Cross-Layer Optimization and Distributed Algorithm Design for Frequency-Agile Radio Networks

Zhenhua Feng

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

Yaling Yang, Chair
Jeffrey H. Reed
Y. Thomas Hou
Allen B. MacKenzie
Douglas R. Bish

November 19, 2010
Blacksburg, Virginia

Keywords: Wireless Networks, Dynamic Spectrum Access, Frequency-Agile Radio,
Cross-Layer Design, Distributed Design

Cross-Layer Optimization and Distributed Algorithm Design for Frequency-Agile Radio Networks

Zhenhua Feng

(ABSTRACT)

Recent advancements in frequency-agile radio technology and dynamic spectrum access network have created a huge space for improving the utilization efficiency of wireless spectrum. Existing algorithms and protocols, however, have not taken full advantage of the new technologies due to obsolete network design idealogies inherited from conventional network design, such as static spectrum access and static channelization. In this dissertation, we propose new resource management models and algorithms that capitalize on the frequency-agility of next generation radios and the dynamic spectrum access concepts to increase the utilization efficiency of wireless spectrum.

We first propose a new analytical model for Dynamic Spectrum Access (DSA) networks. Compared to previous models, the new model is able to include essential DSA mechanisms such as spectrum sensing and primary interference avoidance into solid mathematical representation and thus drastically increase the accuracy of our model. The subsequent numerical study conforms well with existing empirical studies and provides fundamental insights on the design of future DSA networks.

We then take advantage of partially overlapped channels (POC) in frequency-agile radio networks and propose a simple joint channel scheduling and flow routing optimization algorithm that maximizes aggregate network throughput. The proposed model quantifies the impact of fundamental network settings, such as node density and traffic load, on the performance of partially overlapped channel based networks.

We then propose a cross-layer radio resource allocation algorithm JSSRC (Joint Spectrum

Sharing and end-to-end data Rate Control) that iteratively adapts a frequency-agile radio network to optimum with regard to aggregate spectrum utilization. Subsequently, we extend JSSRC to include routing and present TRSS (joint Transport, Routing and Spectrum Sharing) to solve the much more complex joint transport, routing and spectrum sharing optimization problem. Both JSSRC and TRSS enjoy theoretical convergence and achieve optimum with appropriate scheduling algorithms.

The work in this dissertation together strives to improve the efficiency of spectrum utilization in frequency-agile radio networks. Numerical and simulation studies show the effectiveness of our approaches to reduce the so-called spectrum shortage problem.

Dedication

I dedicate this dissertation to my beloved parents Jiming Feng and Guixiang Wang for their unconditional love and support throughout my life.

Acknowledgments

First and foremost, I would like to thank my advisor, Dr. Yaling Yang, for her guidance, help and commitment throughout my study here at Virginia Tech. Dr. Yang's dedication to excellence in research has always been the inspiration for me to pursue higher quality of research. When not in her advisor mode, Dr. Yang has always been a great colleague and a friend who is willing to listen and ready to help. I would also like to thank all my committee members Dr. Reed, Dr. MacKenzie, Dr. Hou and Dr. Bish for their sharp advices and generous help throughout the entire period of my Ph.D study and the composition of this dissertation.

I would also like to extend my thanks to all my friends in and out of Virginia Tech who have enriched me and made my life colorful in all seasons. Many thanks also goes to my colleagues in the SHINE lab and the CESCA research center for their companionship and inspirational discussions.

Finally, special thanks to my beloved parents and my extended family who have always been there for me. Without them, this work would have not been possible. To them, I dedicate this dissertation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Spectrum Shortage Problem

Thanks to the proliferation of wireless voice/data services, we are enjoying unprecedented convenience in personal communication and networking. Once only possible in James Bond movies, business/personal usage of mobile device for data/video/voice is now almost ubiquitous today. A person with only 3G connection can stream videos of live event on the web in a remote town with ease [1]. Business people use wireless services to receive email, to check stocks , and/or even make deals [2] [3]. Doctors use remote wireless sensors to collect key health information from their patients, and use video conference to oversee/advise sophisticated medical operations [4]. Industry leaders, such as Cisco and Ericsson have claimed bold growth in wireless data traffic [5] [6]. Cisco predicted that data traffic will increase by a factor of 39 from 2009 to 2014. Ericsson claimed an even more ambitious growth of 50 folds by 2015. The consensus within the industry is that there will be substantial growth in demand for wireless traffic over the next five years.

Unfortunately, the seemingly inexorable and exponential growth in wireless data traffic has its side effect. It has already set alarm for the decreasing availability of quality wireless

spectrum and caused the so-called spectrum shortage problem [7]. The increasingly crowded wireless spectrum [8] is unable to keep pace with the gigantic growth of wireless data service while available clean spectrum is running out fast. For example, AT & T phased out its unlimited internet data plan for smart phones and other mobile devices to ease congestion on its network [9] [10]. In the ISM band, the relatively small bands has long being crowded by various applications ranging from industry monitoring sensor network based on ZigBee, ad hoc gaming connection using Bluetooth, to WLAN deploying the prominent 802.11 standard.

The limited availability of quality spectrum constrains the wireless systems from improving and evolving. As the number of users increases, more users have to share the same spectrum band. This will bring the overall service quality down. For example, WiFi network users may experience frequent packet drop and low throughput due to interference from other users. Cellular network users may experience more blocked calls due to limited cell capacity. The problem will be even more challenging when video contents became prevalent on wireless. The substantially higher data rate required to support video contents delivery will require substantially more spectrum to support.

1.2 Dynamic Spectrum Access

Although there is a perceived spectrum shortage problem, recent studies [11] [12] reveal that most of the spectrum is actually largely underutilized. It is not as striking as it seems given the static spectrum allocation policy adopted by spectrum regulation agencies around the world. The static spectrum allocation system works sufficiently well in the last century. It only becomes problematic when wireless data service exploded and the limited availability of quality spectrum is insufficient to support the increasing number of systems.

