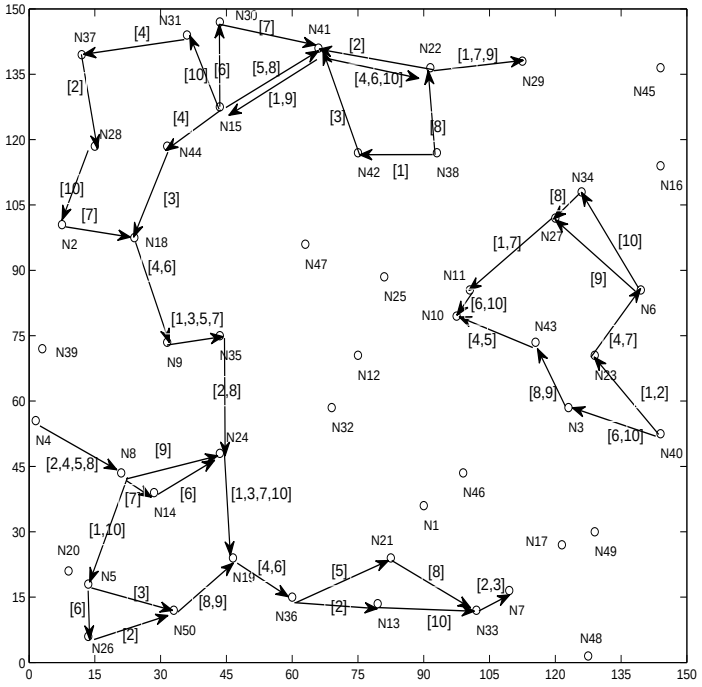


Figure 5.12: The routing and scheduling results under pure interference avoidance model for the 50-node network.

Again, we apply CPLEX solver for the TMP formulation. The optimal objective value (maximum weighted sum throughput) is 16.5, with respective data rates for sessions 1, 2, 3, 4, and 5 being 0.2, 0.9, 0.3, 0.5, and 0.3. Fig. 5.11 shows the optimal routing and scheduling in the solution. Table 5.6 shows the set of active links in each time slot.

For comparison, we use CPLEX to solve TMP-Pure for the same 50-node network. The optimal objective value (maximum weighted sum of throughput) is now 11.5 (vs. 16.5 for TMP), with the data rates for the five sessions being 0.3, 0.4, 0.2, 0.4, and 0.2, respectively. Comparing to the pure interference avoidance model, joint interference exploitation-avoidance can increase throughput by $\frac{16.5-11.5}{11.5} = 43.5\%$.

The optimal routing and scheduling results are shown in Fig. 5.12. The active links in each time slot are given in Table 5.7. Compare Fig. 5.12 and Table 5.7 to Fig. 5.11 and Table 5.6, respectively. We can also see that without SIC, fewer number of links are active in a pure interference avoidance

Table 5.7: Active links in each time slot under pure interference avoidance model for the 50-node network.

Time slot	Active links
1	$8 \to 5$, $9 \to 35$, $22 \to 29$, $24 \to 19$, $27 \to 11$, $38 \to 42$, $40 \to 23$, $41 \to 15$
2	$4 \to 8$, $22 \to 41$, $26 \to 50$, $33 \to 7$, $35 \to 24$, $36 \to 13$, $37 \to 28$, $40 \to 23$
3	$5 \to 50$, $9 \to 35$, $24 \to 19$, $33 \to 7$, $42 \to 41$, $44 \to 18$
4	$4 \to 8$, $15 \to 44$, $18 \to 9$, $19 \to 36$, $23 \to 6$, $31 \to 37$, $41 \to 22$, $43 \to 10$
5	$4 \to 8$, $9 \to 35$, $15 \to 41$, $36 \to 21$, $43 \to 10$
6	$5 \to 26$, $11 \to 10$, $14 \to 24$, $15 \to 30$, $18 \to 9$, $19 \to 36$, $40 \to 3$, $41 \to 22$
7	$2 \to 18$, $8 \to 14$, $9 \to 35$, $22 \to 29$, $23 \to 6$, $24 \to 19$, $27 \to 11$, $30 \to 41$
8	$3 \to 43$, $4 \to 8$, $15 \to 41$, $21 \to 33$, $34 \to 27$, $35 \to 24$, $38 \to 22$, $50 \to 19$
9	$3 \to 43$, $6 \to 27$, $8 \to 24$, $22 \to 29$, $41 \to 15$, $50 \to 19$
10	$6 \to 34$, $8 \to 5$, $11 \to 10$, $13 \to 33$, $15 \to 31$, $24 \to 19$, $28 \to 2$, $40 \to 3$, $41 \to 22$

solution.

5.9 Chapter Summary

Interference avoidance has long been the dominant strategy for wireless medium access. Advances at the physical layer show that interference exploitation has the potential to improve capacity significantly. However, we show that interference exploitation such as SIC has its limitations and cannot handle interference alone, particularly in a complex multi-hop wireless network. In this chapter, we advocated a joint interference exploitation and avoidance approach to handle interference in a multi-hop wireless network. The proposed joint approach combines the best of both worlds while avoids each other's pitfalls. We discussed new challenges that such a joint approach would face in a multi-hop wireless network and proposed a formal optimization framework, with cross-layer formulation of physical, link, and network layers. This optimization framework paves the way to study a broad class of network throughput optimization problems. As a case study, we

demonstrated how to apply such a framework to a network throughout optimization problem. Our numerical results affirmed the efficacy of this framework and gave insights on the optimal interaction between interference exploitation and interference avoidance. The findings in this chapter made a concrete step forward to position a joint interference exploitation and avoidance approach as the main stream approach to handle interference in wireless networks in the near future.

Chapter 6

Throughput Optimization with Network-wide Energy Constraint

6.1 Introduction

With the proliferation of wireless networks, the concern of energy consumption is becoming increasingly important for network operators. Conserving network-wide energy consumption not only can help reducing CO_2 emissions [30] and protect the environment, but can also significantly reduce the operating cost for network providers [31, 93]. Consequently, in addition to pressure from government regulating agencies and citizen environmentalists, network operators now have every incentive to tackle energy conservation in their networks. As a result, energy conservation has quickly evolved from merely an attractive but secondary feature to a key design benchmark. The so-called *Green communications* has also come to the center stage of both national and international research agenda [21, 129, 130] and consequently research efforts in this area is flourishing [71, 83].

In this chapter, we study network-wide energy conservation problem in a multi-hop wireless network which we hope will offer insights to both network operators and end users. Specifically,

in the first part of this work, we will show how to maximize network throughput under a given network-wide total energy consumption budget. This may correspond to a scenario where a network operator has a budget on total energy consumption. In the second part, we generalize the problem in the first part by studying how to optimize both network throughput and network-wide energy consumption through a multicriteria optimization framework. This allows us to characterize the trend of throughput when the total energy consumption budget changes.

We recognize that there is a wealth of literature on optimizing network throughput with energy considerations. A major branch of these prior efforts followed various heuristic approaches in developing physical, link, and network layer schemes and algorithms (see, e.g., [110, 128]). This is in contrast to our work in this chapter, which follows a formal optimization framework with the goal of offering performance guarantee of the final solution.

We recognize that there is a wealth of literature on optimizing network throughput with energy considerations. A major branch of these prior efforts followed various heuristic approaches in developing physical, link, and network layer schemes and algorithms (see, e.g., [110, 128]). This is in contrast to our work in this chapter, which follows a formal optimization framework with the goal of offering performance guarantee of the final solution.

Within the branch of related work that followed formal optimization framework in studying network throughput maximization with energy consideration (see, e.g., [4, 13, 44, 51, 102, 103]), we find that most of these works only considered per-link power constraint or per-node power constraint. Although these constraints are important to characterize local energy consumption, it is not clear how to extend results for local link/node energy conservation to *network-wide* energy conservation, due to the complex inter-dependencies among the layers. Therefore, these prior results cannot directly benefit network operators, who are more concerned with total network-wide energy consumption.

Our work is complementary to a branch of previous work that addressed how to minimize network-wide energy consumption while satisfying some traffic demands (see, e.g., [9, 24, 74, 88, 89]). These works are orthogonal to the problem that we shall study in the first part of this chapter.

It will soon be clear that our mathematical formulation and proposed solution differ from all these seemingly similar efforts. Further, in the second part of this chapter, we consider joint optimization of throughput and network-wide energy, which explores the domain of multi-criteria optimization that is not well studied in the wireless networking community. In our recent work in [57], we explored multicriteria optimization of network energy and throughput. However, power control was not considered in [57]. In this work, we shall consider power control at each node, which is more interesting.

The main contributions of this chapter are the following:

- First, we study how to maximize network throughput under a total network-wide energy consumption constraint. We show that this problem involves both network and physical layer variables and can be formulated as a mixed-integer nonlinear program (MINLP). To solve this problem efficiently, we propose a novel piece-wise linear approximation to transform the nonlinear constraints into linear constraints. We prove that the solution developed under this linear approximation is near-optimal in the sense that the performance gap between our solution and the optimal solution (despite unknown) can be made arbitrary narrow depending on required accuracy.

- Second, we generalize the problem in the first part by exploring joint optimization of both network throughput and network energy consumption via a multicriteria optimization framework, i.e., *maximizing* network throughput while *minimizing* network-wide energy consumption. We find that all the weakly Pareto-optimal points characterize an optimal throughput-energy curve. This curve shows how the maximum network throughput changes as total network-wide energy budget changes. We offer some interesting properties of this optimal throughput-energy curve that are useful to both network operators and end users.

The remainder of this chapter is organized as follows. In Section 6.2, we describe our network model. In Section 6.3, we study how to maximize network throughput under a given total network-wide energy budget. In Section 6.4, we study how to optimize both network throughout and energy

under a multicriteria framework. Section 6.5 presents some numerical results that illustrate our theoretical findings. Section 6.6 concludes this chapter.

6.2 Network Model

Consider a multi-hop wireless ad hoc network, represented by a directed graph $\mathcal{G} = \{\mathcal{N}, \mathcal{L}\}$, where $\mathcal{N}$ and $\mathcal{L}$ are the sets of nodes and directional links, respectively. A link between two nodes exists if and only if the distance between the two is within a certain transmission range. If two nodes are not within one-hop of each other, then a node has to resort to multi-hop to relay messages. We assume orthogonal channels on all links (similar to that in [16, 65, 82]). This can be done by some interference avoidance mechanism (e.g., OFDMA). Note that orthogonal channels do not require as many channels as the number of active links in the network since one can reuse channels on links that are spatially far away from each other. This is called spatial reuse and is commonly used in wireless networks to improve channel efficiency. Note that designing a channel assignment algorithm to achieve orthogonality has been well studied in the literature and its discussion is beyond the scope of this chapter.

We assume there is a set of $\mathcal{F}$ active (unicast) communication sessions in the network. Denote $s(f)$ and $d(f)$ the source and destination nodes of session $f \in \mathcal{F}$, respectively. To differentiate the importance of these user sessions, each session f is assigned a weight $w(f)$. Denote $r(f)$ the data rate of session f. The network throughput U in this chapter is represented by the sum of weighted session rates, which is $\sum_{f \in \mathcal{F}} w(f) \cdot r(f)$. Table 6.1 lists all the notations in this chapter.

6.2.1 Energy Consumption and Power Control

When a wireless link is active for communications, its energy consumption includes transmission power and device power [25, 84], where transmission power is for data transmission over a distance and device power is consumed by device electronics for encoding, modulation, decoding,

Table 6.1: Notation in Chapter 6.

Symbol	Definition
B_l	Channel bandwidth on link l
c_l	Capacity of link l
d_l	Distance between link l's transmitting node and receiving node
$d(f)$	Destination node of session $f \in \mathcal{F}$
$\mathcal{F}$	The set of user sessions in the network
h_l	Channel gain on link l
$\mathcal{L}$	The set of links in the network
$\mathcal{L}_i^{\text{In}}$	The set of incoming links at node i
$\mathcal{L}_i^{\text{Out}}$	The set of outgoing links at node i
$\mathcal{N}$	The set of nodes in the network
p_l	Transmission power of link l
P_d	The circuits power consumption of an active link
P	$= \sum_{l \in \mathcal{L}} (p_l + y_l P_d)$, network-wide energy consumption rate
P_{net}	Network-wide energy budget
$r(f)$	Data rate of session $f \in \mathcal{F}$
$r_l(f)$	Data rate on link l that is attributed to session f
$s(f)$	Source node of session f
U	$= \sum_{m \in \mathcal{M}} w(f) r(f)$, the network throughput
$w(f)$	A weight assigned to session $f \in \mathcal{F}$
y_l	A binary variable indicating whether or not link l is active
η	Ambient Gaussian noise density

demodulation, etc. Denote P_d as device power, which we assume is a constant if link is active. Denote p_l the transmission power on link l, which is a tunable (variable) system parameter.

Denote y_l a binary variable indicating whether or not link l is active, i.e.,

$$y_l = \begin{cases} 1 & \text{if link } l \text{ is active;} \\ 0 & \text{otherwise.} \end{cases}$$

The energy consumption rate of link l, including transmission power and device power, is $p_l + y_l P_d$.

Assume that the maximum transmission power of a node is $P_{\max}$. Then, we have the following relationship between p_l and y_l:

$$p_l \leq y_l \cdot P_{\max} \quad (l \in \mathcal{L}) \ . \tag{6.2.1}$$

For all active links at a node, we have the following node-level transmission power constraint:

$$\sum_{l \in \mathcal{L}_i^{\text{Out}}} p_l \leq P_{\max} \quad (i \in \mathcal{N}) \ , \tag{6.2.2}$$

where $\mathcal{L}_i^{\text{Out}}$ is the set of potential outgoing links at node i.

Denote P as the total energy consumption rate on all active links in the network. Then, the network-wide energy consumption rate P can be written as $P = \sum_{l \in \mathcal{L}} (p_l + y_l P_d)$.

6.2.2 Routing and Link Capacity

To transport data from a source node to its destination node that is more than one-hop away, multi-hop relaying is necessary. Since single-path flow routing is overly restrictive and is unlikely to offer optimal solution, we allow flow splitting so that data can be delivered on multi-path routes. We model multi-path flow routing as follows. Denote $r_l(f)$ the amount of flow rate on link l that is attributed to session $f \in \mathcal{F}$. Denote $\mathcal{L}_i^{\text{In}}$ the set of potential incoming links at node i. If node i is the source node of session f, i.e., $i = s(f)$, then

$$\sum_{l \in \mathcal{L}_i^{\text{Out}}} r_l(f) = r(f) \ . \tag{6.2.3}$$

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

$$\sum_{l \in \mathcal{L}_i^{\mathrm{Out}}}^{l \neq (i,s(f))} r_l(f) = \sum_{m \in \mathcal{L}_i^{\mathrm{In}}}^{m \neq (d(f),i)} r_m(f) \ . \tag{6.2.4}$$

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

$$\sum_{l \in \mathcal{L}_i^{\mathrm{In}}} r_l(f) = r(f) \ . \tag{6.2.5}$$

It can be easily verified that if (6.2.3) and (6.2.4) are satisfied, then (6.2.5) must be satisfied. As a result, it is sufficient to list only (6.2.3) and (6.2.4) in the formulation.

Under the above flow routing scheme, the aggregate flow rate at link l is $\sum_{f \in \mathcal{F}} r_l(f)$. Since aggregate flow rate on any link cannot exceed the link's capacity, we have the following link capacity constraint:

$$\sum_{f \in \mathcal{F}} r_l(f) \leq c_l \quad (l \in \mathcal{L}) \ , \tag{6.2.6}$$

where c_l is the capacity on link l. Given that we are employing orthogonal channels among the links in the network, we have:

$$c_l = B_l \log_2(1 + \frac{p_l \cdot h_l}{\eta B_l}) \ , \tag{6.2.7}$$

where B_l is the bandwidth of link l under a given channel assignment, h_l is channel gain between the transmitter and receiver of link l and η is the ambient Gaussian noise density. Combining (6.2.6) and (6.2.7), we have:

$$\sum_{f \in \mathcal{F}} r_l(f) \leq B_l \log_2(1 + \frac{p_l \cdot h_l}{\eta B_l}) \quad (l \in \mathcal{L}) \ . \tag{6.2.8}$$

Note that constraint (6.2.8) couples network flow variables (i.e., $r_l(f)$) and physical layer power variable p_l.

6.3 Throughput Maximization Under Network-wide Energy Constraint

In this section, we study how to maximize network throughput under a given network-wide energy budget. This problem is motivated by the scenario where we have a strict total energy consumption limit in the network (e.g., due to a given operating budget on energy). The question that we pose is: Given the network-wide energy operating budget P_{net}, i.e.,

$$P = \sum_{l \in \mathcal{L}} (p_l + y_l P_d) \leq P_{\text{net}}, \tag{6.3.1}$$

how to adjust the power on each link and multi-path routing for each session so that the maximum network throughput is achieved?

Mathematically, this problem can be formulated as follows:

$$\begin{aligned} \textbf{OPT:} \quad \max \quad & U = \sum_{f \in \mathcal{F}} w(f) r(f) \\ \text{s.t.} \quad & \text{Constraints } (6.2.1), (6.2.2), (6.2.3), (6.2.4), (6.2.8), (6.3.1) \\ & \text{Variables } y_l \in \{0,1\}, p_l, r_l(f), r(f) \geq 0 \ (l \in \mathcal{L}, f \in \mathcal{F}) \ , \end{aligned}$$

where y_l is a binary variable, p_l, $r(f)$ and $r_l(f)$ are continuous variables and all the other parameters are constants. OPT is a mixed-integer nonlinear program (MINLP), which in general is NP-hard [39]. Note that the network-wide energy constraint complicates overall problem by bringing in integer variables.

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 [32], branch-and-bound [45], extended cutting plane methods [127], and generalized Benders' decomposition [41]). But these techniques do not exploit our problem-specific structures and properties, and hence can only handle small-size problems.

In this chapter, we exploit the structure of our MINLP problem and develop a novel near-optimal solution with performance guarantee. Note that in OPT's formulation, the only set of nonlinear constraints are the link capacity constraints in (6.2.8), which involve the $\log$ function. To address this problem, we propose a piece-wise linear approximation technique to transform the nonlinear constraints to linear constraints. Our main idea is as follows. We first use a set of linear segments to approximate the $\log$ term in (6.2.8) and guarantee the linear approximation error will not exceed a threshold ϵ. Subsequently, the nonlinear constraints in OPT are replaced by a set of linear constraints. Denote the linearized optimization problem as OPT-R, which is a MILP problem. Since MILP problems are much easier than MINLP problems, we can apply a solver such as CPLEX [23] to obtain a solution efficiently.

We will show that solving OPT-R can give us a near-optimal solution to the original problem OPT. Denote γ as desired performance gap of our 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 performance gap γ and the linear approximation error ϵ (see details in Section 6.3.2). Specifically, for a desired performance gap γ, we compute the maximum allowed linear approximation error ϵ. After obtaining ϵ, we can compute the linear approximation constraints and construct OPT-R (see details in Section 6.3.1). Solving the OPT-R will give us a near-optimal solution with performance guarantee γ. We summarize the above steps in Fig. 6.1. In the rest of this section, we fill in the details of these steps.

6.3.1 Piece-wise Linear Approximation

The nonlinear constraint in (6.2.8) can be written as

$$\sum_{f \in \mathcal{F}} r_l(f) \leq \frac{B_l}{\ln 2} \ln(1 + \frac{p_l \cdot h_l}{\eta B_l}). \tag{6.3.2}$$

To simplify notation, denote

$$s_l = \frac{p_l h_l}{\eta B_l} \, . \tag{6.3.3}$$

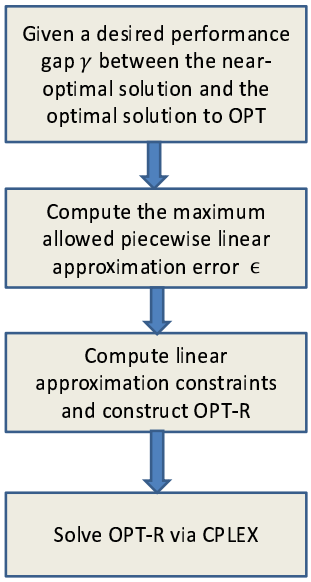


Figure 6.1: A flow chart to develop a near-optimal solution to OPT.

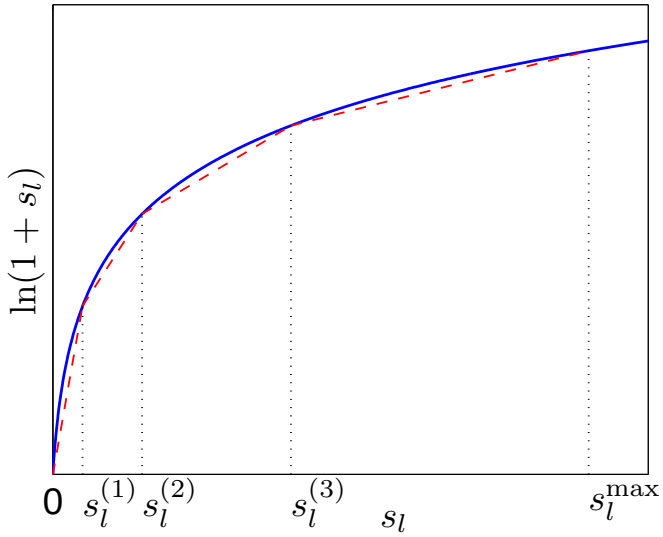


Figure 6.2: An illustration of piece-wise linear approximation with four linear segments.

Then, the nonlinear term in (6.3.2) can be written as $\ln(1+s_l)$. The range of s_l is $[0, s_l^{\max}]$, with $s_l^{\max} = (P_{\max} h_l)/(\eta B_l)$. Our piece-wise linear approximation is to use a set of consecutive linear segments to approximate $\ln(1+s_l)$ for $s_l \in [0, s_l^{\max}]$ (see Fig. 6.2). Denote ϵ the maximum allowed error of this linear approximation. Denote K_l the number of linear segments that is needed to meet this error requirement. (K_l will be determined later.) Denote $s_{l,0}, s_{l,1}, \ldots, s_{l,K_l}$ the X-axis values of the endpoints of these K segments, with $s_{l,0} = 0$ and $s_{l,K_l} = s_l^{\max}$.

A naive approach to generate a linear approximation is making $s_l^{(k)}, k = 0, \ldots, K_l$, evenly distributed between $[0, s_l^{\max}]$. When setting K_l sufficiently large, the linear approximation error requirement will be satisfied. Although this approach is straightforward and easy to implement, it will generate too many linear segments to approximate $\ln(1+s_l)$. Note that the derivative of curve $\ln(1+s_l)$ decreases as s_l increases. This motivates us to enlarge the size of an interval as s_l increases. Thus, we want to pursue an algorithm that optimally divides the K_l intervals within $[0, s_l^{\max}]$. By "optimally", we refer to finding the *minimum* K_l such that the maximum approximation error of each line segment is no more than ϵ.

Denote $m_l^{(k)}$ as the slope of the k-th linear segment, i.e.,

$$m_l^{(k)} = \frac{\ln(1+s_l^{(k)}) - \ln(1+s_l^{(k-1)})}{s_l^{(k)} - s_l^{(k-1)}}. \tag{6.3.4}$$

Denote $g_l^{(k)}(s_l)$ as the k-th linear approximation segment (see Fig. 6.3), which can be represented as follows:

$$g_l^{(k)}(s_l) = m_l^{(k)} \cdot \left(s_l - s_l^{(k-1)}\right) + \ln\left(1 + s_l^{(k-1)}\right), \text{ for } s_l^{(k-1)} \le s_l \le s_l^{(k)}. \tag{6.3.5}$$

Our algorithm computes the values of $s_l^{(0)}, \ldots, s_l^{(K_l)}$ sequentially (for a given ϵ) based on Algorithm 6.1 as follows.