The apparent inefficiency of static spectrum regulation has motivated the wireless industry to consider Dynamic Spectrum Access (DSA) technology, a system that allocates spectrum in a dynamic and/or opportunistic fashion. Theoretically, Dynamic Spectrum Access (DSA)

can achieve near-optimal spectrum efficiency by allowing devices to detect and dynamically utilize any spectrum that is not actively used by other users. It has the potential to significantly reduce or even solve the so called spectrum shortage problem.

1.3 Research Motivation

1.3.1 Advanced Throughput Model for DSA networks

DSA network has since been extensively researched from many angles, such as spectrum sensing, resource management, protocol design, policy/regulation, etc [13][14][15][16][17]. Theoretical development of dynamic spectrum access networking, however, has received much less attention. Such development is important. The theory is needed to justify the development of DSA networks and insights gained from such research can guide the design and standardization of future DSA technology.

Some papers have discussed the capacity of cognitive radio channels from information theory perspective [18] [19] [20]. These papers shed light on the potential capacity gain on a single cognitive radio channel. However, such studies do not address the system level performance of a DSA network and provide little information on how we can utilize frequency-agile radios at the network level. We refer the readers to [20] for a comprehensive survey on the topic. In [18], the authors investigated the capacity of multihop DSA networks and derived a capacity bound. However, the tightness of this bound is unknown and is not discussed. Moreover, the simplified graph theory based model is not suitable for refined analysis and does not consider spectrum sensing which is a fundamental issue in DSA networks.

In general, a theoretical foundation that enables the analysis and evaluation of the throughput performance of DSA networks at the system level is still lacking.

1.3.2 Efficient Dynamic Radio Resource Management

On the other side, from a protocol and algorithm prospective, even though advanced radio technology such as Software Defined Radio (SDR) [21] and Cognitive Radio (CR) [22][23] have been around for a while, most existing spectrum allocation works are still based on the static channelization assumption where the center frequency and the bandwidth of wireless channels are predefined and programmed into radio firmware.

The assumption of static channelization is well established in traditional radio designs, where a radio is only designed for a specific type of communication, such as a cellular system, GPS system, TV broadcasting system or a WLAN. Devices used in these systems operate in predefined and stable spectrum ranges. Since the boundary of the available spectrum range is known to both system designers and device manufacturers, fixed channelization of this predefined spectrum range is appropriate for these systems. A pre-calculated channelization of the available spectrum and a carefully designed channel allocation scheme, e.g. [24][25], can hence be used to optimize the network performance.

In dynamic spectrum access networks, however, either the radio has to be able to operate in a very broad range of frequencies or the boundary and characteristics of the available spectrum are not exactly known to system designers. For example, depending on the activities of primary users, a spectrum whitespace may be located in any spectrum range with any size. Therefore, a spectrum access scheme based on predefined channelization is often not flexible enough and leads to low spectrum efficiency [11]. Hence, new algorithms and protocols are needed to exploit the vastly improvable spectrum utilization by reconfiguring both channel spectrum width and central frequency.

1.4 Research Contributions

The main research goal of this dissertation is to exploit frequency-agile radios and to increase the efficiency of spectrum utilization in wireless networks.

We first present a mathematical model for multihop DSA networks based on two dimensional Poisson point processes [26]. Compared to previous models, we are able to include essential DSA mechanisms such as spectrum sensing and primary interference avoidance into an elegant mathematical representation while keeping a closed form expression. The simulation results conform with empirical studies. Moreover, the numerical results provide fundamental insights on the design of future DSA networks and show potential as an evaluation tool for practical implementations.

We then capitalize on partially overlapped channel based frequency-agile radio networks [27] [28]. We propose simple joint channel scheduling and flow routing optimization algorithm that maximizes network throughput in POC (Partially Overlapped Channel) based wireless networks. The model quantifies the impact of fundamental network settings, such as node density and traffic load, on the performance of POC based networks. The quantitative analysis leads to the following discoveries:

- The use of partially overlapped channels significantly improves network performance in most scenarios.
- The benefit of using POC is higher in denser networks with higher traffic load.
- In extreme cases, such as when the radius of the network is comparable to the interference range, the marginal gain of using POCs can diminish or even reversed (e.g., negatively affect the network throughput). The use of POC should be avoided in such networks.

We then propose JSSRC [29] (Joint Spectrum Sharing and Rate Control) a distributed and cross-layer spectrum sharing algorithm for wireless networks. JSSRC distributively deter-

mines spectrum sharing policy, flow route and transmission rate for nodes in a wireless network to achieve optimal aggregate spectrum utilization. To obtain a computationally tractable solution, we divide the spectrum sharing problem into two subproblems: channel width allocation and link layer frequency scheduling. We model the channel width allocation problem as a joint end-to-end rate-control (transport layer) and bandwidth allocation (link layer) optimization problem which is solved through dual decomposition. We propose a greedy scheduling algorithm TWSR (Timing Window Based Spectrum Reservation) to solve the link layer frequency scheduling problem. TWSR approximates the optimal spectrum width obtained in JSSRC. The two algorithms JSSRC and TWSR together provide a distributed and efficient solution for the inherently cross-layer problem of spectrum sharing in frequency-agile radio networks.

We then extend JSSRC to incorporate routing and present TRSS [30] to solve the much more complex joint Transport, Routing and Spectrum Sharing optimization problem. Both JSSRC and TRSS enjoy theoretical convergence and achieve optimal with appropriate scheduling algorithms and will drastically improve efficiency of spectrum utilization in frequency-agile radio networks.

The work presented in this dissertation together represents a systematic initiative to explore the potentials of recent technology advancements in signal processing, radio design and network design to push efficient spectrum utilization in wireless networks from a variety of angles such as system modeling, cross-layer optimization and distributed algorithm design. The designs represent state of the art results in each of the corresponding areas. The numerical and simulation results obtained show the potential of our designs to address the spectrum shortage problem that has become increasingly prominent and urgent.

1.5 Dissertation Organization

The rest of the proposal is organized as follows. In Chapter 2, we introduce some background technologies and terminologies used in this dissertation. In Chapter 3, we present a general throughput model that enables analytical study of DSA networks. In Chapter 4, we present a simple linear optimization model for POC based frequency-agile radio networks. In Chapter 5, we describe the framework of JSSRC and its main operations. In Chapter 6, we present TRSS algorithm and its main operations. We conclude this dissertation in Chapter 7.

Chapter 2

Background

2.1 Frequency-Agile Radio Technology

Frequency-agility has been historically used for some radar systems that are able to quickly move to a secondary frequency when attacked by a jammer on the primary frequency [32]. Like many technologies in the wireless communication field, there is no universally accepted definition of frequency-agility. Researchers in different domain of the field may have completely different views on the actual requirements on frequency-agility, e.g., the range, speed, accuracy of the frequency change. Since defining frequency agility is not the primary goal of this dissertation, we only use Frequency-Agile Radio (FAR) as an umbrella term for software defined radio and cognitive radio. In the following, we briefly discuss these two technologies.

A software-defined radio system, or SDR, is a radio communication system that uses software for most of its signal processing operations instead of using special-purpose hardware such as mixers, filters, amplifiers, modulators/demodulators, detectors. The major advantage of SDR is its ability to transmit and receive widely different radio protocols/waveforms with only software tuning. Software radios have been already widely deployed for the military and cell phone services, both of which require a variety of different radio protocols on the

fly. SDR is considered an important enabler of cognitive radio.

Cognitive radio is considered to be the center piece of a DSA network. In this dissertation, we define a "Cognitive Radio" as a radio that can learn from its radio environment and adapt accordingly to improve performance. The performance here can either be the performance of the radio itself or the performance of the network that the radio belongs to. The two defining features of a cognitive radio are its Cognitive Capability and Software Adaptability. Cognitive Capability means that the radio can learn essential information from its radio environments such as primary user existence and secondary interference level. Currently, such information is acquired through spectrum sensing and/or information exchange among cognitive radios. Software Adaptability means that the radio can adapt its radio parameters through fast software reconfiguration according to the information it acquired from the network with the goal to improve network/radio performance.

2.2 Network Utility Optimization

Kelly et al. [33] presented an innovative optimization framework where a network resource allocation problem is mathematically represented as a utility maximization problem. The framework has a general form as follows.

$$\begin{aligned} Maximize \quad & \Sigma_i U_i(x_i) \\ Subject \quad to \quad & \mathbf{Rx} \leq \mathbf{c} \end{aligned}$$

In the formulation, the source rate vector $\mathbf{x}$ is the set of optimization variables, one for each data source or node indexed by i. $\mathbf{R}$ is a binary matrix representing the routing in the network, i.e., 1 indicates a link is selected for transmission and 0 indicates a wireless link is idle. $\mathbf{c}$ is the link capacity vector. $U_i(\cdot)$ is the utility function of source i. The choice of utility function determines the optimization objective of the whole formulation. It can take input in the forms of data rate, latency, energy, etc. In general, it is assumed that the utility function is smooth, concave and non-decreasing.

This formulation and its variants have been extensively researched over the past decade to study resource allocation in communication networks [34]-[37]. The framework is important that it brings the traditionally empirical study of network protocols into the solid mathematical domain. For example, the framework has been used to reverse-engineer the Transport Control Protocol (TCP) [38] which can be viewed as a dual-decomposition based solution to a convex optimization problem that maximizes the aggregate date rate of all data flows [39]-[42]. With a deeper understanding of TCP gained from these mathematical derivations, Low etc., from California Institute of Technology have invented FAST-TCP [40], a new TCP congestion control protocol for high-speed long-latency networks. The related algorithm has been successfully commercialized [41] and deployed in real networks.

The framework of network utility optimization captures and quantizes the cross-layer interaction between layers and among nodes. The dual-decomposition form of the framework decomposes the complex primal problem into many smaller sub-problems in dual form. The decomposed dual form matches almost perfectly with the distributed and layered computing architecture of the Internet. These subproblems are internally weaved together by certain shared network parameters such as queue length at each node in the form of primal or dual variables.

Despite all the merits of this amazingly elegant framework, the key question, however, is whether the optimization based approaches can be directly applied to wireless networks and incur the same success it had in wireline networks. Indeed, there are many challenges in the wireless context that do not allow a direct copy of such techniques from its well-defined wireline applications. More specifically, wireless network is in general a multi-access system. In such system, the transmissions of users interfere with each other. The incurred interdependencies among users and network layers are simply not present in wireline systems. Moreover, the channel capacity is time-varying because of node mobility, multi-path and shadowing effect, etc. Despite the challenges, researchers have made significant achievements in this field and successfully demonstrated that wireless resources across multiple layers (such as time, frequency, power, link data rates and end-user data rates), can also be included into

the framework [43]-[50].

2.3 Partially Overlapped Channels

Utilizing Partially Overlapped Channels (POCs) is a straightforward way to take advantage of frequency-agile radios. POCs refer to wireless channels that have spectrum overlap with other channels. For example, in the IEEE 802.11b/g wireless standard, only channel 1, 6 and 11 are considered non-overlapping (orthogonal). Other channels are considered partially overlapped with either one or two of the non-overlapping channels.

From a signal processing prospective, a closely related concept is the so-called Adjacent Channel Interference (ACI) which refers to the physical signal impairment to one frequency band (channel) due to interference from signal on adjacent frequency bands (channels). The design method of classical wireless communication systems emphasizes on channel separation and orthogonality and considers ACI as hardware and software defects caused by incomplete filtering, improper tuning or poor frequency control. In the innovative POC based channel assignment designs, however, spectrum overlapping of different channels are is considered harmful. Spectrum overlap in POC based system can be a result of ACI. But more generally, we refer to POC based design as an approach to intentionally employ channels with partially or fully (co-channel) overlapped spectrum band to take full advantage of all available spectrum. The channel overlap in POC based design is a natural result of spectrum segmentation/channelization methods being used in the existing systems such as IEEE 802.11b/g. Instead of prohibiting the usage of channels with overlapped spectrum, POC based design let nodes to decide by themselves on whether a specific channel is usable based on their local observations. The primary idea is to provide nodes with full access to all wireless channels in the available spectrum to increase channel diversity and leverage overall network capacity.

A handful of papers on POC based design have been published in literature. In [53], Mishra, et al., introduce the concept of POC based network design. The authors evaluate the per-

formance of POC in WLAN and discuss methods to adapt existing protocols to use POCs. More recently, in [54], Liu, et al., propose a POC-aware channel allocation scheme using genetic algorithm. In [55], Rad, et al. formulate POC allocation as a linear mixed integer problem.