Algorithm 6.1. *Initialization:* $k := 0$ *and* $s_l^{(0)} := 0$.

1. $k := k + 1$.
2. *Compute* $m_l^{(k)}$ *satisfying*

$$-\ln(m_l^{(k)}) + m_l^{(k)}(1 + s_l^{(k-1)}) - 1 - \ln(1 + s_l^{(k-1)}) = \epsilon. \tag{6.3.6}$$

3. *After obtaining $m_l^{(k)}$, compute $s_l^{(k)}$ satisfying (6.3.4).*

4. *If $s_l^{(k)} < s_l^{\max}$, go back to Step 1.*

5. *$K_l := k$; $s_l^{(K_l)} := s_l^{\max}$.*

6. *Update $m_l^{(K_l)}$ using (6.3.4).*

The values of $m_l^{(k)}$ in (6.3.6) and $s_l^{(k)}$ in (6.3.4) can be solved by numerical methods such as bisection method or Newton's method [94, Chapter 2].

Our linear approximation method (Algorithm 6.1) satisfies the linear approximation error requirement with the minimum number of linear segments to approximate $\ln(1+s_l)$ for $s_l \in [0, s_l^{\max}]$. We formalize these two claims in the following two lemmas.

Lemma 6.1. *For the piece-wise linear approximation generated by Algorithm 6.1, the maximum approximation error of each linear segment is at most ϵ.*

Proof. Denote $\epsilon_l^{(k)}$ the maximum linear approximation error for the k-th linear segment, i.e.,

$$\epsilon_l^{(k)} = \max_{s_l^{(k-1)} \le s_l \le s_l^{(k)}} \left| \ln(1+s_l) - g_l^{(k)}(s_l) \right| = \max_{s_l^{(k-1)} \le s_l \le s_l^{(k)}} \left\{ \ln(1+s_l) - g_l^{(k)}(s_l) \right\},$$

where the equality holds since $\ln(1+s_l)$ is a convex function of s_l and all linear segments lie beneath the $\ln(1+s_l)$ curve.

Consider the k-th linear segment. Referring to Fig. 6.3, we can move $g_l^{(k)}(s_l)$ upward until it is tangential to the $\ln(1+s_l)$ curve. It is easy to see that the tangential point achieves the maximum approximation error $\epsilon_l^{(k)}$. Denote $\hat{s}_l^{(k)}$ the X-axis value of that tangential point. Since the derivative of $\ln(1+s_l)$ is $\frac{1}{1+s_l}$, we have $\frac{1}{1+\hat{s}_l^{(k)}} = m_l^{(k)}$, i.e,

$$\hat{s}_l^{(k)} = \frac{1}{m_l^{(k)}} - 1, \tag{6.3.7}$$

where $m_l^{(k)}$ is slope of linear segment $g_l^{(k)}(s_l)$. Therefore, the maximum approximation error $\epsilon_l^{(k)}$

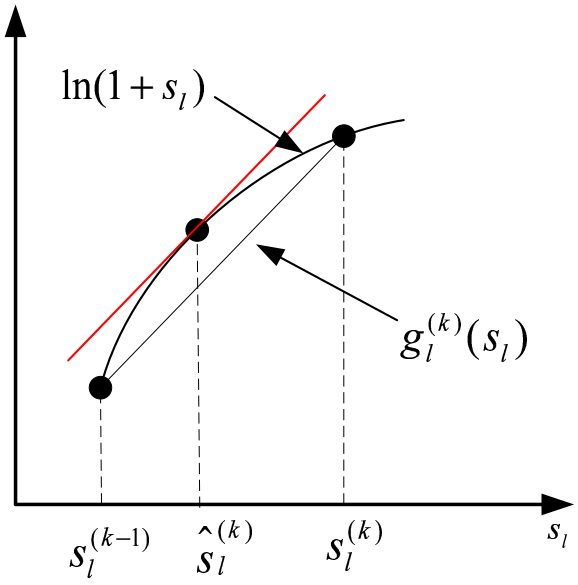


Figure 6.3: An illustration of the maximum approximation error for the k-th linear segment.

can be written as

$$\begin{aligned}\epsilon_l^{(k)} &= \ln(1+\hat{s}_l^{(k)}) - g_l^{(k)}(\hat{s}_l^{(k)}) = \ln(1+\hat{s}_l^{(k)}) - [m_l^{(k)} \cdot (\hat{s}_l^{(k)} - s_l^{(k-1)}) + \ln(1+s_l^{(k-1)})] \\ &= \ln\left(1 + \frac{1}{m_l^{(k)}} - 1\right) - \left\{m_l^{(k)} \cdot \left[\frac{1}{m_l^{(k)}} - 1 - s_l^{(k-1)}\right] + \ln(1+s_l^{(k-1)})\right\} \\ &= -\ln(m_l^{(k)}) + m_l^{(k)}(1+s_l^{(k-1)}) - 1 - \ln(1+s_l^{(k-1)}),\end{aligned}$$

where the second equality holds due to (6.3.5) and the third equality holds due to (6.3.7).

In Algorithm 6.1, we set $-\ln(m_l^{(k)}) + m_l^{(k)}(1+s_l^{(k-1)}) - 1 - \ln(1+s_l^{(k-1)}) = \epsilon$. Thus, the maximum linear approximation error for the k-th linear segment is ϵ. This result holds for all $k = 1, \cdots, K_l$. This completes the proof. □

Lemma 6.2. *For a given approximation error bound ϵ for each linear segment, the number of linear segments to approximate $\ln(1+s_l)$ for $s_l \in [0, s_l^{\max}]$ is minimized by Algorithm 6.1.*

Proof. Our proof is based on contradiction. Assume that the number of linear segments that Algorithm 6.1 generates is K_l, and $s_l^{(k)}, k = 0, \ldots, K_l$, are the corresponding X-axis values of the

endpoints. Suppose that there is another piece-wise linear approximation that needs $K_l' < K_l$ linear segments and $t_l^{(k)}$s (with $k = 0, \ldots, K_l$, $t_l^{(0)} = 0$ and $t_l^{(K_l')} = s_l^{\max}$) are the corresponding X-axis values of the endpoints.

Since $s_l^{(1)}$ is the largest X-axis value of the second endpoint, we have $t_l^{(1)} \leq s_l^{(1)}$. By induction, we can show that $t_l^{(k)} \leq s_l^{(k)}, k = 1, \ldots, K_l'$. For $k = K_l'$, we have $t_l^{(K_l')} \leq s_l^{(K_l')}$. Further, since $K_l' < K_l$, we also have $s_l^{(K_l')} < s_l^{\max}$. Therefore, we conclude that $t_l^{(K_l')} \leq s_l^{(K_l')} < s_l^{\max}$, which is a contradiction to $t_l^{(K_l')} = s_l^{\max}$. This completes our proof. □

With the proposed piece-wise linear approximation of $\ln(1 + s_l)$, constraint (6.2.8) can be replaced by the following set of constraints:

$$\sum_{f \in \mathcal{F}} r_l(f) \leq \frac{B_l}{\ln 2} g_l^{(k)}(s_l) \qquad (k = 1, \ldots, K_l, l \in \mathcal{L}) \ ,$$

where s_l and $g_l^{(k)}(s_l)$ are given in (6.3.3) and (6.3.5), respectively. Substituting (6.3.3) and (6.3.5) into the above equation, we have

$$\sum_{f \in \mathcal{F}} r_l(f) \leq \frac{B_l}{\ln 2} \left\{ m_l^{(k)} \left[\frac{p_l h_l}{\eta B_l} - s_l^{(k-1)} \right] + \ln \left[1 + s_l^{(k-1)} \right] \right\} \quad (k = 1, \ldots, K_l, l \in \mathcal{L}) \ . \quad (6.3.8)$$

By replacing the nonlinear constraints in (6.2.8) with the set of linear constraints in (6.3.8), we have a revised formulation for OPT, which we denote as OPT-R.

$$\begin{aligned} \textbf{OPT-R:} \quad \max \quad & \sum_{f \in \mathcal{F}} w(f) r(f) \\ \text{s.t.} \quad & \text{Constraints } (6.2.1), (6.2.2), (6.2.3), (6.2.4), (6.3.1), (6.3.8) \\ & \text{Variables } y_l \in \{0, 1\}, p_l, r_l(f), r(f) \geq 0 \ (l \in \mathcal{L}, f \in \mathcal{F}) \ . \end{aligned}$$

We have the following lemma on the relationship between OPT-R and OPT.

Lemma 6.3. *A feasible solution to OPT-R is a feasible solution to OPT.*

Proof. Note that the only difference between OPT and OPT-R is the link capacity constraints. Each link capacity constraint for link l in OPT is replaced by a set of linear constraints in OPT-R.

Since these linear constraints are generated by the piece-wise linear segments lying beneath the $\log$ curve, the feasible region of OPT-R falls inside in the feasible region of OPT. Thus, a feasible solution to OPT-R is also a feasible solution to OPT. □

6.3.2 A Near-Optimal Solution

OPT-R is a mixed-integer linear program (MILP) and can be solved efficiently by CPLEX solver [23]. Now we give a bound for the gap between the optimal objective values of OPT and OPT-R, despite that the optimal objective value of OPT is unknown.

To proceed, we need the following notation. For a given power assignment (y_l, p_l) to OPT (i.e., satisfying constraints (6.2.1), (6.2.2), (6.3.1)), define $\bar{\mathbf{x}} = (\bar{r}(f), \bar{r}_l(f), y_l, p_l)$ as a feasible solution to OPT, where $(\bar{r}(f), \bar{r}_l(f))$ is the optimal solution to the following linear program (LP).

$$
\begin{aligned}
\textbf{OPT}(y_l, p_l)\textbf{:}\quad \max & \sum_{f\in\mathcal{F}} w(f) r(f) & \\
\text{s.t.} & \sum_{l\in\mathcal{L}_i^{\mathrm{Out}}} r_l(f) = r(f) & (f\in\mathcal{F}, i\in\mathcal{N}, i = s(f)) \\
& \sum_{l\in\mathcal{L}_i^{\mathrm{Out}}}^{l\neq(i,s(f))} r_l(f) = \sum_{l\in\mathcal{L}_i^{\mathrm{In}}}^{l\neq(d(f),i)} r_l(f) & (f\in\mathcal{F}, i\in\mathcal{N}, i\neq s(f), d(f)) \\
& \sum_{f\in\mathcal{F}} r_l(f) \le \bar{c}_l & (l\in\mathcal{L})\ ,
\end{aligned}
$$

where $\bar{c}_l = B_l \log_2(1 + \frac{p_l \cdot h_l}{\eta B_l})$. Note that OPT$(y_l, p_l)$ is an LP once we set the power variables in OPT to values (y_l, p_l).

For a feasible solution $\bar{\mathbf{x}} = (\bar{r}(f), \bar{r}_l(f), y_l, p_l)$ to OPT, we define a feasible solution $\mathbf{x}^\dagger = (r^\dagger(f), r_l^\dagger(f), y_l, p_l)$ to OPT-R as follows. In $\mathbf{x}^\dagger = (r^\dagger(f), r_l^\dagger(f), y_l, p_l)$, we let $(r^\dagger(f), r_l^\dagger(f))$ be the optimal flow routing solution to OPT-R with given (y_l, p_l). That is, $(r^\dagger(f), r_l^\dagger(f))$ is the optimal

solution to the following LP, in which the power variables in OPT-R are set to given values (y_l, p_l).

$$\begin{aligned}
\textbf{OPT-R}(y_l, p_l)\textbf{:}\quad \max & \sum_{f\in\mathcal{F}} w(f)r(f) \\
\text{s.t.} & \sum_{l\in\mathcal{L}_i^{\text{Out}}} r_l(f) = r(f) && (f\in\mathcal{F}, i\in\mathcal{N}, i = s(f)) \\
& \sum_{l\in\mathcal{L}_i^{\text{Out}}}^{l\neq(i,s(f))} r_l(f) = \sum_{l\in\mathcal{L}_i^{\text{In}}}^{l\neq(d(f),i)} r_l(f) && (f\in\mathcal{F}, i\in\mathcal{N}, i\neq s(f), d(f)) \\
& \sum_{f\in\mathcal{F}} r_l(f) \le c_l^{\dagger} && (l\in\mathcal{L})\ ,
\end{aligned}$$

where $c_l^{\dagger}$ is a linear approximation of link l's capacity under transmission power p_l.

Remark 6.1. *Recall that we use constraints (6.3.8) to replace constraints (6.2.8) in OPT-R. When link l's power is fixed at p_l, we can determine which line segment is involved in our linear approximation of $\ln(1+s_l)$. Suppose the k-th linear segment is used, i.e., $s_l^{(k-1)} \le \frac{p_l\cdot h_l}{\eta B_l} \le s_l^{(k)}$. Then, link l's approximated capacity can be written as $c_l^{\dagger} = \frac{B_l}{\ln 2}\cdot g_l^{(k)}(\frac{p_l\cdot h_l}{\eta B_l})$.* □

To quantify the performance gap between our solution to OPT-R and the optimal solution to OPT, we will first show that for any feasible power assignment (p_l, y_l), the objective value gap between $\bar{\mathbf{x}}$ and $\mathbf{x}^{\dagger}$ is at most $\epsilon\cdot\sum_{f\in\mathcal{F}}\sum_{l\in\mathcal{L}_{s(f)}^{\text{Out}}}\frac{B_l}{\ln 2}w(f)$. Then, we will show that the gap between the optimal objective values of OPT and OPT-R is also bounded by $\epsilon\cdot\sum_{f\in\mathcal{F}}\sum_{l\in\mathcal{L}_{s(f)}^{\text{Out}}}\frac{B_l}{\ln 2}w(f)$.

Lemma 6.4. *For given (y_l, p_l), denote $\bar{z}$ and $z^{\dagger}$ the objective values of solution $\bar{\mathbf{x}}$ (to OPT) and solution $\mathbf{x}^{\dagger}$ (to OPT-R), respectively. Then we have $\bar{z} - z^{\dagger} \le \epsilon\cdot\sum_{f\in\mathcal{F}}\sum_{l\in\mathcal{L}_{s(f)}^{\text{Out}}}\frac{B_l}{\ln 2}w(f)$.*

We find that it is not easy to characterize the gap between $\bar{z}$ and $z^{\dagger}$ directly. Since $\bar{z}$ is the optimal value of OPT(y_l, p_l) and $z^{\dagger}$ is the optimal objective value of OPT-R(y_l, p_l), we study the dual problems of OPT(y_l, p_l) and OPT-R(y_l, p_l) and quantify $\bar{z} - z^{\dagger}$ in the dual domain.

Proof. Note that $\bar{z}$ is the optimal objective value of OPT(y_l, p_l) and $z^{\dagger}$ is the optimal objective value of OPT-R(y_l, p_l). Consider the dual problems of OPT(y_l, p_l) and OPT-R(y_l, p_l). Denote

D(y_l, p_l) and D-R(y_l, p_l) as the dual problems of OPT(y_l, p_l) and OPT-R(y_l, p_l), respectively. Note that D(y_l, p_l) and D-R(y_l, p_l) will have the same constraints, but different objective functions.

Denote the dual variables corresponding to the first group of constraints in OPT(y_l, p_l) and OPT-R(y_l, p_l) as $u(f), f \in \mathcal{F}$. Denote the dual variables corresponding to the second group of constraints in OPT(y_l, p_l) and OPT-R(y_l, p_l) as $v_i(f), f \in \mathcal{F}, i \in \mathcal{N}, i \neq s(f), d(f)$. Denote the dual variables corresponding to the third group of constraints in OPT(y_l, p_l) and OPT-R(y_l, p_l) as $q_l, l \in \mathcal{L}$. Then, D(y_l, p_l) can be written as

$$\begin{aligned}
\text{D}(y_l, p_l): \quad & \min \sum_{l \in \mathcal{L}} \bar{c}_l q_l \\
& \text{s.t.} -u(f) \geq w(f) \ (f \in \mathcal{F}) && (6.3.9) \\
& \quad v_i(f) + q_l \geq 0 \ (f \in \mathcal{F}, l \in \mathcal{L}^{\text{Out}}(i), i \neq s(f), d(f)) \\
& \quad -v_i(f) + q_l \geq 0 \ (f \in \mathcal{F}, l \in \mathcal{L}^{\text{In}}(i), i \neq s(f), d(f)) \\
& \quad u(f) + q_l \geq 0 \ (f \in \mathcal{F}, l \in \mathcal{L}^{\text{Out}}(s(f))) && (6.3.10) \\
& \quad u(f), v_i(f) \text{ unrestricted}, q_l \geq 0.
\end{aligned}$$

Dual problem D-R(y_l, p_l) can be written as

$$\begin{aligned}
\text{D-R}(y_l, p_l): \quad & \min \ \sum_{l \in \mathcal{L}} c_l^\dagger q_l \\
& \text{s.t. All constraints in D}(y_l, p_l).
\end{aligned}$$

Combining (6.3.9) and (6.3.10) gives us $q_l \geq w(f), l \in \mathcal{L}^{\text{Out}}(s(f)), f \in \mathcal{F}$. Since these two dual problems are both minimization problems, it is easy to see that the solution with $q_l^* = w(f), (l \in \mathcal{L}^{\text{Out}}(s(f)), f \in \mathcal{F})$ and all the other variables equal to zero is the optimal solution to both D(y_l, p_l) and D-R(y_l, p_l). That is

$$q_l^* = \begin{cases} w(f) & \text{if link } l \text{ is an outgoing link from } s(f); \\ 0 & \text{otherwise.} \end{cases} \tag{6.3.11}$$

Then, we have

$$\bar{z} - z^\dagger = \sum_{l \in \mathcal{L}} \bar{c}_l q_l^* - \sum_{l \in \mathcal{L}} c_l^\dagger q_l^* = \sum_{l \in \mathcal{L}} (\bar{c}_l - c_l^\dagger) q_l^* = \sum_{f \in \mathcal{F}} \sum_{l \in \mathcal{L}^{\text{Out}}(s(f))} (\bar{c}_l - c_l^\dagger) w(f) \ , \tag{6.3.12}$$

where the first equality holds due to the strong duality property [6, Chapter 6] and the third equality holds due to (6.3.11). Note that the gap between $\bar{c}_l$ and $c_l^\dagger$ is

$$\bar{c}_l - c_l^\dagger \leq \frac{B_l}{\ln 2}\epsilon, \tag{6.3.13}$$

since the maximum error of our linear approximation is ϵ. Combining (6.3.12) and (6.3.13) gives us

$$\bar{z} - z^\dagger \leq \epsilon \cdot \sum_{f\in\mathcal{F}} \sum_{l\in\mathcal{L}^{\mathrm{Out}}(s(f))} \frac{B_l}{\ln 2} w(f) \ .$$

This completes the proof. □

Now we are ready to characterize the performance gap between the optimal objective values of OPT-R and OPT as follows.

Theorem 6.1. *The gap between the optimal objective values of OPT and OPT-R is no more than* $\epsilon \cdot \sum_{f\in\mathcal{F}} \sum_{l\in\mathcal{L}^{\mathrm{Out}}_{s(f)}} \frac{B_l}{\ln 2} w(f)$.

Proof. Denote $\mathbf{x}^*$ and z^* the optimal solution and the optimal objective value of OPT, respectively. From Lemma 6.4, since $\mathbf{x}^*$ is a particular case of $\bar{\mathbf{x}}$, we know that there exists a feasible solution of OPT-R $\mathbf{x}_R$ corresponding to $\mathbf{x}^*$ such that the performance gap between $\mathbf{x}^*$ and $\mathbf{x}_R$ is at most $\epsilon \cdot \sum_{f\in\mathcal{F}} \sum_{l\in\mathcal{L}^{\mathrm{Out}}_{s(f)}} \frac{B_l}{\ln 2} w(f)$. Denote z_R the objective value of solution $\mathbf{x}_R$ to OPT-R. Then, we have

$$z^* - z_R \leq \epsilon \cdot \sum_{f\in\mathcal{F}} \sum_{l\in\mathcal{L}^{\mathrm{Out}}_{s(f)}} \frac{B_l}{\ln 2} w(f). \tag{6.3.14}$$

Denote z_R^* the optimal objective value of OPT-R. Since z_R is the objective value of a feasible solution to OPT-R while z_R^* is the optimal objective value of OPT-R, we have

$$z_R^* \geq z_R. \tag{6.3.15}$$

Combining (6.3.14) and (6.3.15), we have $z^* - z_R^* \leq \epsilon \cdot \sum_{f\in\mathcal{F}} \sum_{l\in\mathcal{L}^{\mathrm{Out}}_{s(f)}} \frac{B_l}{\ln 2} w(f)$. □

Based on Theorem 6.1, we are able to give an algorithm to obtain a near-optimal solution to OPT with performance guarantee as follows.

Algorithm 6.2. Input: *Given a desired performance gap γ for the solution.*

1. *Compute ϵ based on*
$$\epsilon \cdot \sum_{f \in \mathcal{F}} \sum_{l \in \mathcal{L}_{s(f)}^{\text{Out}}} \frac{B_l}{\ln 2} w(f) = \gamma. \tag{6.3.16}$$
2. *Compute $m_l^{(k)}$ and $s_l^{(k)}$ by Algorithm 6.1.*
3. *Construct OPT-R based on $m_l^{(k)}$ and $s_l^{(k)}$.*
4. *Solve OPT-R optimally with CPLEX.*

Upon the completion of Algorithm 6.2, we will have a near-optimal solution to OPT with a guaranteed performance bound (no more than γ from the optimal objective value).

6.4 Maximizing Throughput and Minimizing Network-wide Energy Consumption

In the previous section, we have shown how to maximize network throughput while satisfying a given total network-wide energy budget. The problem was formulated as a *single objective* optimization problem OPT. In this section, we take one step further. We are interested in maximizing network throughput while minimizing energy consumption. We cast this problem into a *multicriteria* optimization problem with two objectives. Mathematically, this problem can be written as follows:

$$\begin{aligned}
\textbf{MP:} \quad & \max \quad \sum_{f \in \mathcal{F}} w(f) r(f) \\
& \min \quad \sum_{l \in \mathcal{L}} (p_l + y_l P_d) \\
& \text{s.t.} \quad \text{Constraints } (6.2.1), (6.2.2), (6.2.3), (6.2.4), (6.2.8) \\
& \qquad \text{Variables } y_l \in \{0,1\}, p_l, r_l(f), r(f) \geq 0 \ (l \in \mathcal{L}, f \in \mathcal{F}).
\end{aligned}$$

As we can see, minimizing network-wide energy consumption and maximizing network throughput are two conflicting objectives. For such a problem, it is in general not possible to find a single feasible solution that is optimal for both objectives at the same time. For example, when P is minimized (i.e., 0), U is also 0 but is not maximized. Therefore, it is important to clarify what we mean by optimal solutions.

In this chapter, we are interested in finding the so-called weakly Pareto-optimal solutions [29]. Weakly Pareto-optimal solutions are optimal in the sense that it is impossible to improve the performance of both objectives simultaneously. Specifically, we say that (P^*, U^*) is a weakly Pareto-optimal point to problem MP if there does not exist another solution to problem MP with (P, U) such that $P < P^*$ and $U > U^*$.