While these works give good demonstration of the potential of POC based network design, a comprehensive investigation to characterize the behavior of partially overlapped channels in various network settings is needed for any practical consideration of such designs. This characterization should be able to analytically evaluate the actual benefit (e.g., capacity improvement) of using POC based design versus more traditional designs. This characterization should work in a wide range of network settings such as channel/frequency separation, transmitting rates, network topology and node density that are present in real problems. To the best of our knowledge, there has not been any work that directly give an in-depth treatment on this topic.

2.4 Protocol Interference Model

Two main interference models have been widely used for wireless network research, the protocol model and the physical model.

In protocol model, the existence of interference is binary and is determined solely by whether a node geographically falls into the interference range of another nodes. For example, any node (other than the intended receiver) within the interference range of a transmitter is considered to be interfered. These nodes cannot receive signals transmitted to them correctly as defined in protocol model. On the other hand, any node that is not within the interference range of a transmitter is consider not to be interfered by this transmitter.

In physical model, interference and noise are treated the same. A transmission is successful only when the SINR (signal-to-interference-and-noise-ratio) at the intended receiver exceeds some threshold. Compared to protocol model, physical model is a closer representation to

real systems. However, the computational complexity incurred by using physical model in cross layer design can be extremely high for large-size networks due to the non-convexity of the SINR function (with regard to transmission power) used in physical model. As a results, a large potion of the existing algorithms and protocols for wireless network use the simpler protocol model for their analysis [34]-[37][43]-[50].

In this dissertation, we use the protocol model for our analysis for tractability. To circumvent some of the drawbacks associated with the protocol model, such as over-simplification and neglect-of-capture-effect, we recommend techniques such as "reality check" [56] to help improving the performance of protocol interference model in real systems. As pointed out in [56], the efficacy of the protocol model depends on the performance gap between the result obtained using reality check and the result obtained under the physical model. If the gap is small, the protocol model is a good approximation and can be used effectively as an alternative for physical model for wireless network analysis.

2.5 System Spectrum Efficiency

The link spectral efficiency of a communication link is defined as the information bit rate over a communication link divided by the bandwidth in hertz of the channel used by the link. Spectral efficiency is jointly determined by many factors such as modulation scheme used, pulse shaping scheme used, channel coding scheme used, etc.

Higher order modulation schemes are generally considered to be more spectral efficient than lower order modulation schemes, as higher order modulation schemes deliver more bits in a symbol time. Channel coding schemes generally decreases channel spectral efficiency by a factor of r, where r is the coding ratio defined as the number of data bit of a code word divided by the total number of bit in that code word. Pulse shaping increases channel spectral efficiency by reducing powers in side lobes.

The Shannon Channel Capacity [87] gives the upper bound on the data rate that can be

reliably received (i.e., with arbitrarily low bit error rate) in a communication channel with a given signal to power ratio. In an Additive White Gaussian Channel (AWGN), this capacity is determined to be

$$C = Blog_2(1 + \frac{S}{N}), \tag{2.1}$$

where B is the bandwidth of the channel, S is the total signal power over the bandwidth and N is the total noise power over the bandwidth. Both S and N are measured in watt.

Link spectral efficiency is relatively easier to measure and is useful as as an indicator of the capability of a one-hop link to deliver information data bit. However, in wireless networks, spectral efficiency on a single link is rarely a good measure of the performance of a entire network system (unless it consists of only one single-hop one-channel link). This is because link spectral efficiency does not take into account of spectrum reuse. Systems with low spectrum efficiency at individual links may achieve higher overall data throughput than systems that have a few links with higher link spectral efficiency. For example, links in a data collecting sensor network that might transmit a few hundred bits per day is not considered spectral efficient on a single link by definition. But a large number of such collecting devices over a few square miles may achieve a combined efficiency that surpass any single link systems.

In wireless network systems, we are more concerned about the performance of the network as a whole but not a single link. In this case, a more relevant measure is per network spectral efficiency [(bit/s/Hz per site/network]. It is computed as the aggregate network throughput or goodput, summed over all nodes in the network, divided by the channel bandwidth that are allocated to the network. Network spectral efficiency is affected not only by spectral efficiency on each single link, but also and in general more importantly by multiple access schemes and radio resource management techniques utilized in the network. If it is defined over the aggregate goodput, traffic from retransmissions caused by co-channel interference and collisions is not included in the computation.

In practice, a combined fairness measure and system spectrum efficiency (Fairly Shared

Spectrum Efficiency or FSSE) [57] is often used to ensure a fair share of the spectrum by all users. In this case, the aggregate throughput in the definition of system spectrum efficiency is replaced by the utility function corresponding to the fairness criterion used. In this proposal, we will use the combined fairness and system spectrum efficiency as the primary performance measure of our designs.

Chapter 3

Spectrum Sharing Model for Dynamic Spectrum Access Networks

DSA network has been considered one of the most important applications of frequency-agile radios. Preliminary results on DSA networks [11] [12] have claimed significant improvements over network based on traditional spectrum access methods in terms of spectrum utilization efficiency. In this chapter, we address the technical void of DSA network characterization and system level throughput analysis. We propose a technical model for DSA network based on two dimensional Poisson Point Processes (P.P.P.) where both the primary network and the secondary network are modeled as Poisson processes. By including essential DSA mechanisms such as spectrum sensing and primary interference avoidance, the model captures the main operations of a real DSA network. The resulting model is capable of quantifying the impact of several key network settings, such as network traffic load, radio interference range and network node density, on the throughput of secondary networks. Based on the presented model, the numerical evaluation presented in this chapter reveals important insights on the design of DSA networks. All the related notations used in this chapter are listed in Table 3.1 for reference.

Table 3.1: Notations for Chapter 3

λ_1	Primary node density
p_1	Probability to transmit per primary node per slot
λ_2	Secondary node density
p_2	Probability to transmit per secondary node per slot given an opportunity is detected
R_i	Interference range of primary transmitters to primary receivers
R_p	Transmission range of primary transmitters to primary receivers
r_i	Interference range of secondary transmitters to secondary receivers
r_p	Transmission range of secondary transmitters to secondary receivers
r_d	Detection range of secondary transmitters
$\tilde{R}_i$	Interference range of primary transmitters to secondary receivers
$\tilde{r}_i$	Interference range of secondary transmitters to primary receivers
N_{R_i}	Average number of primary nodes in a primary interference disk
N_{R_p}	Average number of primary nodes in a primary transmission disk
N_{r_i}	Average number of secondary nodes in a secondary interference disk
N_{r_p}	Average number of secondary nodes in a secondary transmission disk
N_{r_d}	Average number of primary nodes in a secondary detection disk
d	Distance from secondary transmitter to secondary receiver
Π	One-hop throughput measured as the average number of successful transmissions per slot per node
$\mathcal{Z}$	Expected Packet Progress per secondary node per slot

3.1 System Model

3.1.1 Primary Network Model

The architecture of a DSA network considered in this chapter is illustrated in Figure 3.1.

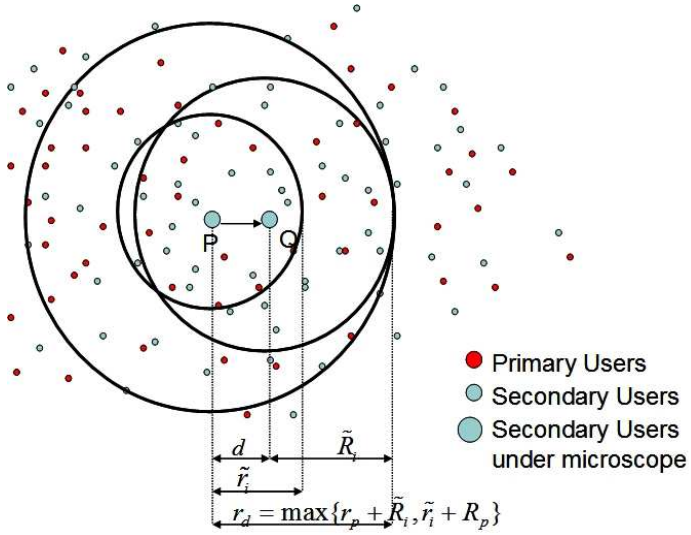


Figure 3.1: 2-D Poisson Aloha DSA Network Model

The nodal distribution of the primary network is modeled as a two-dimensional Poisson point process with density λ_1. The time axis is divided into equal-length slots with size equals to transmission time of a (fixed length) packet plus the guard time. The transmitting protocol of the primary network is slotted Aloha. All primary nodes are assumed to have packets ready to be sent at all times. Each primary node attempts to transmit with probability $p_1 \in (0, 1]$ at the beginning of each slot. If a collision occurs, affected transmissions are rescheduled to some random slot in the future. We assume that a new sample of the spatial distribution is given for every slot, i.e., the distribution of the network is independent from slot to slot. We further assume that a packet will be destined to any node with equal probability with no differentiation between a new packet and a retransmitted packet (Two dimensional isotropic/ uniform traffic over the plane). We assume that all primary nodes use the same transmission power with transmission range R_p and mutual interference range R_i. We define nodes within distance R_p as *transmission neighbors* and nodes within R_i as *interference neighbors*. We define $N_{R_p} = \lambda_1 \pi R_p^2$ and $N_{R_i} = \lambda_1 \pi R_i^2$ respectively to be the average number of transmission

neighbors and the average number of interference neighbors. A successful transmission occurs if the intended receiver and the intended receiver's interference neighbors do not transmit in the same slot. Otherwise, a collision occurs.

3.1.2 Secondary Network Model

The secondary network is deployed on top of the primary network and has a similar node distribution and traffic model as the primary network. For example, secondary network has a two-dimensional Poisson point process with density λ_2 which is regenerated for every time slot. All secondary nodes always have packets to send. All packets are equally destined to all directions in a two dimensional plane. All secondary users use the same transmission power with transmission range r_p and interference range r_i. N_{r_p} and N_{r_i} is defined as the average number of *secondary transmission neighbor* and *secondary interference neighbors*.

The secondary network is an opportunistic network that is allowed to access the spectrum when it does not impact the primary network. As a result, a secondary node senses the spectrum and only attempts to transmit if an opportunity (a spectrum white space) is detected. We denote the conditional probability that a secondary node attempts to transmit given a detection of opportunity as p_2. In this chapter, we assume all nodes (primary or secondary) use the same spectrum band. Our model can be straightforwardly extended to the cases of multi-channel DSA system, where nodes can switch between different channels.

3.1.3 Spectrum Sensing and Opportunity Detection

Define OP as the opportunity indicator. In a DSA network, for the transmission from a secondary transmitter to a secondary receiver, we say that an opportunity exist (OP=1) when the transmitted signal does not interfere any primary receiver and the reception at the intended secondary receiver is not interfered by any primary transmitter. Otherwise we say that opportunity does not exist for this transmission (OP=0).

In practice, spectrum sensing (cooperative or non-cooperative) is done only by the secondary transmitters. Moreover only primary transmitter activity can be detected. In other words, opportunity detection is based only on local observations of primary transmissions at secondary transmitters. Define H as the spectrum sensing detector. In most non-cooperative detection schemes, a secondary node claims an opportunity detection (H=1) if there is no primary transmission sensed in its detection range r_d. Otherwise, the secondary node detects no opportunity (H=0). Due to the lack of information from the receiver, this strategy is prone to false alarms (decide H=0 while OP=1) and miss detections (decide H=1 while OP=0) errors. The choice of r_d plays a central role to the tradeoff between these two phenomena. In general, detection strategies with smaller r_d see more miss detections and less false alarms. These strategies are more aggressive in secondary spectrum access but provide less protection to primary users and the intended receivers. On the other hand, strategies with larger r_d see less miss detections but more false alarms. These strategies provide better protection to primary receivers but may compromise secondary network throughput due to false alarm. We refer the readers to [68] for an interesting discussion on the issue.

In this chapter, we take a conservative strategy for opportunity detection. We choose r_d to be large enough that whenever an opportunity detection (H=1) is claimed at the secondary transmitter, both primary receivers and the intended secondary receiver is strictly protected. That is primary receivers will not be interfered by this secondary transmission and the intended receiver is guaranteed to have no primary interference.