To find weakly Pareto-optimal points, we transform the multicriteria optimization problem into a single objective optimization problem. This can be done by moving the second objective (i.e., $\sum_{l\in\mathcal{L}}(p_l + y_l P_d)$) into the constraints as follows.

$$
\begin{aligned}
\textbf{SP}(P_{\text{net}})\textbf{:} \quad \text{max} \quad & \sum_{f\in\mathcal{F}} w(f) r(f) \\
\text{s.t.} \quad & \sum_{l\in\mathcal{L}} (p_l + y_l P_d) \le P_{\text{net}} \\
& \text{Constraints} (6.2.1), (6.2.2), (6.2.3), (6.2.4), (6.2.8) \\
& \text{Variables } y_l \in \{0,1\}, p_l, r_l(f), r(f) \ge 0 \ (l \in \mathcal{L}, f \in \mathcal{F}).
\end{aligned}
$$

We see that this single objective optimization problem is precisely the same as OPT that we studied earlier. For a fixed value of P_{net}, solving SP(P_{net}) will give us *one* weakly Pareto-optimal point of problem MP [29]. By varying P_{net} from 0 to $P_{\text{net}}^{\max} = |\mathcal{L}| \cdot (P_{\max} + P_d)$, we can obtain all the weakly Pareto-optimal points. These points provide a mapping from the network-wide energy budget P_{net} to the maximum network throughput U, which we denote as $\pi : P_{\text{net}} \to U$. This mapping $U = \pi(P_{\text{net}})$ is an optimal throughput-energy curve, which characterizes how the maximum network throughput changes as the total network-wide energy consumption rate varies. This curve is useful for network operators to have a holistic view of the entire optimal trade-off curve and decide which point to choose so as to meet their needs.

We have several interesting properties about this optimal throughput-energy curve $U = \pi(P_{\text{net}})$, which are shown in Property 6.1.

Property 6.1. *The optimal throughput-energy curve $U = \pi(P_{\text{net}})$ has the following properties.*

1. *$\pi(P_{\text{net}})$ is a nondecreasing function of P_{net}.*
2. *$\pi(P_{\text{net}})$ has a starting point $(P_{\text{start}}, 0)$, i.e., $\pi(P_{\text{net}}) = 0$ for $P_{\text{net}} \leq P_{\text{start}}$ and $\pi(P_{\text{net}}) > 0$ for $P_{\text{net}} > P_{\text{start}}$.*
3. *$\pi(P_{\text{net}})$ has a saturation point $(P_{\text{sat}}, U_{\text{sat}})$, i.e., $\pi(P_{\text{net}}) = U_{\text{sat}}$ for $P_{\text{net}} \geq P_{\text{sat}}$ and $\pi(P_{\text{net}}) < U_{\text{sat}}$ for $P_{\text{net}} < P_{\text{sat}}$.*

Proof. We prove each property as follows.

1. Assume $P_{\text{net}}^{(1)} < P_{\text{net}}^{(2)}$. We need to show that $U(P_{\text{net}}^{(1)}) \leq U(P_{\text{net}}^{(2)})$. Note that $U(P_{\text{net}}^{(1)})$ and $U(P_{\text{net}}^{(2)})$ are the optimal objectives of SP($P_{\text{net}}^{(1)}$) and SP($P_{\text{net}}^{(2)}$), respectively. Since $P_{\text{net}}^{(1)} < P_{\text{net}}^{(2)}$, the feasible region of SP($P_{\text{net}}^{(1)}$) falls inside the feasible region of SP($P_{\text{net}}^{(2)}$). Thus, we have $U(P_{\text{net}}^{(1)}) \leq (P_{\text{net}}^{(2)})$.
2. Such starting point exists because when a link is active, it must consume a constant power P_d. For a session to have positive throughput, it must activate all the links along the path that are used by this session for transporting data. Thus, P_{start} can be determined by the session that uses the minimum number of hops from its source to its destination. Denote m_f the minimum hops of session f. Then, P_{start} can be written as $P_{\text{start}} = P_d \cdot \min\{m_f : f \in \mathcal{F}\}$.
3. The saturation point $(P_{\text{sat}}, U_{\text{sat}})$ can be determined as follows. We can first compute the maximum network throughput without network-wide energy constraint, i.e., solving the following optimization problem.

$$\begin{aligned} \max \quad & \sum_{f \in \mathcal{F}} w(f) r(f) \\ \text{s.t.} \quad & \text{Constraints } (6.2.1), (6.2.2), (6.2.3), (6.2.4), (6.2.8). \end{aligned}$$

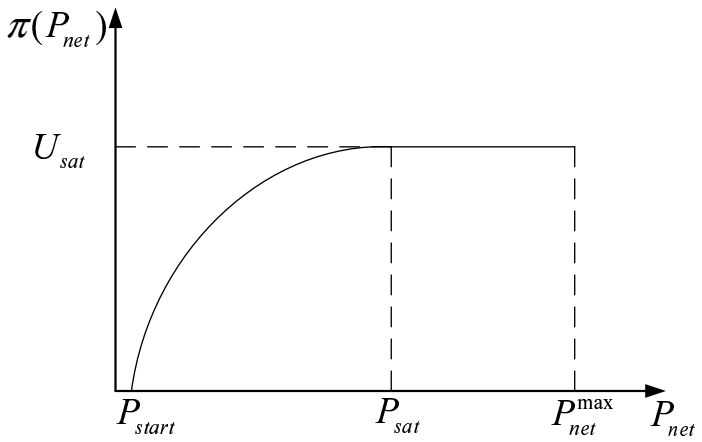


Figure 6.4: An illustration of optimal throughput-energy curve.

The optimal objective value of the above problem is $U_{\rm sat}$. Then, we determine the minimum energy that can achieves this throughput by solving the following optimization problem.

$$\begin{aligned} \text{min} \quad & \sum_{l\in\mathcal{L}}(p_l + y_l P_c) \\ \text{s.t.} \quad & w(f)r(f) = U_{\rm sat} \\ & \text{Constraints } (6.2.1), (6.2.2), (6.2.3), (6.2.4), (6.2.8). \end{aligned}$$

□

Based on Property 6.1, Fig. 6.4 illustrates a typical optimal throughput-energy curve for a multi-hop wireless network.

6.5 Numerical Results

In this section, we present some numerical results to illustrate our theoretical findings in Section 6.3 and 6.4.

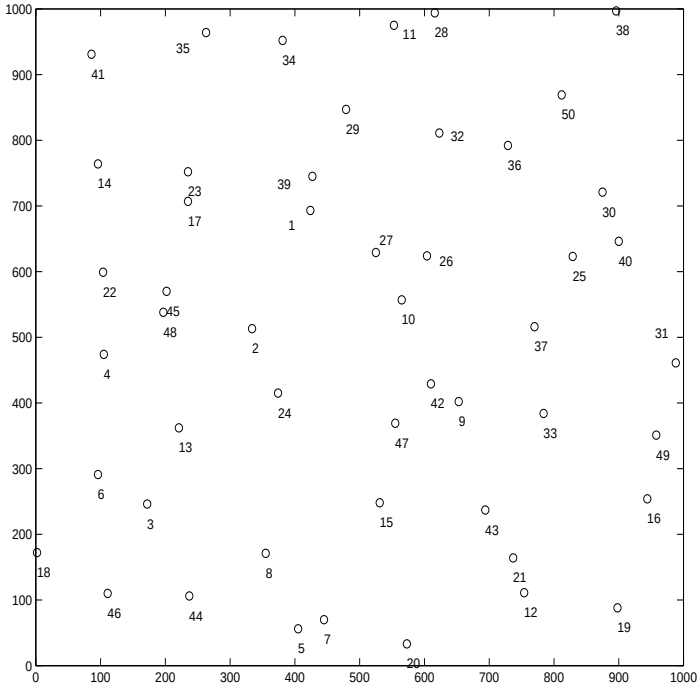


Figure 6.5: The topology for a 50-node network.

6.5.1 Simulation Settings

We consider a 50-node network deployed in a 1000 $\times$ 1000 square area and a 100-node network deployed in a 1500 $\times$ 1500 square area. The topologies of the 50-node network and 100-node network are shown in Fig. 6.5 and Fig. 6.6, respectively. We assume that all units are normalized with appropriate dimensions. We assume the maximum transmission range is 200 and the maximum transmission power is $P_{\max}$ = 2. We assume node device power consumption is P_d = 0.2. The channel bandwidth is $B_l = 1$ for all links and channel gain is $h_l = d_l^{-4}$, where d_l is the distance between link l's transmitting node and receiving node.

6.5.2 Results for the 50-node Network

Within this network, we assume there are $|\mathcal{F}| = 5$ user sessions, with source node and destination node of each session chosen randomly. Table 6.2 lists the source node, destination node, and

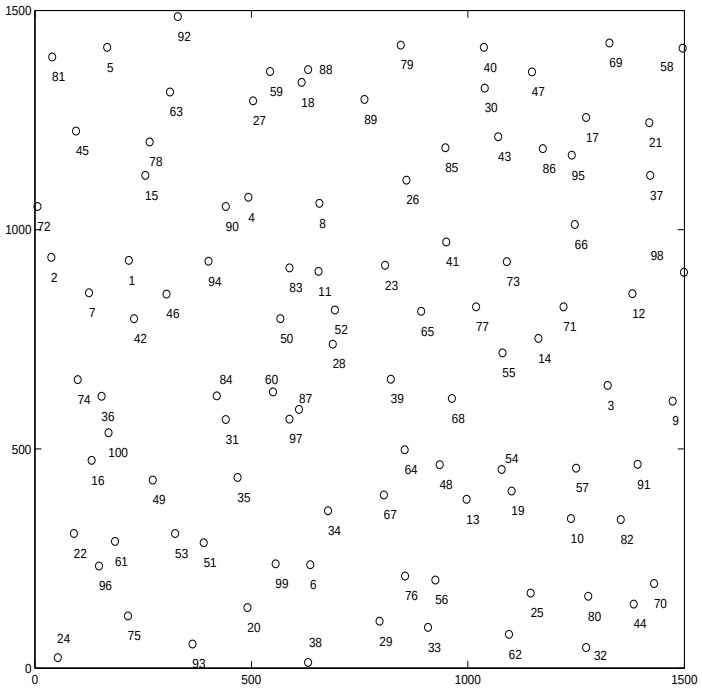


Figure 6.6: The topology for a 100-node network.

weight for each session in the network.

Near-Optimal Solution for OPT

In this case study, we assume the maximum network-wide energy consumption rate $P_{\text{net}} = 40$. We set the maximum acceptable performance gap between the optimal objectives of OPT and linear approximation OPT-R as $\gamma = 0.1$. We apply Algorithm 6.2 here. Based on (6.3.16), we compute $\epsilon = \frac{\gamma \cdot \ln 2}{\sum_{f \in \mathcal{F}} \sum_{l \in \mathcal{L}_{s(f)}^{\text{Out}}} \frac{B_l}{\ln 2} w(f)} = 0.0046$. Based on ϵ, we compute the piece-wise linear approximation according to Algorithm 6.1.

Then we can use CPLEX to solve OPT-R. We obtain that the maximum network throughput is $U = 22.12$. The achieved session data rates are $r_1 = 4.41$, $r_2 = 6.39$, $r_3 = 9.37$, $r_4 = 3.89$, and $r_5 = 6.62$. Our algorithm gives power control and flow routing solutions for the network. We list the power assignment for each active link in Table 6.3, and the flow routing results in Table 6.4.

Table 6.2: Each session's source node, destination node, and weight for the 50-node network.

Session f	Source node $s(f)$	Dest. node $d(f)$	Weight $w(f)$
1	10	35	0.5
2	35	21	0.9
3	5	23	0.7
4	43	14	0.6
5	29	7	0.8

The Optimal Throughput-Energy Curve

For the same 50-node network instance, we characterize its optimal throughput-energy curve based on our theoretical results in Section 6.4. We show the optimal throughput-energy curve in Fig. 6.7. From the figure, we can see all three properties as stated in Property 6.1. As shown in the figure, the curve is nondecreasing. The network throughput keeps at zero when the network energy consumption rate is no greater than P_{start}. For the starting point $(P_{\text{start}}, 0)$, since session 1 needs at least 5 hops , we have $P_{\text{start}} = 5 \cdot P_d = 1$. For the saturation point $(P_{\text{sat}}, U_{\text{sat}})$, we get $(P_{\text{sat}}, U_{\text{sat}}) = (106.20, 36.14)$. The network throughput stops increasing and keeps as 36.14 when the network energy consumption rate exceeds $P_{\text{sat}} = 106.20$.

6.5.3 Results for the 100-node Network

For the 100-node network, we assume that there are $|\mathcal{F}| = 10$ active sessions in the network, with each session's source node, destination node, and weight given in Table 6.5.

We assume that maximum network-wide energy consumption rate $P_{\text{net}} = 100$. By employing our method, we obtain that the maximum network throughput is $U = 42.00$. The achieved session data rates are $r_1 = 10.86$, $r_2 = 1.63$, $r_3 = 7.09$, $r_4 = 4.03$, $r_5 = 9.71$, $r_6 = 4.00$, $r_7 = 9.90$, $r_8 = 4.91$, and $r_9 = 8.08$, and $r_{10} = 7.19$. The detailed results for power assignment and flow routing are omitted to conserve space. The optimal throughput-energy curve for the 100-node network is similar to that for the 20-node network, which is omitted here.

Table 6.3: Power assignment on each active link in the final solution for the 50-node network.

Link	Power	Link	Power	Link	Power
$1 \to 27$	0.1819	$1 \to 23$	0.4317	$1 \to 17$	0.4658
$2 \to 45$	0.1958	$2 \to 24$	0.0370	$3 \to 44$	0.2050
$3 \to 13$	0.1805	$3 \to 6$	0.0350	$4 \to 45$	0.2083
$4 \to 22$	0.1692	$4 \to 13$	0.2441	$5 \to 44$	0.2652
$5 \to 8$	0.1775	$5 \to 7$	0.0313	$6 \to 4$	0.6487
$7 \to 15$	0.4290	$7 \to 8$	0.1534	$8 \to 44$	0.0707
$8 \to 15$	0.2924	$8 \to 7$	0.1209	$8 \to 3$	0.5524
$9 \to 43$	0.1794	$9 \to 10$	0.2835	$10 \to 47$	0.5756
$10 \to 42$	0.0952	$10 \to 27$	0.3033	$10 \to 26$	0.0101
$10 \to 9$	0.2166	$10 \to 1$	0.2355	$11 \to 34$	0.2547
$11 \to 32$	0.1617	$13 \to 4$	0.4853	$13 \to 3$	0.0908
$14 \to 22$	0.2196	$15 \to 47$	0.2544	$15 \to 8$	0.4918
$15 \to 7$	0.5515	$17 \to 45$	0.1424	$17 \to 23$	0.0151
$17 \to 14$	0.1431	$22 \to 45$	0.0628	$22 \to 17$	0.3000
$22 \to 14$	0.2092	$24 \to 47$	0.3587	$24 \to 2$	0.0575
$25 \to 37$	0.1283	$26 \to 32$	0.3506	$27 \to 39$	0.4733
$27 \to 10$	0.3033	$27 \to 1$	0.1177	$29 \to 39$	0.5181
$29 \to 34$	0.1950	$29 \to 32$	0.0776	$29 \to 1$	0.6365
$30 \to 25$	0.0791	$32 \to 36$	0.0774	$32 \to 11$	0.2840
$33 \to 43$	0.5081	$34 \to 35$	0.4009	$34 \to 29$	0.3055
$34 \to 11$	0.1450	$35 \to 41$	0.3099	$35 \to 34$	0.4009
$36 \to 30$	0.3999	$37 \to 33$	0.1787	$39 \to 29$	0.0787
$39 \to 27$	0.3061	$39 \to 23$	0.1054	$39 \to 17$	0.5299
$41 \to 14$	0.2310	$42 \to 15$	0.4273	$42 \to 10$	0.2433
$43 \to 47$	0.3793	$43 \to 21$	0.4274	$43 \to 9$	0.2347
$44 \to 5$	0.3408	$44 \to 3$	0.6235	$45 \to 23$	0.1872
$45 \to 22$	0.0306	$45 \to 17$	0.2270	$45 \to 4$	0.1253
$45 \to 2$	0.1260	$47 \to 43$	0.3982	$47 \to 42$	0.0315
$47 \to 24$	0.5573	$47 \to 15$	0.1061		

Table 6.4: Flow routing results for the 50-node network.

Session f	Flow rate on each link attributed to session f
1	$r_{10\to27}(1) = 2.48$, $r_{10\to26}(1) = 1.93$, $r_{11\to34}(1) = 1.93$, $r_{26\to32}(1) = 1.93$, $r_{27\to39}(1) = 2.48$ $r_{29\to34}(1) = 2.48$, $r_{32\to11}(1) = 1.93$, $r_{34\to35}(1) = 4.41$, $r_{39\to29}(1) = 2.48$
2	$r_{1\to27}(2) = 1.65$, $r_{2\to24}(2) = 1.98$, $r_{9\to43}(2) = 1.65$, $r_{10\to9}(2) = 1.65$, $r_{11\to32}(2) = 1.38$ $r_{14\to22}(2) = 1.98$, $r_{22\to45}(2) = 1.98$, $r_{24\to47}(2) = 1.98$, $r_{25\to37}(2) = 2.76$ $r_{27\to10}(2) = 1.65$ $r_{29\to32}(2) = 1.38$,$r_{29\to1}(2) = 1.65$, $r_{30\to25}(2) = 2.76$, $r_{32\to36}(2) = 2.76$, $r_{33\to43}(2) = 2.76$ $r_{34\to29}(2) = 3.03$, $r_{34\to11}(2) = 1.38$, $r_{35\to41}(2) = 1.98$, $r_{35\to34}(2) = 4.41$, $r_{36\to30}(2) = 2.76$ $r_{37\to33}(2) = 2.76$, $r_{41\to14}(2) = 1.98$, $r_{43\to21}(2) = 6.39$, $r_{45\to2}(2) = 1.98$ $r_{47\to43}(2) = 1.98$
3	$r_{1\to23}(3) = 1.93$, $r_{1\to17}(3) = 0.28$, $r_{2\to45}(3) = 0.55$, $r_{3\to13}(3) = 3.03$, $r_{3\to6}(3) = 2.76$ $r_{4\to45}(3) = 2.80$, $r_{4\to22}(3) = 2.98$, $r_{5\to44}(3) = 1.93$, $r_{5\to8}(3) = 3.03$ $r_{5\to7}(3) = 4.41$ $r_{6\to4}(3) = 2.76$, $r_{7\to15}(3) = 1.93$, $r_{7\to8}(3) = 2.48$, $r_{8\to44}(3) = 1.65$, $r_{8\to15}(3) = 1.65$ $r_{8\to3}(3) = 2.21$, $r_{10\to27}(3) = 1.65$, $r_{10\to1}(3) = 1.38$, $r_{13\to4}(3) = 3.03$, $r_{15\to47}(3) = 3.58$ $r_{17\to23}(3) = 5.24$, $r_{22\to45}(3) = 0.78$, $r_{22\to17}(3) = 2.21$, $r_{24\to2}(3) = 0.55$, $r_{27\to39}(3) = 0.83$ $r_{27\to1}(3) = 0.83$, $r_{39\to23}(3) = 0.83$, $r_{42\to10}(3) = 3.03$, $r_{44\to3}(3) = 3.58$, $r_{45\to23}(3) = 1.38$ $r_{45\to17}(3) = 2.76$, $r_{47\to42}(3) = 3.03$, $r_{47\to24}(3) = 0.55$
4	$r_{1\to17}(4) = 1.93$, $r_{2\to45}(4) = 1.93$, $r_{9\to10}(4) = 1.93$, $r_{10\to27}(4) = 1.93$, $r_{17\to14}(4) = 1.93$ $r_{22\to14}(4) = 1.93$, $r_{24\to2}(4) = 1.93$, $r_{27\to1}(4) = 1.93$, $r_{43\to47}(4) = 1.93$, $r_{43\to9}(4) = 1.93$ $r_{45\to22}(4) = 1.93$, $r_{47\to24}(4) = 1.93$
5	$r_{1\to27}(5) = 1.65$, $r_{3\to44}(5) = 2.21$, $r_{4\to13}(5) = 2.21$, $r_{5\to7}(5) = 2.21$, $r_{8\to7}(5) = 2.21$ $r_{10\to47}(5) = 2.48$, $r_{10\to42}(5) = 1.93$, $r_{13\to3}(5) = 2.21$, $r_{15\to8}(5) = 2.21$, $r_{15\to7}(5) = 2.21$ $r_{17\to45}(5) = 2.21$, $r_{27\to10}(5) = 4.41$, $r_{29\to39}(5) = 4.96$, $r_{29\to1}(5) = 1.65$, $r_{39\to27}(5) = 2.76$ $r_{39\to17}(5) = 2.21$, $r_{42\to15}(5) = 1.93$, $r_{44\to5}(5) = 2.21$, $r_{45\to4}(5) = 2.21$, $r_{47\to15}(5) = 2.48$

Table 6.5: Each session's source node, destination node, and weight for the 100-node network.

Session f	Source node $s(f)$	Dest. node $d(f)$	Weight $w(f)$
1	40	26	0.9
2	27	17	0.8
3	4	55	0.7
4	31	41	0.6
5	78	100	0.8
6	7	83	0.6
7	73	91	0.3
8	12	10	0.4
9	64	38	0.6
10	51	56	0.5

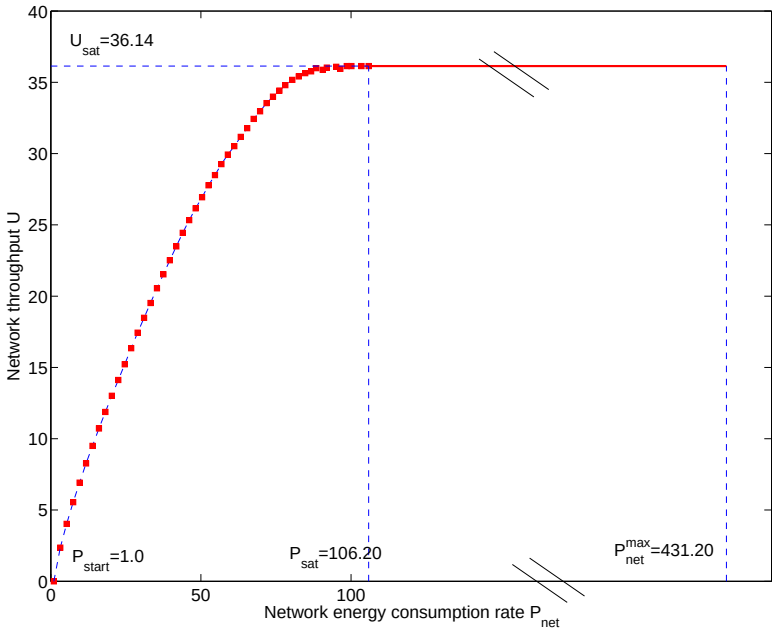


Figure 6.7: The optimal throughput-energy curve for the 50-node network, where the "\\" sign in the figure indicates nonlinear scale for $P_{\text{net}} \in [106.20, 431.20]$.

6.6 Chapter Summary

Network-wide energy consumption has become an important concern for network operators. In this chapter, we studied two tightly coupled problems for network-wide energy conservation. In the first problem, we studied how to maximize network throughput under a network-wide energy constraint. We formulated this problem into a mixed-integer nonlinear program (MINLP) and proposed a near-optimal solution with guaranteed performance bound. In the second problem, we explored joint optimization of both network throughput and energy consumption via a multicriteria optimization framework. We showed that the weakly Pareto-optimal points in the solution can characterize an optimal throughput-energy curve. The results in this chapter offer both solutions and insights to network operators when total energy consumption for the entire network is of greater concern than local energy consumption.