The disadvantage of this strategy as discussed above is that the secondary network will lose some transmission opportunity when false alarm happens. For more efficient detection strategy, the throughput derived in this chapter can be used as a lower bound. It is shown in the appendix of this chapter that the smallest detection range to guarantee no primary interference at the intended receiver and no primary receiver within the interference range of a second transmitter is $r_d = \max\{r_p + \tilde{R}_i, \tilde{r}_i + R_p\}$, where $\tilde{R}_i$ is defined as the interference range of primary transmitter to secondary receivers and $\tilde{r}_i$ is defined as the interference range of secondary transmitter to primary receivers.

3.2 One-Hop Performance Model

3.2.1 Performance Model

We derive the one-hop throughput measured as the average number of successful transmissions per slot per node denoted as Π.

Note that a primary network in general has larger transmission range, interference ranges and power than a secondary networks. This means we normally have $\tilde{R}_i \geq \tilde{r}_i$ and $R_i \geq r_i$. In this chapter, for practical reasons, we only discuss scenarios where the network hold this property. For a similar discussion on scenarios where $\tilde{R}_i < \tilde{r}_i$ and $R_i < r_i$, we refer the readers to our technical report [26] for a detailed discussion.

The one-hop throughput Π equals the average probability that a node successfully transmits a packet at a time slot. For a transmission from a secondary node P to succeed in slot i, the following conditions have to be fulfilled:

c1 There is at least one secondary node in P's transmission range r_p;

c2 P detects an opportunity in slot i;

c3 P attempts to transmit in slot i given a detected opportunity;

c4 Secondary interference neighbors of P's intended receiver do not transmit in slot i.

Since secondary nodes are Poisson distributed with density λ_2, we have

$$p(c1) = 1 - e^{-N_{r_p}} \tag{3.1}$$

where $N_{r_p} = \lambda_2 \pi r_p^2$ and $e^{-N_{r_p}}$ is the probability that no secondary node exists within r_p of P.

An opportunity detection is claimed (H=1) when there is no primary users transmitting in the detection range r_d of P. Conditioned on the number of primary nodes in P's detection range and primary transmission probability p_1, we have

$$
\begin{aligned}
& \text{p(c2)} \\
= \; & \textstyle\sum_{k=0}^{\inf} Prob\{\text{there are } k \text{ primary node in } r_d\} \\
& \cdot Prob\{\text{none of the } k \text{ node transmits}\} \qquad \text{where } N_{r_d} = \lambda_1 \pi r_d^2. \\
= \; & \textstyle\sum_{k=0}^{\inf} (1-p_1)^k \frac{N_{r_d}^k}{k!} e^{-N_{r_d}} \\
= \; & e^{-p_1 N_{r_d}}
\end{aligned}
$$

Given condition $c1$ and $c2$ is true, we simply have

$$p(c3) = p_2 \tag{3.2}$$

We now proceed to derive $p(c4)$ which turns out to to be the most complex.

Let us start with the simplest case where Q has only one secondary interferer $\tilde{Q}$. We compute the probability that both Q and $\tilde{Q}$ kept silent given there is an opportunity detection claimed at P. This simple case is non-trivial since the event that $\tilde{Q}$ claims an opportunity detection is not independent of P and Q's opportunity detection due to the overlap of P, Q and $\tilde{Q}$'s detection disk as illustrated in Figure 3.2. In the scenarios considered in this chapter ($\tilde{R}_i > \tilde{r}_i$,$r_d = \max\{r_p + \tilde{R}_i, \tilde{r}_i + R_p\}$, $\frac{r_p}{r_i} = 0.5$), we normally have r_d much larger than d and the difference between P and Q's detection zone is very small. As a result, $\tilde{Q}$'s detection probability conditioned on both P and Q's detection is very close to the result obtained by conditioning only on P's detection. To simplify our analysis, in this chapter we will ignore the effect of Q's detection on its neighbor's detection. It is shown from our simulation results that this is a good approximation.

Intuitively, if P detects an opportunity, the probability that $\tilde{Q}$ detects an opportunity should be higher since it is known that there has no transmitting primary user in the overlapped portion of their detection ranges. Actually, the probability that $\tilde{Q}$ can detect opportunity then is solely determined by the primary users in the shaded region $S_c(P, \tilde{Q})$ as shown in Figure 3.2. Suppose $\tilde{Q}$ is located at (ρ,θ) in a polar coordinate system that is originated at Q and use the line PQ as the polar axis. Then the distance between P and $\tilde{Q}$ is

$$x = \sqrt{d^2 + \rho^2 - 2\rho d \cos(\pi - \theta)},$$

where d is the distance between P and Q.

The size of the shaded region is determined as

$$S_c(P, \tilde{Q}) = \pi r_d^2 - 2r_d^2 q(\frac{x}{2r_d}), \tag{3.3}$$

where $q(t) = \cos^{-1}(t) - t\sqrt{1-t^2}$ as defined in [78]. The conditional probability that $\tilde{Q}$ detects an opportunity is then determined by the following equation

$$\begin{aligned} & \eta(d, \rho, \theta) \\ = \; & e^{-p_1 \lambda S_c(X,Y,Z)} \\ = \; & e^{-p_1 \lambda [\pi r_d^2 - 2r_d^2 q(\frac{\sqrt{d^2+\rho^2-2d\rho\cos(\pi-\theta)}}{2r_d})]} \end{aligned}$$

Since $\tilde{Q}$ is equally likely to be located anywhere inside Q's interference range, the average probability that an interference neighbor of Q detects an opportunity is

$$\eta(d) = \int_\rho \int_\theta \eta(d, \rho, \theta) \rho d\rho d\theta. \tag{3.4}$$