Chapter 7

Bicriteria Optimization in Multi-Hop Wireless Networks: Characterizing Throughput-Energy Envelope

7.1 Introduction

Since the inception of multi-hop wireless networks, throughput and energy are two key performance metrics that bear in the minds of network designers and operators. Throughput is clearly the first and foremost performance consideration, as users of a multi-hop wireless network increasingly wish such network can offer comparable experience as its counterpart wireline networks. On the other hand, energy consumption is also regarded as a key performance consideration, as many types of multi-hop wireless networks (e.g., ad hoc network, sensor network) are battery-powered and are constrained with limited energy at each node.

To date, there is a vast amount of literature on optimizing throughput or energy. For network throughput, people have been trying to maximize it either at different layers (e.g., throughput-efficient scheduling algorithms [14, 72, 112, 131], throughput-efficient routing algorithms [12, 22,

89]) or jointly across multiple layers (e.g., [1, 2, 13, 16, 70, 73]). For energy, people are trying to conserve/minimize its consumption while meeting certain service requirements (e.g., energy-efficient scheduling and MAC schemes [36, 64, 109, 111, 117, 133], or energy-efficient routing protocol [28, 50, 67]).

We have also witnessed quite a few studies exploring the interaction between network throughput and energy consumption in the context of either maximizing network throughput under energy (or power) constraints (e.g., [17, 24, 88]) or minimizing energy consumption while satisfying some throughput constraints (e.g., [9, 25, 70, 74, 88]). The only one previous work that studied the relationship between throughput and energy is [116], which considered a particular type of cell partitioned network. Although many of these prior efforts were able to offer some optimal solutions, there is still a critical need to have a systematic study on how to optimize both objectives simultaneously. In particular, none of the existing efforts is able to offer a holistic view on how the maximum network throughput changes as a function of network energy consumption for general multi-hop wireless networks, i.e., the so-called optimal throughput-energy curve (or envelope) in this chapter.

The significance of optimal throughput-energy curve is three-fold. First, it gives an envelope of the entire throughput-energy region, which offers a global perspective on the achievable throughput-energy tradeoff. In contrast, a solution to traditional problems such as maximizing throughput under energy constraints or minimizing energy under throughput constraints only represents a point on this curve or inside this region. Second, each time when the requirement on either network throughput or energy consumption changes, one can use the optimal throughput-energy curve to find a new optimal tradeoff between throughput and energy instantly, rather than resorting to solving a new optimization problem. Finally, the optimal throughput-energy curve shows us the existence of a *saturation point*, beyond which the throughput can no longer be further increased, regardless of how much additional energy is used.

In this chapter, we conduct a systematic study on the optimal relationship between network throughput and energy consumption for a multi-hop wireless network. We tackle this problem

through a multicriteria optimization formulation, i.e., maximizing network throughput *while* minimizing total power in the network. Our main contributions can be summarized as follows.

- By solving the multicriteria optimization problem, we find the the entire throughput-energy curve.
- We find a number of important properties associated with the optimal throughput-energy curve, such as non-decreasing, concave, the existence of a saturation point, and strictly increasing between zero and the saturation point.
- For the case study, we consider two cases where the throughput functions are linear and nonlinear, respectively.
  - For the linear case, we show that the optimal throughput-energy curve can be characterized exactly via parametric analysis.
  - For the nonlinear case, we show that the optimal throughput-energy curve can be approximated by piece-wise linear segments with arbitrary desired accuracy.

The remainder of this chapter is organized as follows. In Section 7.2, we describe our network model. In Section 7.3, we present a multicriteria formulation that maximizes network throughput while minimizing energy consumption in a multi-hop wireless network. We show that finding the optimal solution to this multicriteria optimization problem is equivalent to finding the optimal throughput-energy curve. We also present some important properties associated with the optimal throughput-energy curve. In Section 7.4, we discuss approaches to obtain throughput-energy curves in practice. Section 7.5 and Section 7.6 present two case studies when the throughput functions are linear and nonlinear, respectively. Section 7.7 concludes this chapter.

Table 7.1: Notation in Chapter 7.

Symbol	Definition
B_l	Bandwidth on link l
C_l	Average rate of link l
d_l	Distance between the transmitter and the receiver of link l
$\mathrm{dst}(m)$	Destination node of session $m \in \mathcal{M}$
$f(P)$	The optimal throughput-energy curve
g_l	Channel gain on link l
$h(\cdot)$	A utility function
$\mathcal{L}$	The set of links in the network
$\mathcal{L}_i^{\mathrm{In}}$	The set of incoming links at node i
$\mathcal{L}_i^{\mathrm{Out}}$	The set of outgoing links at node i
$\mathcal{M}$	The set of user sessions in the network
$\mathcal{N}$	The set of nodes in the network
P	Rate of energy consumption in the network,
	$P = \sum_{l \in \mathcal{L}} \alpha_l \cdot (P_T + P_R)$
P_T	Rate of energy consumption for transmission at a node
P_R	Rate of energy consumption for reception at a node
$r(m)$	Data rate of session $m \in \mathcal{M}$
$r_l(m)$	Data rate on link l that is attributed to session m
$\mathrm{src}(m)$	Source node of session m
U	$= \sum_{m \in \mathcal{M}} h[r(m)]$, the network throughput utility
$w(m)$	A weight associated with session $m \in \mathcal{M}$
$\mathbf{x}$	$= \{r(m), r_l(m), \alpha_l \mid l \in \mathcal{L}, m \in \mathcal{M}\}$, a solution to our optimization problems
α_l	The fraction of time within a time frame when link l is active
γ	Path loss index
η	Ambient Gaussian noise density

7.2 Network Model

We consider a general multi-hop wireless network with a set of $\mathcal{N}$ nodes. A directed link $(i, j), i, j \in \mathcal{N}$ from node i to node j exists if and only if node j is within the transmission range of node i. Denote $\mathcal{L}$ the set of directed links in the network. To focus on throughput and energy performance, we simplify link layer scheduling by employing orthogonal channels among the links,[1] similar to that in [16, 65, 82]. Table 7.1 lists all notations used in this chapter.

Denote $\mathcal{M}$ a set of user (unicast) communication sessions in the network. Denote $\mathrm{src}(m)$ and $\mathrm{dst}(m)$ the source and destination nodes of session $m \in \mathcal{M}$, respectively. Denote $r(m)$ the rate of session $m \in \mathcal{M}$. We consider a general flow routing strategy where flow splitting (i.e., multi-path) is allowed. On link l, denote $r_l(m)$ the data rate that is attributed to session $m \in \mathcal{M}$. Denote $\mathcal{L}_i^{\mathrm{Out}}$ and $\mathcal{L}_i^{\mathrm{In}}$ the sets of potential outgoing and incoming links at node i, respectively. Then we have the following flow balance equations for multi-hop routing.

- If node i is the source node of session m, i.e., $i = \mathrm{src}(m)$, then

$$\sum_{l \in \mathcal{L}_i^{\mathrm{Out}}} r_l(m) = r(m) \ . \tag{7.2.1}$$

- If node i is an intermediate relay node along the path of session m, i.e., $i \neq \mathrm{src}(m)$ and $i \neq \mathrm{dst}(m)$, then

$$\sum_{l \in \mathcal{L}_i^{\mathrm{Out}}}^{l \neq (i, \mathrm{src}(m))} r_l(m) = \sum_{l \in \mathcal{L}_i^{\mathrm{In}}}^{l \neq (\mathrm{dst}(m), i)} r_l(m) \ . \tag{7.2.2}$$

- If node i is the destination node of session m, i.e., $i = \mathrm{dst}(m)$, then

$$\sum_{l \in \mathcal{L}_i^{\mathrm{In}}} r_l(m) = r(m) \ . \tag{7.2.3}$$

[1] An upper bound on the number of required orthogonal channels is $1 + d_v$, where d_v is the maximum vertex degree in the conflict graph in the final flow routing solution. More efficient channel assignment algorithms may further reduce the number of required channels. But problem of channel assignment is beyond the scope of this chapter.

It can be easily verified that once (7.2.1) and (7.2.2) are satisfied, then (7.2.3) is also satisfied. As a result, it is sufficient to list only (7.2.1) and (7.2.2) in a formulation.

For power control at each node, we employ a simple "on/off" control, which has been used for energy-saving in wireless networks (see e.g., [77, 87]). When a link is "on", the transmitter of this link transmits at a fixed power level P_T; when the link is "off" (for energy conservation), the transmitter of this link does not expend any power for transmission. To quantify the percentage of time that the link is in different states, denote α_l ($0 \leq \alpha_l \leq 1, l \in \mathcal{L}$) the fraction of time within a time frame that link l is "on".

Based on this on/off energy conservation model, the average rate of link l can be computed as

$$C_l = \alpha_l \cdot B_l \log_2(1 + \frac{P_T \cdot g_l}{\eta B_l}) \ , \tag{7.2.4}$$

where B_l is the bandwidth of link l under a given channel assignment, g_l is channel gain between the transmitter and receiver of link l and η is the ambient Gaussian noise density. Note the absence of an interference term in (7.2.4), which is due to our use of orthogonal channels in the network.

On link l, we have the following flow rate constraint:

$$\sum_{m \in \mathcal{M}} r_l(m) \leq C_l, \text{ for all } l \in \mathcal{L} \ , \tag{7.2.5}$$

which states that the aggregate flow rates from all sessions traversing link l cannot exceed the achievable rate of this link.

7.3 Throughput-Energy Curve and Its Properties

7.3.1 Multicriteria Formulation

In this chapter, we are interested in a multicriteria optimization problem, i.e., how to maximize network throughput while minimizing energy consumption at the same time. We now give a formulation of this problem.

Denote $h(\cdot)$ as a continuous, concave, and nondecreasing utility function. We define the network throughput utility U as follows:

$$U = \sum_{m \in \mathcal{M}} h[r(m)] \ ,$$

where $r(m)$ is the rate of session $m \in \mathcal{M}$. Note that in the special case when $h[r(m)] = r(m)$, then U is simply the sum of throughput in the network; in the case when $h[r(m)] = \ln[r(m)]$, U is called proportional fairness [63].

Now we consider energy consumption. Note that when a link is active, the rate of energy consumption includes energy consumption for both transmission and reception. Then the network energy consumption rate P in the network can be defined as follows:

$$P = \sum_{l \in \mathcal{L}} \alpha_l \cdot (P_T + P_R) \ ,$$

where α_l is the fraction of time within a time frame that link l is active, P_T is the transmission power, and P_R is the reception power. For simplicity, we assume that all nodes have the same transmission power and reception power.

With the above two definitions, our multicriteria optimization problem can be formulated as follows.

$$\begin{aligned} \text{MOPT} \quad \min \quad & P = \sum_{l \in \mathcal{L}} \alpha_l \cdot (P_T + P_R) \\ \max \quad & U = \sum_{m \in \mathcal{M}} h[r(m)] \\ \text{s.t.} \quad & \text{Constraints } (7.2.1), (7.2.2), (7.2.4) \text{ and } (7.2.5) \\ & r(m), r_l(m) \geq 0, 0 \leq \alpha_l \leq 1. \end{aligned}$$

Note that the two objective functions, P and U, are *conflicting* objectives. For example, when P is minimized (i.e., 0), U is also 0 and is not maximized. So there does not appear to exist an optimal solution to our problem that optimizes both objectives simultaneously.

Given that an optimal solution does not exist, a natural question to ask is what kind of solutions should we pursue when investigating problem MOPT? Before answering this question, it is

important to clarify how we compare two feasible solutions. We use

$$\mathbf{x} = \{r(m), r_l(m), \alpha_l | l \in \mathcal{L}, m \in \mathcal{M}\}$$

to represent a solution. Denote (P_1, U_1) and (P_2, U_2) the objective pairs of two different feasible solutions $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively. We say objective pair (P_1, U_1) dominates (P_2, U_2) if $P_1 \leq P_2$ and $U_1 \geq U_2$. This means that solution $\mathbf{x}_1$ uses no more energy than solution $\mathbf{x}_2$ to achieve the same or more throughput, i.e., $\mathbf{x}_1$ is better than $\mathbf{x}_2$. With this clarification, it is clear that our goal should be to find solutions that are not dominated by any other solutions. That is, we want to find solutions with their objective pair $(P^\dagger, U^\dagger)$ such that there does *not* exist another solution with objective pair (P, U) such that $P \leq P^\dagger$ and $U \geq U^\dagger$. Such solutions are called *Pareto optimal solutions* (also called efficient solutions in [29]) and the objective value pair $(P^\dagger, U^\dagger)$ corresponding to a Pareto optimal solution is called a *Pareto optimal point*. Pareto optimal solutions are those that any further improvement in one objective will lead to a deterioration in the other objective.

For our problem, we find that it is difficult to obtain all Pareto optimal solutions directly. Instead, we can find a solution $\mathbf{x}^*$ with its objective pair (P^*, U^*) such that there does not exist another solution $\mathbf{x}$ with its objective pair (P, U) satisfying $P < P^*$ and $U > U^*$. That is, there does not exist a solution $\mathbf{x}$ that can use less energy than solution $\mathbf{x}^*$ to achieve more throughput. Such solutions are called *weakly Pareto optimal solutions* (also called weakly efficient solutions in [29]) and the objective value pair (P^*, U^*) corresponding to such a solution is called a *weakly Pareto optimal point*. Note that Pareto optimal points are also weakly Pareto optimal, but weakly Pareto optimal points are not always Pareto optimal. Weakly Pareto optimal solutions are those for which improvement in both objectives simultaneously is impossible, but improvement on one objective without deteriorating the other is possible. Once we find all the weakly Pareto optimal solutions, we can identify a subset of solutions that are Pareto optimal based on its definition.

7.3.2 Throughput-Energy Curve

Instead of solving MOPT directly, let's consider a simpler single objective optimization problem for a given P (i.e., fixing one of the objective values). That is,

$$\begin{aligned} \text{OPT}(P) \quad \max \quad & \sum_{m\in\mathcal{M}} h[r(m)] \\ \text{s.t.} \quad & \sum_{l\in\mathcal{L}} \alpha_l(P_T+P_R) = P \\ & \text{All constraints in MOPT} \\ & r(m), r_l(m) \geq 0, 0 \leq \alpha_l \leq 1 \ . \end{aligned} \tag{7.3.1}$$

We now show that the optimal solution to OPT(P) is a weakly Pareto optimal solution to MOPT.

Lemma 7.1. *Let $\mathbf{x}^* = \{r^*(m), r_l^*(m), \alpha_l^* | l \in \mathcal{L}, m \in \mathcal{M}\}$ be an optimal solution to OPT(P) for a given value of P^* with a corresponding objective value U^*, then $\mathbf{x}^*$ is a weakly Pareto optimal solution to MOPT.*

Proof. Our proof is based on contradiction. Suppose $\mathbf{x}^*$ is not a weakly Pareto optimal solution to MOPT. Then there must exist a feasible solution $\hat{\mathbf{x}} = \{\hat{r}(m), \hat{r}_l(m), \hat{\alpha}_l | l \in \mathcal{L}, m \in \mathcal{M}\}$ to MOPT such that $\hat{P} < P^*$ and $\hat{U} > U^*$, where $\hat{U} = \sum_{m\in\mathcal{M}} h[\hat{r}(m)]$ and $\hat{P} = \sum_{l\in\mathcal{L}} \hat{\alpha}_l(P_T+P_R)$.

Based on this feasible solution $\hat{\mathbf{x}}$, we can construct another feasible solution $\hat{\mathbf{y}} = \{\hat{r}(m), \hat{r}_l(m), \hat{\beta}_l | l \in \mathcal{L}, m \in \mathcal{M}\}$ to OPT(P) as follows. We keep data rates of all sessions and on all links unchanged, but increase $\hat{\alpha}_l$ on some links to $\hat{\beta}_l$ such that $\sum_{l\in\mathcal{L}} \hat{\beta}_l(P_T+P_R) = P^*$. Clearly, $\hat{\mathbf{y}}$ is also a feasible solution, despite a waste of energy resource. Since the data rates of all sessions are unchanged in solution $\hat{\mathbf{y}}$, the objective value of this new solution remains the same as that for solution $\hat{\mathbf{x}}$, i.e., $\hat{U}$. That is, corresponding to the same P^* value, we have two feasible solution $\hat{\mathbf{y}}$ (with objective $\hat{U}$) and $\mathbf{x}^*$ with U^* and that $\hat{U} > U^*$. This contradicts the fact that $\mathbf{x}^*$ is an optimal solution to OPT(P) under a given P^* value. This completes our proof. □

Denote the range of P to be $[0, P_{max}]$, where P_{max} can be obtained by setting $\alpha_l = 1$ for all $l \in \mathcal{L}$. That is, $P_{max} = \sum_{l\in\mathcal{L}}(P_T+P_R) = |\mathcal{L}| \cdot (P_T+P_R)$. If one can enumerate all

possible $P \in [0, P_{max}]$ and obtain their corresponding optimal solutions via OPT(P), then based on Lemma 7.1, all these solutions are weakly Pareto optimal solutions.

Now we show the converse is also true, i.e., any weakly Pareto optimal point (P, U) of MOPT can be obtained by a corresponding problem of OPT(P).

Lemma 7.2. *Each weakly Pareto optimal point (P, U) of MOPT can be obtained by solving an instance of OPT(P).*

Proof. Suppose (P, U) is a weakly Pareto optimal point for MOPT and the solution to MOPT corresponding to (P, U) is $\mathbf{x} = \{r(m), r_l(m), \alpha_l | l \in \mathcal{L}, m \in \mathcal{M}\}$. We can conclude that (P, U) can be obtained by solving OPT(P) and the optimal objective value of OPT(P) is U.

Assume that the optimal solution of OPT(P) is $\mathbf{x}^* = \{r^*(m), r_l^*(m), \alpha_l^* | l \in \mathcal{L}, m \in \mathcal{M}\}$. If the above conclusion doesn't hold, we have that the optimal objective value of OPT(P) is $U^* \neq U$. For the case where $U^* < U$, since $\mathbf{x}$ is feasible to MOPT and $\sum_{l \in \mathcal{L}} \alpha_l \cdot (P_T + P_R) = P$, $\mathbf{x}$ is also a feasible solution to OPT(P). But we have that the objective value U of solution $\mathbf{x}$ is greater than U^*, which is a contradiction to $\mathbf{x}^*$ being the optimal solution to OPT(P).

Now, we discuss the case where $U^* > U$. For a small value δ, it is easy to see that the difference between optimal objective values of OPT(P) and OPT$(P - \delta)$ goes to zero as δ goes to zero. Denote $\bar{U}$ the optimal objective value of OPT$(P - \delta)$. Then, there exists a δ such that $\bar{U}$ can be very close to U^* and $\bar{U} > U$. Based on Lemma 7.1, we know that $(P - \delta, \bar{U})$ is a weakly Pareto optimal point of MOPT. However, point $(P - \delta, \bar{U})$ achieves more throughput by using less energy compared to point (P, U). This is a contradiction to (P, U) being a weakly Pareto optimal point of MOPT.

Thus, we have shown that $U^* \neq U$ does not hold and any weakly Pareto optimal point (P, U) of MOPT can be obtained by solving OPT(P). □

Based on Lemmas 7.1 and 7.2, we conclude that each weakly Pareto optimal point (P, U) of MOPT uniquely corresponds to the same (P, U) generated by an optimal solution of OPT(P).

Thus, by finding the optimal U for each OPT(P), $P \in [0, P_{max}]$, we can obtain all the weakly Pareto optimal points of MOPT. This gives us a mapping from P to U, which we denote as $f : P \rightarrow U$. Intuitively, this says that for any weakly Pareto optimal point (P, U), $U = f(P)$ is the maximum throughput utility that the network can deliver. Note that function $U = f(P)$ defines the envelope of the entire throughput-energy region, which we formally define as follows.

Definition 7.1. (Optimal Throughput-Energy Curve) *For all $P \in [0, P_{max}]$, the mapping $f : P \rightarrow U$ via solving OPT(P) constitutes an optimal throughput-energy curve $U = f(P)$.*

7.3.3 Key Properties

In this section, we present several interesting properties for the optimal throughput-energy curve. These properties are important for us to understand the fundamental behavior of this curve and to characterize this curve under specific throughput utility functions in the next section.

Property 7.1. *$U = f(P)$ is a nondecreasing function over $0 \leq P \leq P_{max}$.*

This property is easy to understand intuitively. It says that the throughput will not decrease when energy is increased. The proof is quite straightforward and is omitted.

Property 7.2. *$U = f(P)$ is a concave function.*

Proof. Based on the definition of a concave function, we need to prove that for any P_1, P_2, and λ ($0 \leq P_1 \leq P_{max}$, $0 \leq P_2 \leq P_{max}$, and $0 \leq \lambda \leq 1$), we have $f[\lambda P_1 + (1-\lambda)P_2] \geq \lambda f(P_1) + (1-\lambda) f(P_2)$.

Suppose the optimal solution to OPT(P_1) is $\mathbf{x}_1 = \{r^{(1)}(m), r_l^{(1)}(m), \alpha_l^{(1)} | l \in \mathcal{L}, m \in \mathcal{M}\}$ with the optimal objective value $f(P_1)$. Suppose the optimal solution to OPT(P_2) is $\mathbf{x}_2 = \{r^{(2)}(m), r_l^{(2)}(m), \alpha_l^{(2)} | l \in \mathcal{L}, m \in \mathcal{M}\}$ with the optimal objective value $f(P_2)$. We will construct a feasible solution to OPT$(\lambda P_1 + (1-\lambda)P_2)$ with an objective value being at least $\lambda f(P_1) + (1-\lambda) f(P_2)$.

Then, for the optimal solution to OPT$(\lambda P_1 + (1-\lambda)P_2)$, its optimal objective value $f(\lambda P_1 + (1-\lambda)P_2)$ is at least $\lambda f(P_1) + (1-\lambda)f(P_2)$.

The constructed solution is $\lambda \cdot \mathbf{x}_1 + (1-\lambda) \cdot \mathbf{x}_2$. We now show that it is a feasible solution to OPT$(\lambda P_1 + (1-\lambda)P_2)$. First, we verify that $\lambda \cdot \mathbf{x}_1 + (1-\lambda) \cdot \mathbf{x}_2$ satisfies constraint (7.3.1). We have

$$\begin{aligned} & \sum_{l \in \mathcal{L}} [\lambda \alpha_l^{(1)} + (1-\lambda)\alpha_l^{(2)}](P_T + P_R) \\ = \; & \lambda \sum_{l \in \mathcal{L}} \alpha_l^{(1)}(P_T + P_R) + (1-\lambda) \sum_{l \in \mathcal{L}} \alpha_l^{(2)}(P_T + P_R) \\ = \; & \lambda P_1 + (1-\lambda)P_2 \ , \end{aligned}$$

i.e., constraint (7.3.1) holds. Second, we note that the region defined by constraints (7.2.1), (7.2.2), (7.2.4) and (7.2.5) is a convex region. Since $\mathbf{x}_1$ and $\mathbf{x}_2$ are both in this region, we have that $\lambda \cdot \mathbf{x}_1 + (1-\lambda) \cdot \mathbf{x}_2$ is also in this region, i.e., $\lambda \cdot \mathbf{x}_1 + (1-\lambda) \cdot \mathbf{x}_2$ satisfies constraints (7.2.1), (7.2.2), (7.2.4) and (7.2.5). Therefore, the constructed solution satisfies all constraints in OPT$(\lambda P_1 + (1-\lambda)P_2)$ and thus is feasible.

Now we calculate the objective value achieved by solution $\lambda \cdot \mathbf{x}_1 + (1-\lambda) \cdot \mathbf{x}_2$, which is

$$\begin{aligned} & \sum_{m \in \mathcal{M}} h\left[\lambda r^{(1)}(m) + (1-\lambda) r^{(2)}(m)\right] \\ \geq \; & \sum_{m \in \mathcal{M}} \left\{\lambda h[r^{(1)}(m)] + (1-\lambda) h[r^{(2)}(m)]\right\} \\ = \; & \lambda \sum_{m \in \mathcal{M}} h\left[r^{(1)}(m)\right] + (1-\lambda) \sum_{m \in \mathcal{M}} h\left[r^{(2)}(m)\right] \\ = \; & \lambda f(P_1) + (1-\lambda) f(P_2) \ , \end{aligned}$$

where the first inequality holds since $h(\cdot)$ is a concave function.

Therefore, we have constructed a feasible solution to OPT$(\lambda P_1 + (1-\lambda)P_2)$ with its objective value greater than or equal to $\lambda f(P_1) + (1-\lambda)f(P_2)$. This completes the proof. □

The next two properties further spell out the shape of the concave throughput-energy curve.

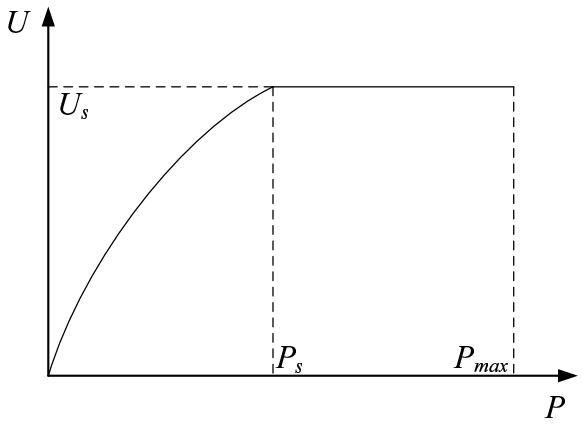


Figure 7.1: The shape of an optimal throughput-energy curve.

Property 7.3. *There is a saturation point (P_s, U_s) on the optimal throughput-energy curve $f(P)$ such that $f(P) = U_s$, for $P \in [P_s, P_{max}]$ and $f(P) < U_s$ for $P < P_s$.*

Proof. We prove this property by construction. We compute the saturation point (P_s, U_s) as follows.

We first compute the maximum achievable network throughput U_s under OPT(P_{max}). Once we have U_s, we can find the minimum network energy consumption rate P_s that can achieve this U_s by solving the following optimization problem:

$$\begin{aligned} P_s = \quad & \min \quad \sum_{l \in \mathcal{L}} \alpha_l \cdot (P_T + P_R) \\ & \text{s.t.} \quad \sum_{m \in \mathcal{M}} h[r(m)] \geq U_s \\ & \quad \text{Constraints } (7.2.1), (7.2.2), (7.2.4) \text{and } (7.2.5). \end{aligned}$$

Since a throughput-energy curve is a non-decreasing function (Property 7.1) and that we have $f(P_s) = f(P_{max}) = U_s$, the throughput-energy curve must be flat between $[P_s, P_{max}]$. Since P_s is the minimum energy that achieves U_s, based on Property 7.1, we have $f(P) < U_s$ for $P < P_s$. □

The above property says that the last segment of the optimal throughput-energy curve is *flat* after the saturation point (see Fig. 7.1)

The following property says that the segment of the optimal throughput-energy curve is strictly increasing for $P \in [0, P_s]$ (see Fig. 7.1).

Property 7.4. *$f(P)$ is a strictly increasing function for $P \in [0, P_s]$.*

Proof. Our proof is based on contradiction. Suppose $f(P)$ is not strictly increasing within $[0, P_s]$. Since $f(P)$ is nondecreasing (Property 7.1), then there must exist $0 < P_1 < P_2 < P_s$ such that $f(P_1) = f(P_2) < f(P_s)$. We can express P_2 as a linear combination of P_1 and P_s as follows:

$$P_2 = \lambda P_1 + (1 - \lambda) P_s \ ,$$

where $0 < \lambda < 1$. Then, we have

$$\begin{aligned} \lambda f(P_1) + (1 - \lambda) f(P_s) &> \lambda f(P_2) + (1 - \lambda) f(P_2) \\ &= f(P_2) \\ &= f[\lambda P_1 + (1 - \lambda) P_s] \ , \end{aligned}$$

where the first inequality holds since $f(P_1) = f(P_2) < f(P_s)$ and the third inequality holds since $P_2 = \lambda P_1 + (1 - \lambda) P_s$. But this contradicts the fact that $f(P)$ is a concave function. □

Recall that all the weakly Pareto optimal points of MOPT coincide with the optimal throughput-energy curve $f(P)$ over $P \in [0, P_{max}]$. It is easy to see that the points on $f(P)$ over $P \in [0, P_s]$ are Pareto optimal points (while those on $f(P)$ over $P \in (P_s, P_{max}]$ are only weakly Pareto optimal points).

7.4 A Naive Approach vs. Performance Guarantee

Although we have successfully analyzed some key properties of the optimal throughput-energy curve, it remains difficult to characterize the entire curve for a given throughput utility function. A naive approach to approximate the curve could be as follows. We can discretize the energy interval $[0, P_s]$ into a large number of equally spaced intervals. For each energy consumption value, P_i, we

can compute its corresponding throughput value $f(P_i)$ by solving OPT(P_i). So we obtain a point $(P_i, f(P_i))$ on the throughput-energy curve. Upon finding all these points on the curve, we can connect them via linear segments. This will give an approximate throughput-energy curve.

Although the above naive approach is simple and straightforward, it does not offer any performance guarantee of the curve. In contrast, one of the goals of this chapter is to characterize the curve with performance guarantee. In the following two sections, we consider two classes of throughput utility functions: the linear case and the non-linear case. In the linear case, we are able to characterize the optimal curve exactly by exploiting some special structures of linear program; for the nonlinear case, we develop a novel technique to approximate the curve with $(1-\varepsilon)$-optimal performance guarantee, where ε is an arbitrary small error reflecting our desired accuracy.

7.5 Case 1: Linear Throughput Function

In this section, we consider the case where the throughput utility function is linear with respect to $r(m)$, $m \in \mathcal{M}$. That is, $U = \sum_{m\in\mathcal{M}} w(m)r(m)$, where $w(m)$ is a constant and can be considered as the weight for session $m \in \mathcal{M}$. In this case, our OPT(P) becomes the following LP.

$$\begin{aligned} \text{LP}(P) \quad \max \quad & U = \sum_{m\in\mathcal{M}} w(m)r(m) \\ \text{s.t.} \quad & \text{All constraints in OPT(P)} \\ & r(m), r_l(m) \geq 0, 0 \leq \alpha_l \leq 1 \ . \end{aligned}$$

Instead of obtaining the $f(P)$ curve by solving LP(P) for all possible $P \in [0, P_{max}]$, which is impractical, we will exploit the special structure of LP and obtain the exact $f(P)$ curve by solving a *finite* number of LPs. In particular, since LP(P) is parametric linear program with respect to P, we propose to employ the so-called *parametric analysis* (PA) technique [6, Ch. 6] to obtain $f(P)$ curve efficiently.

7.5.1 Finding $f(P)$ Curve via Parametric Analysis

The main idea of PA is to investigate how a perturbation on parameter P will affect the optimality of LP(P). For a given value of P, the current optimal basis of LP(P) could still be optimal when there is a perturbation on P. Thus, the interval $[0, P_s]$ can be partitioned into small consecutive intervals, each corresponding to a different optimal basis. Within each small interval, the optimal basis to LP(P) is the same even when P varies. Further, we will show that $f(P)$ is linear within each small interval.

Partition $[0, P_s]$ into Smaller Intervals. We now show how to partition interval $[0, P_s]$ into small intervals. Rewrite LP(P) in the standard form **Max** $\mathbf{cx}$, s.t. $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$. Here we use boldface to denote vectors and matrices. For a fixed value of P, this LP can be solved via standard technique in polynomial time [6, Ch. 8]. Corresponding to this P, we assume that the optimal solution to LP(P) is $(\mathbf{x}_\mathcal{B}, \mathbf{x}_\mathcal{Q})$, where $\mathbf{x}_\mathcal{B}$ and $\mathbf{x}_\mathcal{Q}$ denote the values of basic and non-basic variables, respectively; $\mathcal{B}$ and $\mathcal{Q}$ denote the sets of basic and non-basic variables, respectively. Denote $\mathbf{B}$ and $\mathbf{Q}$ the optimal basic matrix and nonbasic matrix for LP(P), respectively. Then, $\mathbf{B}$ is the nonsingular sub-matrix of $\mathbf{A}$, which includes the columns of $\mathbf{A}$ corresponding to the basic variables in set $\mathcal{B}$, and $\mathbf{Q}$ is the sub-matrix of $\mathbf{A}$, which includes the columns of $\mathbf{A}$ corresponding to the nonbasic variables in set $\mathcal{Q}$. Denote $\mathbf{c}_\mathcal{B}$ and $\mathbf{c}_\mathcal{Q}$ the objective function coefficient vectors of throughput utility U for the basic and non-basic variables, respectively. Then we can write the corresponding canonical equations as follows [6, Ch. 6]:

$$U + (\mathbf{c}_\mathcal{B}^T \mathbf{B}^{-1}\mathbf{Q} - \mathbf{c}_\mathcal{Q})\mathbf{x}_\mathcal{Q} = \mathbf{c}_\mathcal{B}^T \mathbf{B}^{-1}\mathbf{b} \,, \tag{7.5.1}$$

$$\mathbf{x}_\mathcal{B} + \mathbf{B^{-1}Q}\mathbf{x}_\mathcal{Q} = \mathbf{B}^{-1}\mathbf{b} \,. \tag{7.5.2}$$

Note that when $\mathbf{x}_\mathcal{B}$ and $\mathbf{x}_\mathcal{Q}$ are optimal solutions, we have $\mathbf{x}_\mathcal{Q} = \mathbf{0}$ [6, Ch. 3]. Thus, based on (7.5.2), we have

$$\mathbf{x}_\mathcal{B} = \mathbf{B^{-1}b} \,.$$

Suppose that we do a perturbation on parameter P, i.e., we change P to $P + \delta$. Then vector $\mathbf{b}$ becomes $\mathbf{b} + (\delta, 0, \ldots, 0)^T$. The only change due to this perturbation is that $\mathbf{B}^{-1}\mathbf{b}$ will be replaced

by $\mathbf{B}^{-1}(\mathbf{b}+\delta\mathbf{I})$, where vector $\mathbf{I}$ has a single 1 on the first element and zero on all the others. Note that $\mathbf{x}_{\mathcal{B}} = \mathbf{B}^{-1}(\mathbf{b}+\delta\mathbf{I})$ is a basic feasible solution (BFS). As long as $\mathbf{B}^{-1}(\mathbf{b}+\delta\mathbf{I})$ is nonnegative, the current basis remains optimal. This is because that changing $\mathbf{b}$ to $\mathbf{b}+\delta\mathbf{I}$ does not affect the correctness of (7.5.1) and (7.5.2).

On the other hand, when one of the elements in $\mathbf{B}^{-1}(\mathbf{b}+\delta\mathbf{I})$ becomes negative, the optimal basis must change. Otherwise, we will have one negative element in $\mathbf{x}_{\mathcal{B}}$, which contradicts $\mathbf{x} \geq \mathbf{0}$ in the LP formulation. The value of δ at which this change occurs can be determined as follows. Denote $\bar{\mathbf{b}} = \mathbf{B}^{-1}\mathbf{b}$ and $\bar{\mathbf{b}}' = \mathbf{B}^{-1}\mathbf{I}$, and let $\mathcal{S} = \{i : \bar{b}'_i < 0\}$, where $\bar{b}'_i$ is the i-th element in vector $\bar{\mathbf{b}}'$. If $\mathcal{S} = \emptyset$, then the current basis is optimal for all values of $\delta \geq 0$ since all elements in vector $\mathbf{B}^{-1}(\mathbf{b}+\delta\mathbf{I})$ are nonnegative. Otherwise, let

$$\hat{\delta} = \min_{i \in \mathcal{S}} \left\{ \frac{\bar{b}_i}{-\bar{b}'_i} \right\} . \tag{7.5.3}$$

For $\delta \in [0, \hat{\delta}]$, the current basis $\mathbf{B}$ remains optimal and its corresponding BFS is $\mathbf{x}_{\mathcal{B}} = \mathbf{B}^{-1}(\mathbf{b}+\delta\mathbf{I})$. When $\delta > \delta'$, the basis $\mathbf{B}$ is no longer optimal. Thus, we need to choose the variable x_r to leave the basis, where the minimum in (7.5.3) is attained for $i = r$. The entering variable x_s is chosen by the dual simplex method rule [6, Ch. 6]. Based on the new optimal basis obtained after the pivot, we can update the corresponding canonical equations and get a (P, U) pair, which is an endpoint of the linear segment of $f(P)$.

Figure 7.2 lists the steps to obtain a new optimal basis for a given optimal basis $\mathcal{B}$. Thus, starting from $P = 0$, we can use this algorithm iteratively to find different bases until we reach P_s. The series of $\hat{\delta}$ for these bases will partition $[0, P_s]$ into small intervals.

The complexity of the basis updating algorithm can be analyzed as follows. The dominant computational complexity occurs in step 2: $\bar{\mathbf{A}} = \mathbf{B}^{-1}\mathbf{A}$. Note that our linear programming LP(P) has $Z = (1 + 2|\mathcal{L}| + |\mathcal{N}| - |\mathcal{M}|)$ constraints and $V = (|\mathcal{L}| \cdot |\mathcal{M}| + 2|\mathcal{L}| + |\mathcal{M}|)$ variables. Since $\bar{\mathbf{A}} = \mathbf{B}^{-1}\mathbf{A}$ involves matrix multiplication of a $Z \times Z$ matrix and a $Z \times V$ matrix, its complexity is $O(Z^2V) = O(|\mathcal{L}|^3|\mathcal{M}| + |\mathcal{N}|^2|\mathcal{L}||\mathcal{M}| + |\mathcal{N}||\mathcal{L}|^2|\mathcal{M}|)$.

Linearity of Each Small Interval. For each small interval with an optimal basis, we now

Basis Updating Algorithm

1. **Input:** An optimal basis $\mathcal{B}$ for a given P.
2. Compute $\bar{\mathbf{A}} = \mathbf{B}^{-1}\mathbf{A}$, $\bar{\mathbf{b}} = \mathbf{B}^{-1}\mathbf{b}$, and $\bar{\mathbf{b}}' = \mathbf{B}^{-1}\mathbf{I}$.
3. If $\mathcal{S} = \{i : \bar{b}'_i < 0\} = \emptyset$, stop.
4. $\hat{\delta} = \min_{i\in\mathcal{S}} \left\{ \frac{\bar{b}_i}{-\bar{b}'_i} \right\}$.
5. $r = \arg\min_i \left\{ \frac{\bar{b}_i}{-\bar{b}'_i} \right\}$.
6. $s = \arg\min_j \left\{ \frac{\bar{b}_j}{\bar{\mathbf{A}}_{jr}}, \bar{\mathbf{A}}_{jr} < 0 \right\}$.
7. Let $\mathcal{B} = (\mathcal{B}\backslash\{r\}) \cup \{s\}$ and $\mathbf{b} = \mathbf{b} + \hat{\delta}\mathbf{I}$.
8. Update $\mathbf{B}$ based on $\mathcal{B}$ ($\mathbf{B}$ consists of $\mathbf{A}$'s columns whose indices are in $\mathcal{B}$).
9. Compute $P = P + \hat{\delta}$, $\mathbf{x} = \mathbf{B}^{-1}\mathbf{b}$ and $U = \mathbf{c}_{\mathcal{B}}^T\mathbf{x}$.
10. **Output:** The new basis $\mathcal{B}$, $\mathbf{x}$, $\hat{\delta}$ and (P, U) pair.

Figure 7.2: The basis updating algorithm.

show that $f(P)$ is linear. Suppose interval $[0, P_s]$ is divided into K small intervals $[P_i, P_{i+1}], i = 1, \ldots, K$, where $P_1 = 0, P_{K+1} = P_s$, and the optimal basis for small $[P_i, P_{i+1}]$ is $\mathbf{B}_i$. Then, for an optimal basis $\mathbf{B}_i$ within a particular small interval $[P_i, P_{i+1}]$, the objective value of throughput $f(P), P_i \le P \le P_{i+1}$ can be computed as follows.

$$f(P) = \mathbf{c}_{\mathcal{B}_i}^T \mathbf{B}_i^{-1}(\mathbf{b} + \delta\mathbf{I}) \ , \tag{7.5.4}$$

where $\delta = P - P_i$. Substituting $\delta = P - P_i$ into (7.5.4), we have

$$f(P) = \mathbf{c}_{\mathcal{B}_i}^T \mathbf{B}_i^{-1}[\mathbf{b} + (P - P_i)\mathbf{I}] \ . \tag{7.5.5}$$

In (7.5.5), since $\mathbf{c}_{\mathcal{B}_i}^T, \mathbf{B}_i^{-1}, \mathbf{b}, \mathbf{I}$ and P_i are constants, and P is the only variable, we conclude that $f(P)$ is a linear function of P for $P_i \le P \le P_{i+1}$, $i = 1, \ldots, K$. We formally state this result in the following lemma.

Lemma 7.3. *For the linear case, the optimal throughput-energy curve $f(P)$ is piece-wise linear within $[0, P_s]$.*

Recall that by executing the basis updating algorithm sequentially, we also obtain a series of (P, U) pair and the solution $\mathbf{x}$ generating (P, U), each corresponding to an optimal basis. Since $f(P)$ is a piece-wise linear line with each linear segment determined by an optimal basis, the series of (P, U) pairs are the endpoints of these linear segments. Then, by connecting these endpoints consecutively, we are able to characterize the entire optimal throughput-energy curve $f(P)$.

7.5.2 From Curve to a Point

By obtaining the entire optimal throughput-energy curve $f(P)$, we also have the endpoints of each line segment on $f(P)$ and the solutions of all endpoints. We now show that the solution for any point on the optimal throughput-energy curve $f(P)$ can be easily calculated through linear combination of the solutions for the endpoints (instead of solving a new LP(P)).

Assume that we want to find the solution $\mathbf{x}$ for a point (P, U) on the optimal throughput-energy curve, which lies in the line segment with two ends (P_1, U_1) and (P_2, U_2), and the optimal solutions for (P_1, U_1) and (P_2, U_2) are $\mathbf{x}_1 = \{r^{(1)}(m), r_l^{(1)}(m), \alpha_l^{(1)} | l \in \mathcal{L}, m \in \mathcal{M}\}$ and $\mathbf{x}_2 = \{r^{(2)}(m), r_l^{(2)}(m), \alpha_l^{(2)} | l \in \mathcal{L}, m \in \mathcal{M}\}$, respectively. Then there exists a constant $0 \leq \lambda \leq 1$ such that $P = \lambda P_1 + (1 - \lambda) P_2$. The corresponding solution $\mathbf{x} = \{r(m), r_l(m), \alpha_l | l \in \mathcal{L}, m \in \mathcal{M}\}$ for point (P, U) can now be computed as $\mathbf{x} = \lambda \mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2$, which means that the optimal session rates $r(m)$, data flow rates $r_l(m)$ on each link l, and the fraction of active time on each link α_l in solution $\mathbf{x}$ is just a simple linear combination of solutions $\mathbf{x}_1$ and $\mathbf{x}_2$.

Thus, after we characterize the optimal throughput-energy curve $f(P)$, we can find an optimal solution for any point on the curve via linear combination of known solutions. We formally state this result in the following theorem.

Theorem 7.1. *Denote $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$ the optimal solutions for the two endpoints $(P_i, f(P_i))$ and $(P_{i+1}, f(P_{i+1}))$ of the i-th linear segment in $f(P)$. The optimal solution $\mathbf{x}$ for any point $(P, f(P))$ between $(P_i, f(P_i))$ and $(P_{i+1}, f(P_{i+1}))$, where $P = \lambda P_i + (1 - \lambda) P_{i+1}, 0 \leq \lambda \leq 1$, can be written as $\mathbf{x} = \lambda \mathbf{x}_i + (1 - \lambda)\mathbf{x}_{i+1}$.*

Table 7.2: Source and destination nodes of each session.

Session m	Source Node	Destination Node
1	N19	N15
2	N8	N12
3	N9	N3
4	N3	N5
5	N5	N1
6	N1	N12
7	N4	N11
8	N6	N10
9	N16	N6
10	N2	N10

Proof. Based on Lemmas 7.1 and 7.2, we know that $(P, f(P))$ can be obtained by solving LP(P). Now, we need to show that the optimal solution of LP(P), where $P = \lambda P_i + (1-\lambda)P_{i+1}, 0 \leq \lambda \leq 1$, is $\lambda \mathbf{x}_i + (1-\lambda)\mathbf{x}_{i+1}$. From the previous analysis, we know that basis $\mathbf{B}_i$ remains optimal for LP$(P), P \in [P_i, P_{i+1}]$. Rewrite LP$(P)$, LP$(P_i)$ and LP(P_{i+1}) under standard forms as **Max** $\mathbf{cx}$, s.t. $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$, **Max** $\mathbf{cx}$, s.t. $\mathbf{Ax} = \mathbf{b}_i$ and $\mathbf{x} \geq \mathbf{0}$ and **Max** $\mathbf{cx}$, s.t. $\mathbf{Ax} = \mathbf{b}_{i+1}$ and $\mathbf{x} \geq \mathbf{0}$, respectively. The only difference among $\mathbf{b}$, $\mathbf{b}_i$ and $\mathbf{b}_{i+1}$ is on the first element. The first elements of $\mathbf{b}$, $\mathbf{b}_i$ and $\mathbf{b}_{i+1}$ are P, P_i and P_{i+1}. Since $P = \lambda P_i + (1-\lambda)P_{i+1}$, we have

$$\mathbf{b} = \lambda \mathbf{b}_i + (1-\lambda)\mathbf{b}_{i+1} \ . \tag{7.5.6}$$

We also know that the optimal solutions of LP(P), LP(P_i) and LP(P_{i+1}) are $\mathbf{x} = \mathbf{B}_i^{-1}\mathbf{b}$, $\mathbf{x}_i = \mathbf{B}_i^{-1}\mathbf{b}_i$ and $\mathbf{x}_{i+1} = \mathbf{B}_i^{-1}\mathbf{b}_{i+1}$. Based on (7.5.6), we can conclude $\mathbf{x} = \lambda \mathbf{x}_i + (1-\lambda)\mathbf{x}_{i+1}$. This completes the proof. □

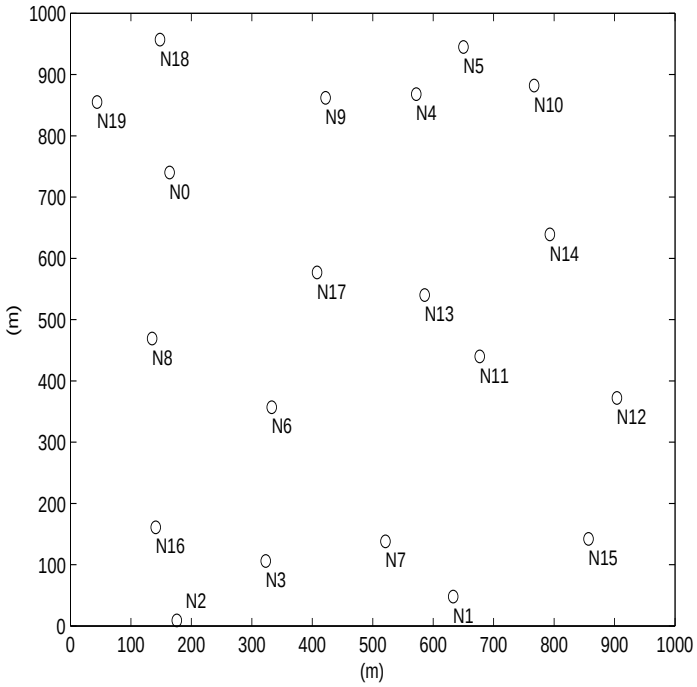
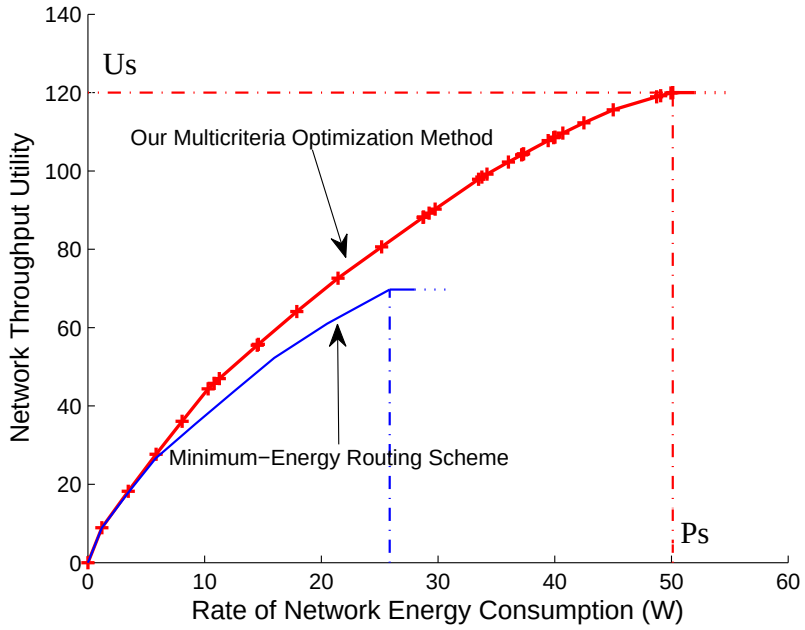


Figure 7.3: Topology for a 20-node network.