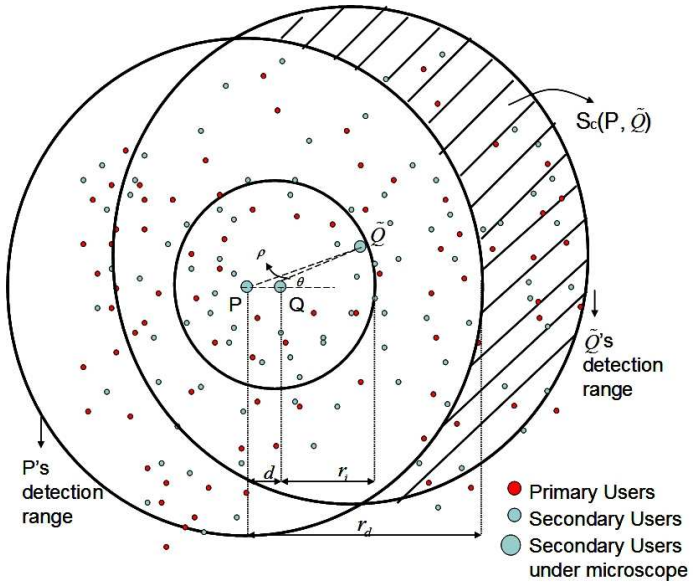


Figure 3.2: Opportunity detections at different nodes are correlated

Now we proceed to consider the more complex case of k interference neighbors. Not surprisingly, the opportunity detection probability of these interference neighbors are not independent due to the overlapping of their detection ranges. Furthermore, the location of

these interference neighbors are random due to the 2-D poisson assumption. To get an exact form of Prob{Transmission from P to Q is successful} is feasible but the expression will be extremely complex and is untractable for theoretical analysis due to the spaghetti-like correlation among opportunity detections and random location of Q's interference neighbors. Fortunately, we have the following proposition,

Proposition 1 Given there are k secondary interference neighbors of Q and d(P,Q)=d, the average number of nodes $\tilde{k}$ that detect opportunity is determined as $\tilde{k} = k\eta$ and is independent of their correlation.

Proof: Define ζ_1, ζ_2, ...,ζ_k as binary detector for interference neighbor 1,2, ..., k.

$$\begin{aligned} & \text{E[number of nodes that detect opportunity]} \\ = \; & E[\zeta_1 + \zeta_2 + ... + \zeta_k] \\ = \; & E[\zeta_1] + E[\zeta_2] + ... + E[\zeta_k] \\ = \; & \eta(d) + \eta(d) + ... + \eta(d) \\ = \; & k\eta(d) \end{aligned}$$

The importance of proposition 1 is that, given the number of interference neighbors k, the average number of interference neighbors that detect opportunities, denoted as $\tilde{k}$ is deterministic. We can then take advantage of this property and approximate the conditional probability that transmission from P to Q is successful given k secondary interference neighbors to be the probability that none of $\tilde{k}$ secondary users with detected opportunities transmits, i.e.,

$$\begin{aligned} & \text{E}[Prob\{Q \text{ has no secondary interference given} \\ & \qquad Q \text{ has } k \text{ secondary interference neighbors}\}] \\ \approx & Prob\{\text{None of the } \tilde{k} \text{ neighbors with detection transmits}\} \\ = & (1-p_2)^{\tilde{k}}. \end{aligned} \tag{3.5}$$

By summing among all possible number of secondary interference neighbors, we get a close approximation of the probability $p(c4)$.

$$
\begin{aligned}
& p(c4) \\
=& Prob\ \{Q \text{ does not transmit}\} \\
& \cdot \sum_{k=0}^{k=\inf} [Prob\ \{\text{k interference neighbors within } r_i\} \\
& \cdot Prob\ \{\tilde{k} \text{ neighbors with detection do not transmits}\}] \\
=& (1 - p_2 e^{-p_1 \lambda_1 [\pi r_d^2 - 2 r_d^2 q(\frac{d}{2r_d})]}) \\
& \cdot \sum_{k=0}^{k=\inf} \frac{N_{r_i}^k}{k!} e^{N_{r_i}} (1 - p_2)^{\tilde{k}}
\end{aligned} \tag{3.6}
$$

Our simulation results in the next section show that this is a good approximation.

We are finally in the position to compute $\Pi(d)$ where d is the distance between X and Y.

$$
\begin{aligned}
& \Pi(d) \\
=& Prob\ [\text{there is at least one teriminal within } r_p] \\
& \cdot Prob[P \text{ detects opportunity}] \\
& \cdot Prob[P \text{ transmits with detection}] \\
& \cdot Prob[\text{Transmission from } P \text{ to } Q \text{ is successful}] \\
=& p(c1) \cdot p(c2) \cdot p(c3) \cdot p(c4) \\
=& (1 - e^{-N_{r_p}}) \cdot e^{-p_1 N_{r_d}} \cdot p_2 \cdot (1 - p_2 e^{-p_1 \lambda_1 [\pi r_d^2 - 2 r_d^2 q(\frac{d}{2r_d})]}) \\
& \cdot \sum_{k=0}^{k=\inf} \frac{N_{r_i}^k}{k!} e^{N_{r_i}} (1 - p_2)^{\tilde{k}}
\end{aligned} \tag{3.7}
$$

Take the expectation of $\Pi(d)$ over probability density function of d we can get Π, the average number of successful transmission per slot per node.

Remarks 1) If primary network density $\lambda_1 = 0$, i.e., there is no presence of primary network,

we have $\lambda_1 = 0$, $N_d = 0$, $\tilde{k} = k$. Then equation (3.7) is reduced to

$$\begin{aligned}&(1 - e^{-N_{r_p}}) \cdot p2 \cdot (1 - p_2) \cdot \sum_{k=0}^{k=\inf} \frac{N_{r_i}^k}{k!} e^{N_{r_i}} (1 - p_2)^k \\ =&(1 - e^{-N_{r_p}}) \cdot p2 \cdot (1 - p_2) \cdot e^{-p_2 N_{r_i}} \end{aligned} \tag{3.8}$$

which is equivalent to the classical results in [78] for one-hop throughput in slotted aloha multihop radio networks.