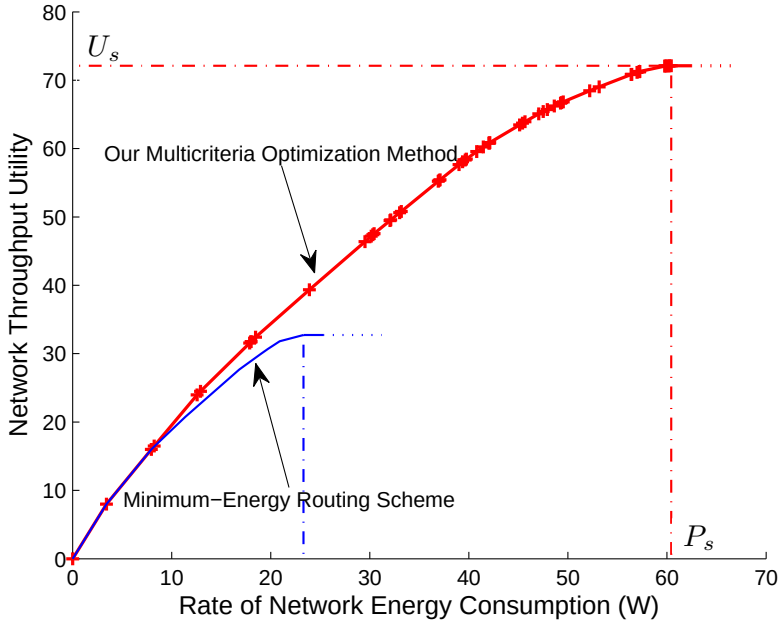
7.5.3 A Numerical Example

In the following, we present some pertinent numerical results to demonstrate our theoretical findings. We first describe our simulation settings. As shown in Fig. 7.3, we consider a randomly generated multi-hop wireless network with 20 nodes, which are distributed in a square region of 1000m$\times$1000m. The transmission power and reception power for each node are set to P_T = 1 W and P_R = 0.2 W. The bandwidth on each link is $B_l = 1$ MHz. We use a simplified channel gain model $g_l = d_l^{-\gamma}$, where d_l is the distance between the transmitter and receiver of link l and γ is the path loss index. We set $\gamma = 3$. There are ten user sessions in the network and Table 7.2 specifies the source and destination nodes of each session. For the weight $w(m)$ of each session $m \in \mathcal{M}$, we consider two scenarios: (i) equal weight, e.g., $w(m) = 1$ for all $m \in \mathcal{M}$; and (ii) random weight for each session.

The top curve in Fig. 7.4(a) shows the throughput-energy curve when each session has an

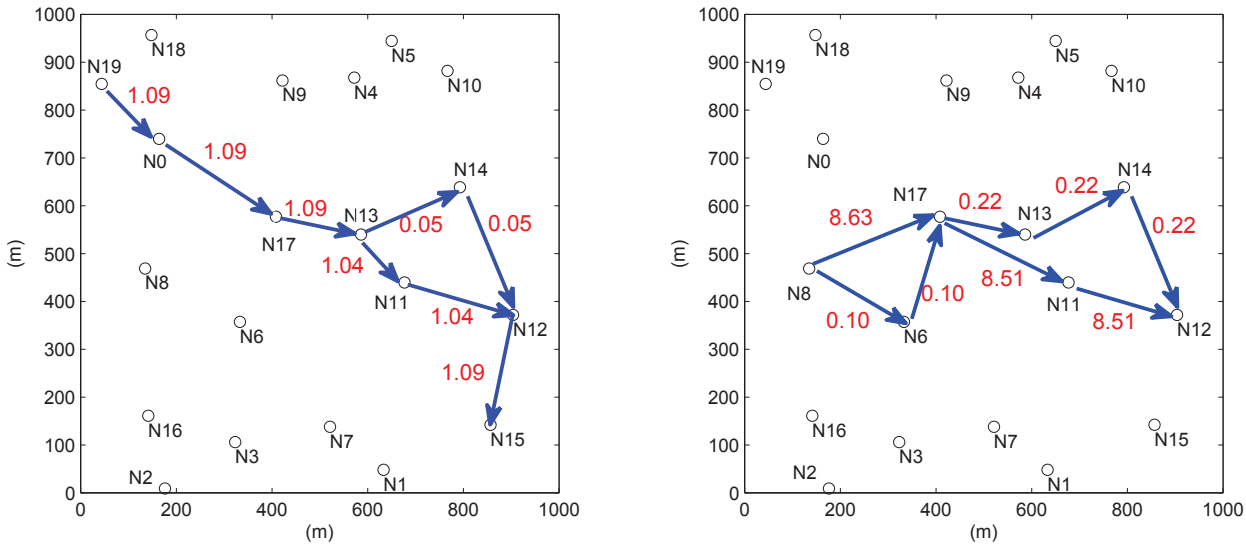


(a) Equal weight



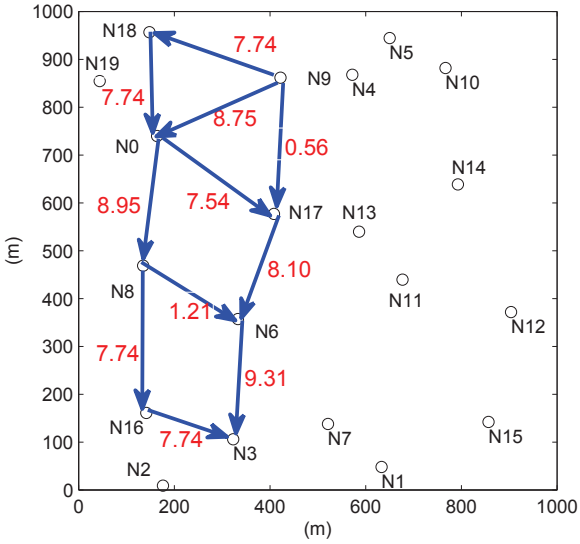
(b) Unequal weight

Figure 7.4: The throughput-energy curves for a 20-node example.



(a) Session 1

(b) Session 2



(c) Session 3

Figure 7.5: The optimal flow routing solutions for Sessions 1, 2, and 3 for the saturation point $(50.12, 120.02)$ in the example.

Table 7.3: α_l for each active link l in the example for the saturation point.

Active link	α_l	Active link	α_l	Active link	α_l	Active link	α_l	Active link	α_l
$(0,8)$	1.00	$(4,10)$	1.00	$(8,16)$	0.92	$(12,15)$	0.11	$(16,6)$	1.00
$(0,9)$	1.00	$(4,14)$	1.00	$(8,17)$	1.00	$(13,11)$	1.00	$(16,8)$	1.00
$(0,17)$	1.00	$(5,10)$	1.00	$(9,0)$	1.00	$(13,14)$	1.00	$(17,6)$	0.84
$(0,18)$	0.02	$(6,3)$	1.00	$(9,4)$	0.69	$(14,10)$	1.00	$(17,9)$	0.97
$(1,15)$	1.00	$(6,8)$	0.92	$(9,5)$	1.00	$(14,11)$	1.00	$(17,11)$	1.00
$(2,3)$	1.00	$(6,13)$	1.00	$(9,17)$	1.00	$(14,12)$	0.64	$(17,13)$	0.99
$(3,6)$	1.00	$(6,17)$	1.00	$(9,18)$	0.89	$(14,13)$	0.28	$(18,0)$	0.78
$(3,7)$	0.85	$(7,6)$	1.00	$(10,14)$	1.00	$(15,12)$	0.98	$(18,9)$	0.02
$(4,5)$	0.54	$(8,0)$	1.00	$(11,12)$	1.00	$(16,2)$	0.95	$(19,0)$	0.10
$(4,9)$	0.71	$(8,6)$	1.00	$(12,11)$	0.55	$(16,3)$	1.00		

equal weight of 1. At the saturation point, we have $P_s = 50.12$ W and $U_s = 120.02$. This curve is obtained by using PA method, which gives 33 endpoints that interconnect the piece-wise linear segments of $f(P)$. For comparison, the bottom curve in Fig. 7.4(a) shows the throughput-energy curve under the popular minimum energy routing scheme [106], where each session chooses the path that consumes the minimum energy. The minimum energy path for a session can be computed by using the well-known shortest path algorithms (e.g., Dijkstra's algorithm [20]), where the cost on link l is set to the total energy consumed to send one bit from a transmitter to a receiver, i.e., $(P_T + P_R)/C_l$. The large gap between throughput utility of the two curves shows that minimum-energy routing is far from optimal in terms of throughput-energy curve. This result affirms the importance of employing multicriteria formulation as we have done in this chapter.

Figure 7.4(b) shows the results for the case when the weight of each session is randomly chosen. The randomly generated weights for the ten sessions are 0.8147, 0.1270, 0.9134, 0.9134, 0.6324, 0.0975, 0.2785, 0.5469, 0.1270 and 0.9058, respectively. Again, the throughput-energy curve is of the same form as that in Fig. 7.4(a), as expected. At the saturation point, we have $P_s = 60.43$ W and $U_s = 72.11$. The bottom curve in Fig. 7.4(b) shows the throughput-energy curve under

minimum energy routing, which is far from optimal.

Note that for each endpoint on the curve, we also obtain its optimal solution for multi-hop routing variables $r(m), r_l(m)$ and α_l at each link. As an example, we show the optimal solution for the saturation point $(P_s, U_s) = (50.12, 120.02)$ under equal weight, which uses the minimum energy consumption of 50.12 W to achieve the maximum network throughput of 120.02 Mb/s. We find that the optimal data rates (in Mb/s) for the ten sessions are 1.09, 8.73, 17.05, 4.40, 0, 9.45, 25.90, 22.44, 30.96 and 0, respectively.[2] In this optimal solution, there are 49 active links in the network. Table 7.3 shows the fraction of time for each active link. Also, we find that some links never need to be activated to maximize throughput utility. Figure 7.5 shows the flow routing solution for Sessions 1, 2, and 3 (others are similar and are thus omitted). The number next to each arrow represents the data rate on that link that is attributed to that session.

7.6 Case 2: Nonlinear Throughput Function

In this section, we consider the case where the throughput utility function $h(\cdot)$ is a concave, but nonlinear function of $r(m), m \in \mathcal{M}$. In particular, we consider $h[r(m)] = \ln[r(m)], m \in \mathcal{M}$, which is called *proportional fairness* in [63]. In this case, for a given P, OPT(P) is a convex, nonlinear program. Although convex program OPT(P) can be solved efficiently for one given P, it is impractical to solve an *infinite* number of such convex problems when P varies from 0 to P_{max}. Further, due to nonlinearity, we cannot take advantage of the PA technique to compute the *exact* optimal throughput-energy curve efficiently.

Instead of finding the exact optimal throughput-energy curve, we propose a piece-wise linear approximation for this curve, where the approximation is guaranteed to be within $(1-\varepsilon)$-optimal, with ε being an arbitrary small number.

[2] We are aware that there is a fairness issue in this solution, due to the network throughput being defined as the weighted sum of all session rates. In the next session, we will show that fairness issue can be addressed when the throughput utility function is defined in terms of $\ln(\cdot)$.

7.6.1 Finding $f(P)$ Curve with $(1-\varepsilon)$ Optimality

Note that for a given P, we can always find a corresponding U on the optimal throughput-energy curve by solving a convex program (see Lemma 7.1). So the question becomes how to choose a set of such points and connect them with piece-wise linear segments so that this piece-wise linear approximation is no more than ε (in percentile) from the unknown optimal throughput-energy curve.

First, we identify the two endpoints on the optimal throughput-energy curve that we want to approximate. On the left side, since the throughput utility is a $\ln(\cdot)$ function, it is negative when P is small. Assuming that we are only interested in the optimal throughput-energy curve when $f(P) \geq 0$, we will pick a P, denoted as P_0, such that $U_0 = f(P_0)$ is just above zero.[3] On the right side, recall that the optimal throughput-energy curve $f(P)$ is flat from $P = P_s$ to $P = P_{max}$. So we can choose the saturation point (P_s, U_s) (see Section 7.3 on how to obtain it) as our right endpoint.

With our two endpoints on the optimal throughput-energy curve being (P_0, U_0) and (P_s, U_s), our approximation method works as follows (see Fig. 7.6). We connect points (P_0, U_0) and (P_s, U_s) with a linear segment a and consider it as our first approximation of the optimal throughput-energy curve. To examine if linear segment a is accurate enough, we compute an error upper bound σ of this approximation (in percentile). This is not trivial and will be shown in Lemma 7.4. If $\sigma \leq \varepsilon$, then our linear approximation is considered accurate enough and we are done. Otherwise, we will find a point (P^*, U^*) on the optimal throughput-energy curve and use two linear segments b and c as a better approximation. Again, finding this point (P^*, U^*) is not trivial (since the complete optimal throughput-energy curve is unknown) and will be explained shortly. Now the same process continues on linear segments b and c. The process continues until $\sigma \leq \varepsilon$ for every linear segment of the piece-wise linear approximation curve.

We first show how to compute (P^*, U^*), since we need (P^*, U^*) when computing σ.

[3]Note that $f(P_0) = 0$ cannot be our left endpoint due to the singularity it presents when we compute the approximation error (in percentile).

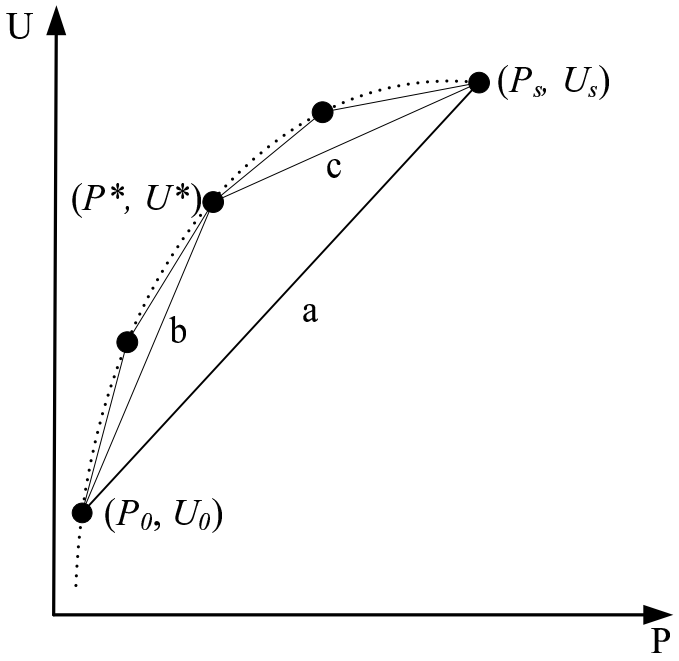


Figure 7.6: An illustration of our piece-wise linear approximation method.

Finding (P^*, U^*)**.** Point (P^*, U^*) has the maximum approximation error when we use a line segment to approximate a segment of the optimal throughput-energy curve (see Fig. 7.7).

Suppose that (P_1, U_1) and (P_2, U_2) are two endpoints of a line segment, which we denote as $\tilde{f}(P)$. Then this line segment $\tilde{f}(P)$ can be characterized as $\tilde{f}(P) = U_1 + \frac{U_2 - U_1}{P_2 - P_1}(P - P_1)$, $P_1 \leq P \leq P_2$. Although the optimal throughput-energy curve $f(P)$ is unknown, we imagine that we move line $\tilde{f}(P)$ upward until it is tangential to the curve. Denote this tangential point as (P^*, U^*), which is the point having the maximum absolute (rather than percentile) approximation error if we were to use $\tilde{f}(P)$ to approximate $f(P)$. Then, we have

$$\begin{aligned} & f(P^*) - \tilde{f}(P^*) \\ =& \max\{f(P) - \tilde{f}(P)\} \\ =& \max\{\sum_{m \in \mathcal{M}} h[r(m)] - [U_1 + \frac{U_2 - U_1}{P_2 - P_1}(P - P_1)]\} \ , \end{aligned}$$

for $P_1 \leq P \leq P_2$. Therefore, the tangential point (P^*, U^*) can be found by solving the following

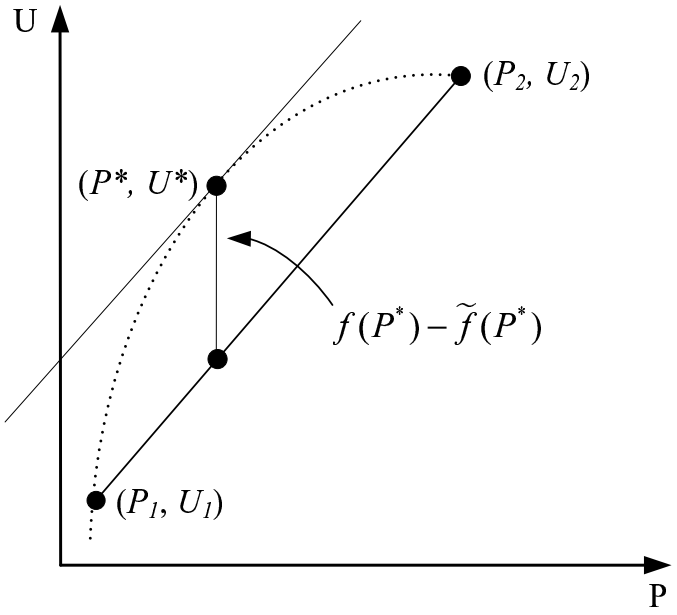


Figure 7.7: An illustration showing how to obtain the tangential point and maximum approximation error on one linear segment.

optimization problem.

$$\begin{aligned} \text{P-MAX} \quad \max \quad & \sum_{m\in\mathcal{M}} h[r(m)] - \left[U_1 + \frac{U_2 - U_1}{P_2 - P_1}(P - P_1)\right] \\ \text{s.t.} \quad & \sum_{l\in\mathcal{L}} \alpha_l (P_T + P_R) - P = 0 \\ & \text{All constraints in MOPT} \\ & P_1 \le P \le P_2 \ . \end{aligned}$$

Note that in the above optimization problem, P is a variable, which is different from OPT(P). The above optimization problem is a convex problem, which can be solved efficiently by using subgradient method [7, Ch. 8].

Finding σ. After obtaining the tangential point (P^*, U^*), we can calculate an upper bound σ of the approximation error (in percentile) with the following lemma.

Lemma 7.4. *By using $\tilde{f}(P)$ to approximate $f(P)$ for $P_1 \le P \le P_2$, an upper bound for this*

approximation error (in percentile) is

$$\sigma = \frac{1}{1 + \frac{U_1}{U^* - \tilde{f}(P^*)}} .$$

Proof. Referring to Fig. 7.7, for any point $(P, f(P))$ within $[P_1, P_2]$, the approximation error (in percentile) is

$$\frac{f(P) - \tilde{f}(P)}{f(P)} = \frac{f(P) - \tilde{f}(P)}{f(P) - \tilde{f}(P) + \tilde{f}(P)} = \frac{1}{1 + \frac{\tilde{f}(P)}{f(P) - \tilde{f}(P)}} .$$

Since $\tilde{f}(P) \geq f(P_1) = U_1$ and $f(P) - \tilde{f}(P) \leq U^* - \tilde{f}(P^*)$, we have

$$\frac{f(P) - \tilde{f}(P)}{f(P)} = \frac{1}{1 + \frac{\tilde{f}(P)}{f(P) - \tilde{f}(P)}} \leq \frac{1}{1 + \frac{U_1}{U^* - \tilde{f}(P^*)}} = \sigma .$$

□

Now given that we can compute σ at each iteration and our process stops when $\sigma \leq \varepsilon$ for each segment, it is not hard to see that our piece-wise linear approximation can guarantee $(1 - \varepsilon)$-optimality. We state this result in the following theorem.

Theorem 7.2. *For any small $\varepsilon > 0$, the proposed piece-wise linear approximation method can approximate the optimal throughput-energy curve $f(P)$ with $(1 - \varepsilon)$-optimality.*

Our proposed piece-wise linear approximation method involves computing a sequence of convex programming problems. In theory, the worst-case complexity of convex programming problems is NP-hard. But in practice, most convex programming problems (including ours) can be solved efficiently. For the numerical example in Section 7.6.3, it only took several seconds for our method to find the approximated curve.

7.6.2 From Approximated Curve to a Point

We have shown how to obtain a throughput-energy curve with $(1 - \varepsilon)$-optimal performance guarantee. Next, we show that for any point $(P, f(P))$ on the optimal (unknown) throughput-energy

curve, we can obtain a solution (which includes session rates, data flow rates, and the fraction of time for each link) with $(1-\varepsilon)$-optimality through linear combination of the solutions that we already have for the endpoints on the approximated curve. Note that this is much faster than solving a new convex programming problem (OPT(P)). We formally state this result in the following theorem.

Theorem 7.3. *Denote $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$ the optimal solutions for the two endpoints $(P_i, \tilde{f}(P_i))$ and $(P_{i+1}, \tilde{f}(P_{i+1}))$ of the i-th linear segment on the approximated curve $\tilde{f}(P)$. Denote $(P, f(P))$ a point on the optimal curve, where $P = \lambda P_i + (1-\lambda)P_{i+1}$, $P_i \leq P \leq P_{i+1}$, $0 \leq \lambda \leq 1$. Then the point $(P, \hat{U})$ generated by the solution $\hat{\mathbf{x}} = \lambda \mathbf{x}_i + (1-\lambda)\mathbf{x}_{i+1}$ is within $(1-\varepsilon)$-optimal from point $(P, f(P))$.*

Proof. We first show that $\hat{\mathbf{x}}$ is a feasible solution to MOPT. Note that solutions $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$ are obtained by solving P-MAX. It is easy to see that $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$ satisfy all the constraints in MOPT. Since the constraints in MOPT define a convex region, $\hat{\mathbf{x}} = \lambda \mathbf{x}_i + (1-\lambda)\mathbf{x}_{i+1}$ is also in this region. Thus, $\hat{\mathbf{x}}$ is feasible to MOPT.

For the energy consumption by solution $\hat{\mathbf{x}}$, it is easy to show that it is equal to P.

Next, we show that throughput $\hat{U}$ under $\hat{\mathbf{x}}$ is at least $\tilde{f}(P)$. That is, the throughput $\hat{U}$ under $\hat{\mathbf{x}}$ is greater than or equal to the throughput corresponding to the same P on the approximated curve.

Denote $\{r^{(i)}(m) | m \in \mathcal{M}\}$ and $\{r^{(i+1)}(m) | m \in \mathcal{M}\}$ the session data rates in solutions $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$, respectively. Then, we have that the data session rates in $\hat{\mathbf{x}}$ is $\{\lambda r^{(i)}(m) + (1-\lambda) r^{(i+1)}(m) | m \in \mathcal{M}\}$. Thus, we get

$$\begin{aligned}
\hat{U} &= \sum_{m \in \mathcal{M}} h[\lambda r^{(i)}(m) + (1-\lambda) r^{(i+1)}(m)] \\
&\geq \sum_{m \in \mathcal{M}} \left\{ \lambda h[r^{(i)}(m)] + (1-\lambda) h[r^{(i+1)}(m)] \right\} \\
&= \lambda \tilde{f}(P_i) + (1-\lambda)\tilde{f}(P_{i+1}) \\
&= \tilde{f}[\lambda P_i + (1-\lambda) P_{i+1}] \\
&= \tilde{f}(P) \ ,
\end{aligned}$$

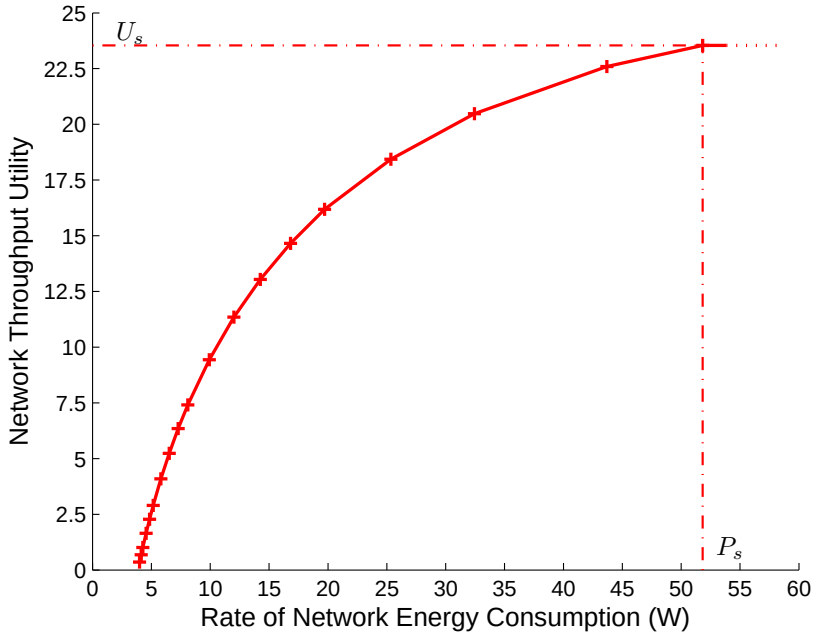


Figure 7.8: A $(1-\varepsilon)$-optimal throughput-energy curve for the nonlinear case. $\varepsilon = 1\%$.

where the second inequality holds due to the concavity of function $h(\cdot)$ and the fourth inequality holds since $\tilde{f}(P)$ is linear for $P_i \leq P \leq P_{i+1}$.

Since $\hat{\mathbf{x}}$ is a feasible solution to MOPT and $(P, f(P))$ is Pareto-optimal, we have $\hat{U} \leq f(P)$. Since $(P, \tilde{f}(P))$ is on the approximated curve with $(1-\varepsilon)$-optimal and $\tilde{f}(P) \leq \hat{U} \leq f(P)$, we can conclude that $(P, \hat{U})$ is also $(1-\varepsilon)$-optimal. □

7.6.3 A Numerical Example

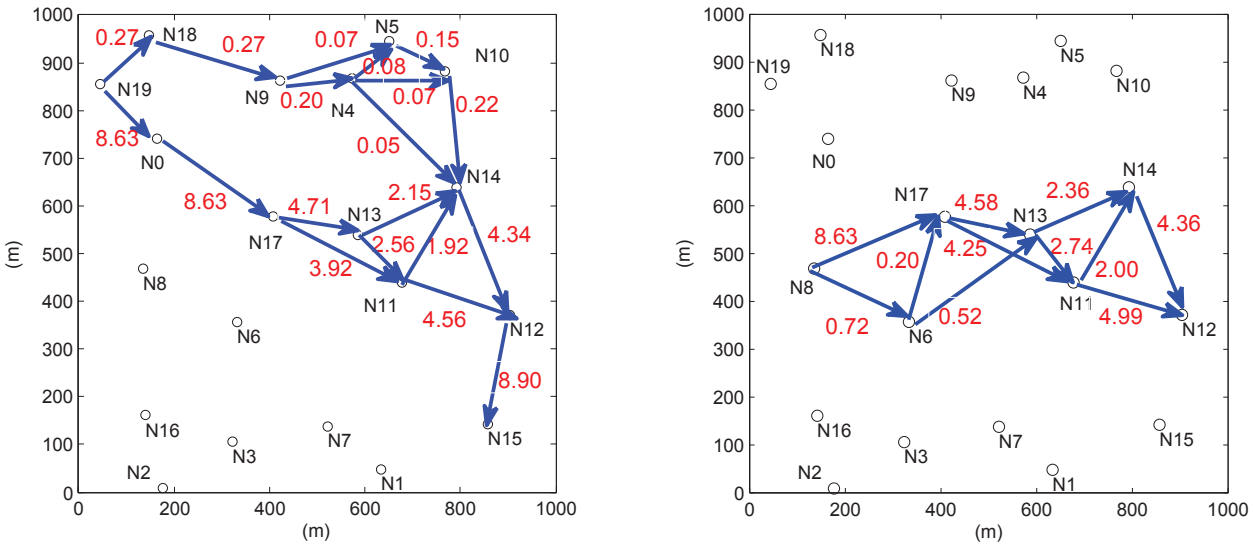
We now use a numerical example to illustrate the optimal throughput-energy curve when the throughput utility function $h[r(m)] = \ln[r(m)]$. We use the same setting as that of the numerical example shown in Section 7.5. The network topology is shown in Fig. 7.3. We first determine the saturation point (P_s, U_s) based on the approach presented in Section 7.3. On the left, we find $P(3.86) = 0$. So we choose $P_0 = 4 > 3.86$ and find its corresponding throughput util-

Table 7.4: α_l for each active link l in the example for the saturation point under nonlinear case.

Active link	α_l	Active link	α_l	Active link	α_l	Active link	α_l	Active link	α_l
$(0,8)$	1.00	$(4,14)$	1.00	$(8,16)$	0.61	$(13,6)$	0.65	$(16,8)$	1.00
$(0,9)$	1.00	$(5,9)$	0.62	$(8,17)$	1.00	$(13,11)$	0.89	$(17,6)$	0.91
$(0,17)$	1.00	$(5,10)$	0.75	$(9,0)$	1.00	$(13,14)$	1.00	$(17,9)$	1.00
$(0,18)$	0.02	$(6,3)$	1.00	$(9,4)$	0.74	$(14,4)$	0.02	$(17,11)$	1.00
$(1,15)$	1.00	$(6,7)$	1.00	$(9,5)$	1.00	$(14,10)$	1.00	$(17,13)$	0.90
$(2,3)$	0.54	$(6,8)$	0.22	$(9,17)$	1.00	$(14,11)$	1.00	$(18,0)$	0.02
$(2,16)$	0.60	$(6,13)$	1.00	$(9,18)$	0.02	$(14,12)$	1.00	$(18,9)$	0.05
$(3,1)$	0.43	$(6,17)$	1.00	$(10,14)$	0.89	$(14,13)$	0.80	$(19,0)$	0.78
$(3,6)$	1.00	$(7,1)$	0.74	$(11,12)$	1.00	$(15,12)$	0.98	$(19,18)$	0.02
$(3,7)$	0.85	$(7,6)$	1.00	$(11,14)$	0.75	$(16,2)$	0.44		
$(4,5)$	0.16	$(8,0)$	1.00	$(12,14)$	1.00	$(16,3)$	1.00		
$(4,10)$	1.00	$(8,6)$	0.64	$(12,15)$	0.12	$(16,6)$	1.00		

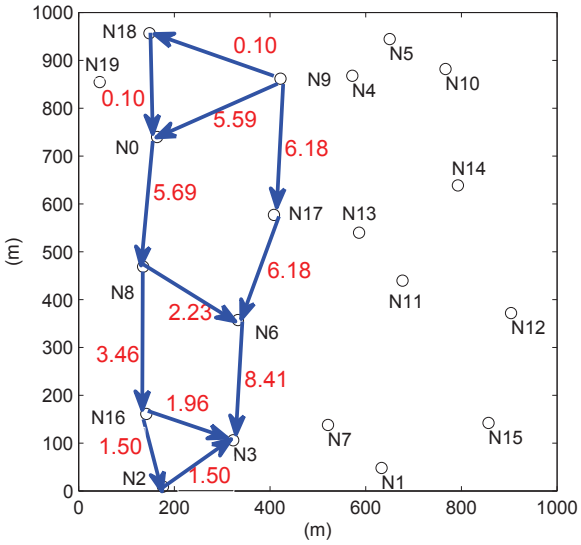
ity $f(P_0) = 0.35$. On the right, we find the saturation point $(P_s, U_s) = (51.83, 23.54)$. Now we will approximate the optimal throughput-energy curve $f(P)$ for $P \in [4.00, 51.83]$. Suppose we set the target approximation error $\varepsilon = 1\%$, i.e., we are pursuing a 99%-optimal piece-wise linear approximation. Using the method described in this section, we obtain 18 piece-wise linear segments shown in Fig. 7.8, corresponding to linear connection of 19 points on the optimal throughput-energy curve. From the figure, we can see that these points are not equally spaced along the horizontal axis. Our method dynamically adds points on the curve to meet the error bound requirement. When the curve grows rapidly at the beginning, we put more points there; when the curve slows its growth toward the end, fewer points are needed. On the other hand, if the naive approach were employed to divide the same interval $[P_0, P_s]$ into 18 equally spaced smaller intervals, the maximum error bound among all intervals would be 48%.

As an example, we show the optimal solution (including session rates, data flow rates and the fraction of time for each active link) for saturation point $(P_s, U_s) = (51.83, 23.54)$. The optimal data rates (in Mb/s) for the ten sessions are 8.90, 9.35, 11.87, 7.84, 11.27, 8.40, 12.00, 11.78, 20.89



(a) Session 1

(b) Session 2



(c) Session 3

Figure 7.9: The flow routing solutions for Sessions 1, 2, and 3 for the saturation point $(51.83, 23.54)$ in the example.

and 7.67, respectively. Note that unlike the linear case under equal weight, where there is a fairness issue, there is no session that has zero rate under the nonlinear case. This is due to our choice of $\ln(\cdot)$ as the throughput utility function. There are 57 active links in the network. In Table 7.4, we show the fraction of time for each active link. We also show the optimal flow routing solutions for the first three sessions in Fig. 7.9 but omit the rest to conserve space.

7.7 Chapter Summary

In this chapter, we explored the relationship between two key performance metrics for a multi-hop wireless network: network throughput and energy consumption. By casting the problem into a multicriteria optimization, we showed that the solution to this problem characterizes the envelope of the entire throughput-energy region. Subsequently, we presented a number of important properties associated with the optimal throughput-energy curve. As for the case study, we considered both the linear and nonlinear throughput functions. For the linear case, we were able to characterize the optimal throughput-energy curve precisely via parametric analysis. For the nonlinear case, we proposed a piece-wise linear approximation that can guarantee $(1-\varepsilon)$-optimal.

In theory, the characterization of optimal throughput-energy curve is a major step beyond the state-of-the-art research, which is limited to either maximizing throughput under some energy constraint or minimizing energy consumption while satisfying some throughput requirement (with each being able to offer only a single point on the optimal throughput-energy curve). In practice, the optimal throughput-energy curve is very useful for a network designer or operator, as it offers a holistic view on the two performance metrics. A network designer/operator can achieve his/her desired tradeoff between the two metrics depending on the specific network application scenarios.

Chapter 8

Summary and Future Work

8.1 Summary

In this dissertation, we investigated a number of optimization problems in the areas of capacity scaling laws of wireless networks and optimization of finite-sized networks. The main contributions of this dissertation are summarized as follows:

1. **Capacity Scaling Laws of Cognitive Radio Ad Hoc Networks.** We first studied the asymptotic scaling laws for CRNs, i.e., how each individual node's maximum throughput scales as the number of nodes in the network increases. Due to the heterogeneity in available frequency bands at each node, the asymptotic capacity is much more difficult to develop than prior efforts for other types of wireless networks. To overcome this difficulty, we introduced two auxiliary networks ζ and α to analyze the capacity upper and lower bounds. We derived capacity scaling law results under both the protocol and the physical models. Further, we showed that the results developed by Gupta and Kumar for the simple single-channel single-radio (SC-SR) networks are special cases under the results for CRNs.

2. **Asymptotic Capacity of Multi-hop MIMO Ad Hoc Networks.** MIMO is a key technology to increase the capacity of wireless networks. Although there had been extensive

work on MIMO at the physical and link layers, there was limited work on MIMO at the network layer (i.e., multi-hop MIMO network), particularly results on capacity scaling laws. In this work, we investigated capacity scaling laws for MIMO ad hoc networks. Our goal was to find the achievable throughput of each node as the number of nodes in the network increases. We employed a MIMO network model that captures spatial multiplexing and interference cancellation. We showed that for a MIMO network with n randomly located nodes, each equipped with γ antennas and a data rate of W on each data stream, the achievable throughput of each node is $\Theta(\frac{\gamma W}{\sqrt{n \ln n}})$.

3. **A Unified and Simple Method for Establishing Capacity Scaling Laws.** Since the seminal work of Gupta and Kumar, there have been tremendous efforts developing capacity scaling laws for ad hoc networks with various advanced physical layer technologies. These efforts led to many custom-designed approaches, most of which were intellectually challenging and lacked universal properties that can be extended to address scaling laws of ad hoc networks with other physical layer technologies. In this work, we presented a set of simple yet general criteria that can be applied to quickly determine the capacity scaling laws for various physical layer technologies under the protocol model. We proved the correctness of our proposed criteria and validated them through a number of case studies, such as ad hoc networks with directional antenna, MIMO, cognitive radio, multi-channel and multi-radio, and multiple packet reception. These simple criteria will serve as powerful tools to networking researchers to obtain throughput scaling laws of ad hoc networks under different physical layer technologies, particularly those to be developed in the future.

4. **Exploiting SIC for Multi-Hop Wireless Networks.** There is a growing interest in exploiting interference to increase network throughput. In particular, the so-called *successive interference cancellation* (SIC) scheme appeared very promising, due to its ability to enable concurrent receptions from multiple transmitters and interference rejection. However, due to some stringent constraints and limit, SIC alone is inadequate to handle all concurrent interference. We advocated a joint interference exploitation and avoidance approach, which

combines the best of interference exploitation and interference avoidance, while avoiding each's pitfalls. We discussed the new challenges of such a new approach in a multi-hop wireless network and propose a formal optimization framework, with cross-layer formulation of physical, link, and network layers. This framework offered a rather complete design space for SIC to squeeze the most out of interference. To demonstrate the practical utility of our model, we conducted a case study. Our numerical results affirmed the validity of our framework and gave insights on how SIC can optimally interact with an interference avoidance scheme.

5. **Throughput Optimization with Network-wide Energy Constraint.** Conserving network-wide energy consumption is becoming an increasingly important concern for network operators. In this work, we studied network-wide energy conservation problem which we hope will offer insights to both network operators and users. Specifically, we studied how to maximize network throughput under a network-wide energy constraint for a general multi-hop wireless network. We formulated this problem as a mixed-integer nonlinear program (MINLP). We proposed a novel piece-wise linear approximation to transform the nonlinear constraints into linear constraints. We proved that the solution developed under this approach is near-optimal with guaranteed performance bound.

6. **Bicriteria Optimization in Multi-Hop Wireless Networks.** Network throughput and energy consumption are two important performance metrics for a multi-hop wireless network. Current state-of-the-art is limited to either maximizing throughput under some energy constraint or minimizing energy consumption while satisfying some throughput requirement. However, the important problem of how to optimize both objectives simultaneously remains open. In this work, we took a multicriteria optimization approach to conduct a systematic study on the relationship between the two performance objectives. We showed that the solution to the multicriteria optimization problem characterizes the envelope of the entire throughput-energy region, i.e., the so-called optimal throughput-energy curve. We proved some important properties of the optimal throughput-energy curve. For case study, we con-

sidered both linear and nonlinear throughput functions. For the linear case, we characterized the optimal throughput-energy curve precisely through parametric analysis, while for the nonlinear case, we used a piece-wise linear approximation to approximate the optimal throughput-energy curve with arbitrary accuracy. Our results offered important insights on exploiting the trade-off between the two performance metrics.

8.2 Future Work

In this dissertation, we explored a number of network optimization problems, both in asymptotic and finite sizes. During the course of our investigation, we found that there remain many problems to be explored. Here, we list a few important and open problems in Chapter 4, 5, and 7 that deserve future research.

- **Open Problems in Unified Approach to Determine Asymptotic Capacity Scaling Laws.** Chapter 4 provides some simple criteria for establishing asymptotic capacity upper bounds for ad hoc networks with different physical layer technologies. However, it remains an open problem whether or not it is possible to develop a unified approach to determine asymptotic capacity lower bounds. The main difficulty in deriving a capacity lower bound for a specific physical layer technology is to find a *feasible* solution, which includes resource allocation at the physical layer, scheduling at the MAC layer, and routing at the network layer. Finding a solution to variables that is feasible at all these layers is much harder than just developing inequality relationships that are used to derive asymptotic upper bounds. It will be interesting to see how a unified approach can be developed to quickly determine the asymptotic capacity lower bounds for various physical layer technologies.

  Also, the results developed in Chapter 4 are under the protocol model. It is well recognized that the physical model offers a much better characterization of physical layer behavior. But developing a unified approach under the physical model is still an open problem. Our success under the protocol model lies in the fact that accounting of interference relationship among

such a model is relatively simple — the successful transmission of a link is solely dependent upon the locations of its nearby links and the determination of interference relationship is binary (based on interference range). In contrast, under the physical model, the successful transmission of a link is based on a SINR threshold. This involves complex calculation of SINR at a receiver, which would require taking into account of interference from many nodes in the network than those under the protocol model.

- **Exploiting SIC with Power Control and Rate Adaption.** In Chapter 5, we showed that SIC is a promising physical layer technology to improve the network throughput in a multi-hop wireless network. Our work was done under the simple assumption that the power level at each node is fixed. If we allow tunable power level at each node, then the design space will be much greater since more simultaneous receptions and interference rejection may be allowed in the network. However, allowing power control also brings many difficulties in the modeling of SIC behavior in a multi-hop network. This is because that in SIC, we need to determine the decoding order of the received signals at a receiver based on their received power levels. When power control is allowed at each node, the received power levels will not only be determined by the channel gain, but also the variable transmit power levels. This makes the modeling of sequential decoding process very difficult.

  Our work in Chapter 5 also assumed that once the SINR of a link exceeds a threshold, we consider that the data transmission on this link is successful and the data rate on this link is fixed as a constant regardless of the SINR value. Another approach is to allow rate adaption on a link, i.e., adjust the data rate on a link based on its SINR. However, allowing rate adaption will make the problem much harder. Rate adaption will involve the log function (for Shannon capacity computation), which introduces nonlinear constraints into our optimization framework. This will change our problem into a Mixed Integer Non-Linear Program (MINLP), which is much more difficult to solve than a MILP problem.

- **Optimal Throughput-Energy Curve under Co-channel Interference Model.** In Chapter 7, we simplified link layer scheduling by employing orthogonal channels to handle in-

terference. A general problem with consideration of co-channel interference will make the problem more interesting and challenging. Incorporating interference management either in time, frequency, or code domains into problem formulation will add much complexity. For example, if we handle interference in time domain, we need to answer the following questions: How many time slots should be divided in each time frame? How long should be a time frame? Should each time slot be of the same size? How to reuse each time slot among different nodes or links in the network? And if we employ some random access mechanism to handel interference, we need to answer another set of questions: When a transmitter is allowed to access the medium? How long should a transmitter wait when it detects packet collisions? How to solve the classic hidden terminal problem in these random access mechanisms? It will be interesting to see how to characterize the optimal throughput-energy curve under those more complex co-channel interference models.

Bibliography

[1] M. Al-Ayyoub and H. Gupta, "Joint routing, channel assignment, and scheduling for throughput maximization in general interference models," *IEEE Trans. on Mobile Computing*, vol. 9, no. 4, pp. 553–565, April 2010.

[2] M. Alicherry, R. Bhatia, and L. Li, "Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks," in *Proc. ACM MobiCom*, pp. 58–72, Cologne, Germany, Aug. 2005.

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

[4] M. Andrews and M. Dinitz, "Maximizing capacity in arbitrary wireless networks in the SINR model: Complexity and game theory," in *Proc. IEEE INFOCOM*, pp. 1332–1340, Rio de Janeiro, Brazil, April 19–25, 2009.

[5] 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.

[6] M.S. Bazaraa, J.J. Jarvis, and H.D. Sherali, *Linear Programming and Network Flows, 3rd edition,* John Wiley & Sons Inc., Hoboken, New Jersey, 2005.

[7] M.S. Bazaraa, H.D. Sherali, and C.M. Shetty, *Nonlinear Programming: Theory and Algorithms, third edition,* John Wiley & Sons Inc., Hoboken, New Jersey, 2006.

[8] V. Bhandari and N.H. Vaidya, "Capacity of multi-channel wireless networks with random (c, f) assignment," in *Proc. ACM MobiHoc*, pp. 229–238, Montreal, Quebec, Canada, Sep. 9–14, 2007.

[9] R. Bhatia and M. Kodialam, "On power efficient communication over multi-hop wireless networks: Joint routing, scheduling and power control," in *Proc. IEEE INFOCOM*, pp. 1457–1466, Hong Kong, China, March 2004.

[10] J. Blomer and N. Jindal, "Transmission capacity of wireless ad hoc networks: Successive interference cancellation vs. joint detection," in *Proc. IEEE ICC*, 5 pages, Dresden, Germany, June 14–18, 2009.

[11] H. Bolcskei, R.U. Nabar, O. Oyman, and A.J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Transactions on Wireless Communications*, vol. 5, no. 6, pp. 1433–1444, June 2006.

[12] S. Chachulski, M. Jennings, S. Katti and D. Katabi, "Trading structure for randomness in wireless opportunistic routing," in *Proc. ACM SIGCOMM*, pp. 169–180, Kyoto, Japan, Aug. 2007.

[13] D. Chafekar, V.S.A. Kumar, M. Marathe, S. Parthasarathy, and A. Srinivasan, "Approximation algorithms for computing capacity of wireless networks with SINR constraints," in *Proc. IEEE INFOCOM*, pp. 1166–1174, Phoenix, AZ, April 13–18, 2008.

[14] P. Chaporkar, K. Kar, and S. Sarkar, "Throughput guarantees through maximal scheduling in wireless networks," in *Proc. 43rd Ann. Allerton Conf. Commun., Control and Comput.*, pp. 1557–1567, Monticello, IL, Sep. 2005.

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

[16] L. Chen, S.H. Low, M. Chiang, and J.C. Doyle, “Cross-layer congestion control, routing and scheduling design in ad hoc wireless networks,” in *Proc. IEEE INFOCOM,* pp. 1–13, Barcelona, Spain, April 2006.

[17] M. Chiang, “Balancing transport and physical layers in wireless multihop networks: Jointly optimal congestion control and power control,” *IEEE Journal on Selected Areas in Communications,* vol. 23, no. 1, pp. 1166–1174, Jan. 2005.

[18] M. Chiani, M.Z. Win, and H. Shin, “MIMO networks: The effect of interference,” *IEEE Transactions on Information Theory,* vol. 56, no. 1, Jan. 2009.

[19] L.-U. Choi and R.D. Murch, “A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach,” *IEEE Transactions on Wireless Communications,* vol. 3, no. 1, pp. 20-24, Jan. 2004.

[20] T.J. Cormen, C.E. Leiserson, R.L. Rivest, and C. Stein, *Introduction to Algorithms, Second Edition,* The MIT Press, 2001.

[21] The First International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies, Venice, Italy, May 22–27, 2011. http://www.iaria.org/conferences2011/ENERGY11.html.

[22] D.S.J. De Couto, D. Aguayo, J. Bicket, and R. Morris, “A high-throughput path metric for multi-hop wireless routing,” *Springer Wireless Networks,* vol. 11, no. 4, pp. 419–434, July 2005.

[23] IBM ILOG CPLEX Optimizer, http://www-01.ibm.com /software/integration/optimization/cplex-optimizer/.

[24] R.L. Cruz and A.V. Santhanam, “Optimal routing, link scheduling, and power control in multi-hop wireless networks,” in *Proc. IEEE INFOCOM,* pp. 702–711, San Francisco, CA, March 30–April 3, 2003.

[25] S. Cui, A.J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Transactions on Wireless Communications,* vol. 4, no. 5, pp. 2349–2360, Sep. 2005.

[26] H. Dai, K. Ng, R.C. Wong, and M. Wu, "On the capacity of multi-channel wireless networks using directional antennas" in *Proc. IEEE INFOCOM,* pp. 1301-1309, Phoenix, AZ, Apr. 13-18, 2008.

[27] D. Dardari, A. Conti, C. Buratti, and R. Verdone, "Mathematical evaluation of environmental monitoring estimation error through energy-efficient wireless sensor networks," *IEEE Transactions on Mobile Computing,* vol. 6, no. 7, pp. 790–802, July 2007.

[28] Q. Dong, S. Banerjee, M. Adler, and A. Misra, "Minimum energy reliable paths using unreliable wireless links," in *Proc. ACM MobiHoc,* pp. 449–459, Urbana-Champaign, IL, May 2005.

[29] M. Ehrgott, *Multicriteria Optimization, Second edition* Springer-Verlag, New York, 2010. (ISBN-13: 978–3–642–05975–9).

[30] Minimizing carbon intensity in telecom networks using TCO techniques. Ericsson white paper, Feb. 2010. http://www.ericsson.com/res/docs/whitepapers/TCO2_0211high.pdf

[31] G. Fettweis and E. Zimmermann, "ICT energy consumption – trends and challenges," in *Proc. of the 11th International Symposium on Wireless Personal Multimedia Communications* (WPMC 2008), Lapland, Finland, Sep. 8–11, 2008.

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

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

[34] M. Franceschetti, O. Dousse, and D.N.C. Tse, "Closing the gap in the capacity of wireless networks via percolation theory," *IEEE Transaction on Information Theory,* vol. 53, no. 3, pp. 1009–1018, March 2007.

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

[36] A. El Gamal, C. Nair, B. Prabhakar, E. Uysal-Biyikoglu, and S. Zahedi, "Energy-efficient scheduling of packet transmissions over wireless networks," in *Proc. IEEE INFOCOM*, pp. 1773–1782, New York, NY, June 2002.

[37] A.E. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-delay trade-off in wireless networks," in *Proc. IEEE INFOCOM,* pp. 464–475, Hong Kong, China, March 7–11, 2004.

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

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

[40] E. Gelal, K. Pelechrinis, T.S. Kim, I. Broustis, S.V. Krishnamurthy, and B. Rao, "Topology control for effective interference cancellation in multi-user MIMO networks," in *Proc. IEEE INFOCOM,* 9 pages, San Diego, CA, March 15–19, 2010.

[41] A.M. Geoffrion, "A generalized Benders' decomposition," *Journal of optimization theory and applications,* vol. 10, no. 4, pp. 237–260, 1972.

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

[43] S. Gollakota and D Katabi, "Zigzag decoding: Combating hidden terminals in wireless networks," in *Proc. ACM SIGCOMM,* pp. 159–170, Seattle, WA, Aug. 17–22, 2008.

[44] O. Goussevskaia and R. Wattenhofer, "Capacity of arbitrary wireless networks," in *Proc. IEEE INFOCOM,* pp. 1872–1880, Rio de Janeiro, Brazil, April 19–25, 2009.

[45] 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.

[46] P. Gupta and P.R. Kumar, "Critical power for asymptotic connectivity in wireless networks," in *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming,* W.M. McEneany, G. Yin, and Q. Zhang, Eds. Boston, MA: Birkhauser, pp. 547–566, 1998.

[47] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Transaction on Information Theory,* vol. 46, no. 2, pp. 388–404, March 2000.

[48] A. Keshavarz-Haddad, V. Ribeiro, and R. Riedi, "Broadcast capacity in multihop wireless networks," in *Proc. ACM MobiCom,* pp. 239–250, Los Angeles, CA, Sep. 23–26, 2006.

[49] 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. 14–19, 2008.

[50] W.R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. of the 33rd Hawaii International Conference on System Sciences,* pp. 3005–3014, Maui, Hawaii, Jan. 2000.

[51] S. Huang, X. Liu, and Z. Ding, "Distributed power control for cognitive user access based on primary link control feedback," in *Proc. IEEE INFOCOM,* pp. 1280–1288, San Diego, CA, March 14–19, 2010.

[52] Y.T. Hou, Y. Shi, and H.D. Sherali, "Optimal base station selection for anycast routing in wireless sensor networks," *IEEE Trans. Vehic. Tech.*, vol. 55, issue 3, pp. 813–821, May 2006.

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

[54] W. Huang and X. Wang, "Throughput and delay scaling of general cognitive networks," in *Proc. IEEE INFOCOM*, Shanghai, China, Apr. 2011.

[55] S.-W. Jeon, N. Devroye, M. Vu, S.-Y. Chung, and V. Tarokh, "Cognitive networks achieve throughput scaling of a homogeneous network," 5 pages, 7th Intl. Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), Seoul, Korea, June 2009.

[56] C. Jiang, Y. Shi, Y.T. Hou, and S. Kompella, "On the asymptotic capacity of multi-hop MIMO ad hoc networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 4, pp. 1032–1037, Apr. 2011.

[57] C. Jiang, Y. Shi, Y.T. Hou, and S. Kompella, "On optimal throughput-energy curve for multi-hop wireless networks," in *Proc. IEEE INFOCOM*, pp. 1341–1349, Shanghai, China, April 10–15, 2011.

[58] C. Jiang, Y. Shi, Y.T. Hou, W. Lou, S. Kompella, and S.F. Midkiff, "Toward simple criteria to establish capacity scaling laws for wireless networks," in *Proc. IEEE INFOCOM*, Orlando, FL, March 25–30, 2012.

[59] 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*, Orlando, FL, March 25–30, 2012.

[60] 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. Vehic. Tech.*, vol. 46, no. 4, pp. 862–870, Apr. 1997.

[61] M. Kang and M-S. Alouini, "Capacity of MIMO Rician channels," *IEEE Transactions on Wireless Communications*, vol. 5, no. 1, pp. 112–122, Jan. 2006.

[62] S. Katti, S. Gollakota, and D Katabi, "Embracing wireless interference: Analog network coding," in *Proc. ACM SIGCOMM*, pp. 397–408, Kyoto, Japan, Aug. 27–31, 2007.

[63] F.P. Kelly, A. Maulloo, and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, no. 3, pp. 237–252, March 1998.

[64] Y. Kim, H. Shin, and H. Cha, "Y-MAC: An energy-efficient multi-channel MAC protocol for dense wireless sensor networks," in *Proc. of the 7th International Conference on Information Processing in Sensor Networks*, pp. 53–63, St. Louis, Missouri, USA, April 2008.

[65] M. Kodialam and T. Nandagopal, "Characterizing achievable rates in multi-hop wireless mesh networks with orthogonal channels," *IEEE/ACM Transactions on Networking*, vol. 13, no. 4, pp. 868–880, August 2005.

[66] S.R. Kulkarni and P. Viswanath, "A deterministic approach to throughput scaling in wireless networks," *IEEE Transaction on Information Theory*, vol. 50, no. 6, pp. 1041–1049, June 2004.

[67] S. Kwon and N.B. Shroff, "Unified energy-efficient routing for multi-hop wireless networks," in *Proc. IEEE INFOCOM*, pp. 430–438, Phoenix, AZ, April 2008.

[68] P. Kyasanur and N.H. Vaidya, "Capacity of multi-channel wireless networks: impact of number of channels and interfaces," in *Proc. ACM MobiCom*, pp. 43–57, Cologne, Germany, Aug. 28–Sep. 2, 2005.

[69] M. Kodialam and T. Nandagopal, "Characterizing the capacity region in multi-radio multi-channel wireless mesh networks," in *Proc. ACM MobiCom,* pp. 73–87, Cologne, Germany, Aug. 28–Sep. 2, 2005.

[70] W. Li and H. Dai, "Optimal throughput and energy efficiency for wireless sensor networks: Multiple access and multipacket reception," *EURASIP Journal on Wireless Communications and Networking,* vol. 5, issue 4, pp. 541-553, Sep. 2005.

[71] G.Y. Li, S. Xu, A. Swami, N. Himayat, and G. Fettweis, (Guest Editors), *IEEE Journal on Selected Areas in Commun. – Special Issue on Energy-Efficient Wireless Communications,* forthcoming.

[72] X. Lin and N.B. Shroff, "The impact of imperfect scheduling on cross-layer congestion control in wireless networks," *IEEE/ACM Trans. on Networking,* vol. 14, no. 2, pp. 302–315, April 2006.

[73] X. Lin and S. Rasool, "Distributed and provably efficient algorithms for joint channel-assignment, scheduling, and routing in multichannel ad hoc wireless networks," *IEEE/ACM Trans. on Networking,* vol. 17, no. 6, pp. 1874–1887, Dec. 2009.

[74] L. Lin, X. Lin, and N.B. Shroff, "Low-complexity and distributed energy minimization in multihop wireless networks," *IEEE/ACM Transactions on Networking,* vol. 18, no. 2, pp. 501–514, April 2010.

[75] J. Liu, D. Goeckel, and D. Towsley, "Bounds on the gain of network coding and broadcasting in wireless networks," in *Proc. IEEE INFOCOM,* pp. 724–732, Anchorage, AK, May 6–12, 2007.

[76] 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. 20–24, 2010.

[77] G. Lu, N. Sadagopan, B. Krishnamachari, and A. Goel, "Delay efficient sleep scheduling in wireless sensor networks," in *Proc. of IEEE INFOCOM*, pp. 2470–2481, Miami, FL, March 2005.

[78] S. Lv, X. Wang, and X. Zhou, "Scheduling under SINR model in ad hoc networks with successive interference cancellation," in *Proc. IEEE GLOBECOM*, 5 pages, Miami, FL, Dec. 6–10, 2010.

[79] 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.

[80] 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.

[81] X. Mao, X. Li, and S. Tang, "Multicast capacity for hybrid wireless networks," in *Proc. ACM MobiHoc*, pp.189–198, Hong Kong, China, May 26–30, 2008.

[82] I. Maric and R.D. Yates, "Cooperative multihop broadcast for wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, issue 6, pp. 1080–1088, Aug. 2004.

[83] G.W. Miao, N. Himayat, G.Y. Li, and A. Swami, "Cross-layer optimization for energy-efficient wireless communications: A survey," *Wireless Commun. and Mobile Computing (Wiley)*, vol. 9, no. 4, pp. 529–542, Apr. 2009.

[84] G.W. Miao, N. Himayat, and G.Y. Li "Energy-efficient link adaptation in frequency-selective channels," *IEEE Transactions on Communications*, vol. 58, no. 2, pp. 545–554, Feb. 2010.

[85] S.M. Mishra, A. Sahai, and R.W. Brodersen, "Cooperative sensing among cognitive radios," in *Proc. IEEE International Conference on Communications*, pp. 1658–1663, Istanbul, Turkey, June 11–15, 2006.

[86] R. Motwani and P. Raghavan, *Randomized Algorithms,* Cambridge University Press, 1995.

[87] H. Nama, M. Chiang, and N. Mandayam, "Utility-lifetime trade-off in self-regulating wireless sensor networks: A cross-layer design approach," in *Proc. IEEE ICC,* pp. 3511–3516, Istanbul, Turkey, June 2006.

[88] M.J. Neely, "Energy optimal control for time varying wireless networks," *IEEE Trans. on Information Theory,* vol. 52, no. 7, pp. 2915–2934, July 2006.

[89] M.J. Neely and R. Urgaonkar, "Optimal backpressure routing in wireless networks with multi-receiver diversity," *Ad Hoc Networks (Elsevier),* vol. 7, no. 5, pp. 862–881, July 2009.

[90] R. Negi and A. Rajeswaran, "Capacity of power constrained ad hoc networks," in *Proc. IEEE INFOCOM,* pp. 443–453, Hong Kong, China, March 2004.

[91] A. Özgür, O. Lévêque and D.N.C. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Transaction on Information Theory,* vol. 53, no. 10, pp. 3549–3572, Oct. 2007.

[92] C. Peraki and S.D. Servetto, "On the maximum stable throughput problem in random networks with directional antennas," in *Proc. ACM MobiHoc,* pp. 76–87, Annapolis, Maryland, June 1–3, 2003.

[93] M. Ramsay, "Green networks keep growing in emerging markets," *Wireless Week Magazine,* September 25, 2010. http://www.wirelessweek.com/Articles/2010/09/Green-Networks-Emerging-Markets/.

[94] S. Rosloniec, *Fundamental Numerical Methods for Electrical Engineering,* Springer, Berlin, 2008. (ISBN-13: 978–3–540–79518-6).

[95] H.R. Sadjadpour, Z. Wang, and J.J. Garcia-Luna-Aceves, "The capacity of wireless ad hoc networks with multi-packet reception," *IEEE Transactions on Communications,* vol. 58, no. 2, pp. 600–610, Feb. 2010.

[96] S. Sambhwani, W. Zhang, and W. Zeng, “Uplink interference cancelation in HSPA: Principles and practice,” QUALCOMM White Paper, 2008.

[97] 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. 20–24, 2010.

[98] A. Schrijver, *Theory of Linear and Integer Programming,* Wiley-Interscience, New York, NY, 1986.

[99] S. Sen, N. Santhapuri, R.R. Choudhury, and S. Nelakuditi, “Successive interference cancellation: A back-of-the-envelope perspective,” in *Proc. Ninth ACM Workshop on Hot Topics in Networks (HotNets-IX),* Monterey, CA, Oct. 20–21, 2010.

[100] S. Shakkottai, X. Liu, and R. Srikant, “The multicast capacity of large multihop wireless networks,” in *Proc. ACM MobiHoc,* pp. 247–255, Montreal, Quebec, Canada, Sep. 9–14, 2007.

[101] S. Sharma, Y. Shi, Y.T. Hou, H.D. Sherali, S. Kompella, and S.F. Midkiff, “Joint flow routing and relay node assignment in cooperative multi-hop networks,” *IEEE Journal on Selected Areas in Commun.*, vol. 30, issue 2, pp. 254–262, February 2012.

[102] Y. Shi and Y.T. Hou, “Optimal power control for multi-hop software defined radio networks,” in *Proc. IEEE INFOCOM,* pp. 1694–1702, Anchorage, AL, May 6–12, 2007.

[103] Y. Shi, Y.T. Hou, S. Kompella, and H.D. Sherali, “Maximizing capacity in multi-hop cognitive radio networks under the SINR model,” *IEEE Transactions on Mobile Computing,* vol. 10, no. 7, pp. 954-967, July 2011.

[104] Y. Shi, J. Liu, C. Jiang, C. Gao, and Y.T. Hou, “An optimal link layer model for multi-hop MIMO networks,” in *Proc. IEEE INFOCOM,* Shanghai, China, April 2011.

[105] Y. Shi, C. Jiang, Y.T. Hou, and S. Kompella, “On capacity scaling law of cognitive radio ad hoc networks,” in *Proc. IEEE ICCCN,* 8 pages, Maui, Hawaii, July 31–Aug. 4, 2011.

[106] S. Singh, M. Woo, and C.S. Raghavendra, “Power-aware routing in mobile ad hoc networks,” in *Proc. ACM MobiCom*, pp. 181–190, Dallas, Texas, Oct. 1998.

[107] Q.H. Spencer, A.L. Swindlehurst, and M. Haardt, “Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels,” *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 388–404, Feb. 2004.

[108] 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. 259–268, New Orleans, LA, May 18–21, 2009.

[109] Y. Sun, S. Du, O. Gurewitz, and D.B. Johnson, “DW-MAC: A low latency, energy efficient demand-wakeup MAC protocol for wireless sensor networks,” in *Proc. ACM MobiHoc*, pp. 53–62, Hong Kong, China, May 2008.

[110] J. Tang, G. Xue, C. Chandler, and W. Zhang, “Link scheduling with power control for throughput enhancement in multihop wireless networks,” *IEEE Trans. on Vehicular Technology*, vol. 55, no. 3, pp. 733–742, May 2006.

[111] A. Tarello, J. Sun, M. Zafer, and E. Modiano, “Minimum energy transmission scheduling subject to deadline constraints,” *Springer Wireless Networks*, vol. 14, no. 5, pp. 633-645, 2008.

[112] L. Tassiulas and A. Ephremides, “Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks,” *IEEE Trans. on Automatic Control*, vol. 37, no. 12, pp. 1936–1948, Dec. 1992.

[113] I.E. Telatar, “Capacity of multi-antenna Gaussian channels,” *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[114] L. Tong, Q. Zhao and G. Mergen, “Multipacket reception in random access wireless networks: From signal processing to optimal medium access control,” *IEEE Communications Magazine*, vol. 39, no. 11, pp. 108–112, Nov. 2001.

[115] D.N.C. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Chapter 6, Cambridge Univ. Press, 2005.

[116] R. Urgaonkar and M.J. Neely, "Capacity region, minimum energy, and delay for a mobile ad-hoc network," in *Proc. WiOpt*, 10 pages, Boston, MA, April 2006.

[117] T. Van Dam and K. Langendoen, "An adaptive energy-efficient MAC protocol for wireless sensor networks," in *Proc. ACM SenSys*, pp. 171–180, Los Angeles, CA, Nov. 2003.

[118] M.K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code-division multiple access communications," *IEEE Trans. Commun.*, vol. 38, no. 4, pp. 509–519, Apr. 1990.

[119] S. Verdu, *Multiuser Detection*, Cambridge Univ. Press, 1998.

[120] A.J. Viterbi, "Very low rate convolutional codes for maximum theoretical performance of spread-spectrum multiple-access channel," *IEEE J. Sel. Areas Commun.*, vol. 8, no. 4, pp. 641–649, May 1990.

[121] M. Vu, N. Devroye, M. Sharif, and V. Tarokh, "Scaling laws of cognitive networks," International Conference on Cognitive Radio Oriented Wireless Networks and Communications (Crowncom), pp. 2–8, Orlando, FL, July 31–Aug. 3, 2007.

[122] X. Wang and H.V. Poor, "Iterative (Turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046–1061, July 1999.

[123] Z. Wang, H.R. Sajadpour, and J.J. Garcia-Luna-Aceves, "The capacity and energy efficiency of wireless ad hoc networks with multi-packet reception," in *Proc. ACM MobiHoc*, pp. 179–188, Hong Kong, China, May 26–30, 2008.

[124] S. Weber, J.G. Andrews, X. Yang, and G. de Veciana, "Transmission capacity of wireless ad hoc networks with successive interference cancellation," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2799–2814, Aug. 2007.

[125] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 388–404, Sep. 2006.

[126] D.B. West, *Introduction to Graph Theory,* Prentice Hall, Upper Saddle River, NJ, 2001.

[127] 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.

[128] J.E. Wieselthier, G.D. Nguyen, and A. Ephremides, "Energy-aware wireless networking with directional antennas: The case of session-based broadcasting and multicasting," *IEEE Trans. on Mobile Computing*, vol. 1, no. 3, pp. 176–192, July–Sept. 2002.

[129] The Green Communications Workshop, Oct. 14–15, 2010, University of Maryland, College Park, MD. http://www.greencomm.umd.edu.

[130] The Third International Workshop on Green Communications, Miami, FL, December 10, 2010. http://www.green-communications.net/globecom10/home.html.

[131] X. Wu and R. Srikant, "Regulated maximal matching: A distributed scheduling algorithm for multi-hop wireless networks with node-exclusive spectrum sharing," in *Proc. IEEE CDC*, pp. 5342–5347, Seville, Spain, Dec. 2005.

[132] A.M. Wyglinski, M. Nekovee, and Y.T. Hou, *Cognitive Radio Communications and Networks: Principles and Practices,* Academic Press/Elsevier, 2010.

[133] W. Ye, J. Heidemann, and D. Estrin, "An energy-efficient MAC protocol for wireless sensor networks," in *Proc. IEEE INFOCOM*, pp. 1567–1576, New York, NY, June 2002.

[134] S. Yi, Y. Pei, and S. Kalyanaraman, "On the capacity improvement of ad hoc wireless networks using directional antennas," in *Proc. ACM MobiHoc*, pp. 108–116, Annapolis, MD, June 1–3, 2003.

[135] C. Yin, L. Gao, and S. Cui, “Scaling laws of overlaid wireless networks: a cognitive radio network vs. a primary network”, in *Proc. IEEE GLOBECOM*, pp. 1235–1239, New Orleans, LA, Nov. 30–Dec. 4, 2008.

[136] C. Yin, L. Gao, and S. Cui, “Scaling laws of overlaid wireless networks: A cognitive radio network vs. a primary network,” *IEEE/ACM Transactions on Networking*, vol. 18, no. 4, pp. 1317–1329, Aug. 2010.

[137] T. Yoo and A. Goldsmith, “On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming,” *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, March 2006.

[138] H. Zhang and J.C. Hou, “Capacity of wireless ad-hoc networks under ultra wide band with power constraint,” in *Proc. IEEE INFOCOM*, pp. 455–465, Miami, FL, March 13–17, 2005.

[139] L. Zheng and D.N.C. Tse, “Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels,” *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[140] R. Zheng, “Information dissemination in power-constrained wireless network,” in *Proc. IEEE INFOCOM*, Barcelona, Catalunya, Spain, April 23–29, 2006.