2) if $\tilde{R}_i \gg \tilde{r}_i$, i.e., the primary network is a high power long range network (e.g. mesh network) and the secondary network is a low power local area network (e.g. home WLAN), then we have $\frac{d}{r_d} \approx 0$, $\tilde{k} = k$. Equation (3.7) is reduced to

$$\begin{aligned}&(1 - e^{-N_{r_p}}) \cdot e^{-p_1 N_{r_d}} \cdot p2 \cdot (1 - p_2) \\ &\cdot \sum_{k=0}^{k=\inf} \frac{N_{r_i}^k}{k!} e^{N_{r_i}} (1 - p_2)^k \\ =&e^{-p_1 N_{r_d}} \cdot [(1 - e^{-N_{r_p}}) \cdot p2 \cdot (1 - p_2) \cdot e^{-p_2 N_{r_i}}]. \end{aligned} \tag{3.9}$$

Note that the latter part (in square brackets) of equation (4.13) is exactly the same as equation (4.12). This indicates that all nodes in the secondary network see the same spectrum opportunity and when such opportunity is detected, the secondary network is not different from a normal multihop radio network.

3.2.2 Numerical Results

In the following, we use the model presented in previous sections to study the one-hop secondary network throughput. In all scenarios, we have $\tilde{R}_i > \tilde{r}_i$,$r_d = \max\{r_p + \tilde{R}_i, \tilde{r}_i + R_p\}$ and $\frac{r_i}{r_p} = 0.5$.

Impact of N_{R_i} and p_1 Firstly we examine how primary network parameters affect secondary network performance. We are especially interested in two representative scenarios. In scenario I, we let $\frac{\tilde{r}_i}{\tilde{R}_i} \ll 1$, e.g., the primary network has a much larger transmission power

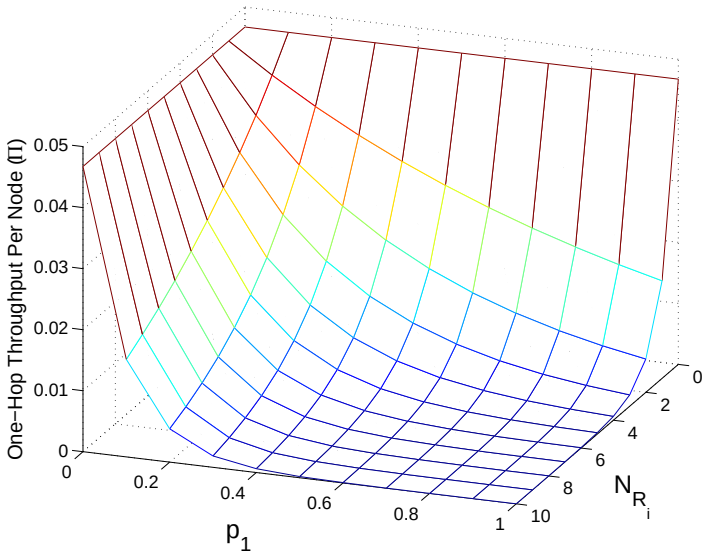


Figure 3.3: Effect of primary network traffic load and primary network density, Single-hop network, Scenario I, Primary network radio signal power is significantly larger than secondary network radio signal power

than the secondary network. In scenarios II, we let $\frac{\tilde{r}_i}{R_i} = 1$, e.g., the primary network and the secondary network is similar in transmission power. In both scenarios, we vary the average number of primary interference neighbors per primary node N_{R_i} (which represents the primary network density) and primary user transmission probability p_1 (which represents the primary network traffic load). For each pair (N_{R_i}, p_1), we numerically derive the maximum achievable one-hop throughput Π_{max} by tuning secondary network settings, e.g. N_{r_i}, p_2. The results are shown in Figure 3.3 and Figure 3.4 for $\frac{\tilde{r}_i}{R_i} = \frac{1}{10}$ and $\frac{\tilde{r}_i}{R_i} = \frac{1}{1}$ separately.

From Figure 3.3 and Figure 3.4, we see that in both cases, Π_{max} is a decreasing function of p_1 and N_{R_i}. The global maximum is obtained when either $p_1 = 0$ or $N_{R_i} = 0$, e.g., the primary network is either not present or is not utilized at all. This global maximum is exactly the optimal throughput we can obtain in a stand alone slotted aloha multihop radio network. Since p_1 is related to primary network traffic load and N_{R_i} is related to primary network density, this means with increased primary network traffic and node density,

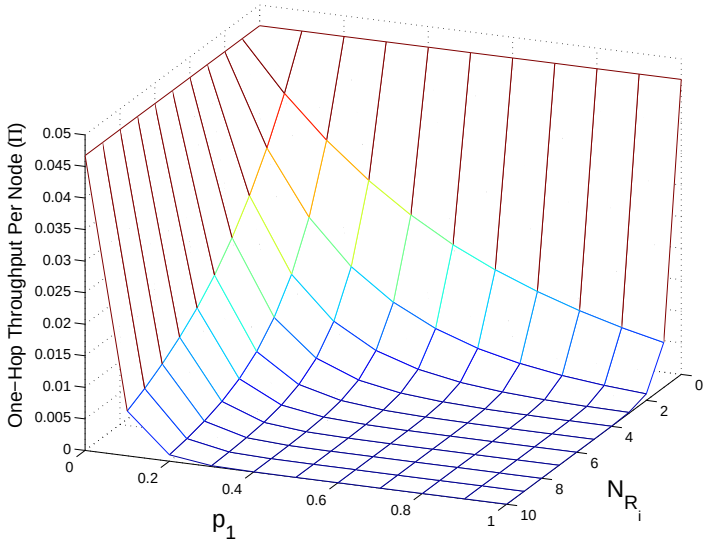


Figure 3.4: Effect of primary network traffic load and primary network density, Single-hop network, Scenario II, Primary network radio signal power is only comparable with secondary network radio signal power

secondary network throughput decreases. Another observation is that , in Scenario II, the secondary one-hop throughput degrades much faster than in Scenario I. This indicates that when primary network and secondary network have comparable transmission ranges and interference ranges, increased primary network density and traffic are more punishing on the capacity of secondary network than in cases where primary network are relatively large.

Impact of N_{r_i} and p_2 We now turn our attention to secondary network itself. In this part, we study secondary network settings and how they affect its own performance given a reasonably designed primary network. For comparison, we take the optimal settings for a stand-alone multihop radio network given in Kleinrock's classical paper [78]. Again, we study two representative scenarios as in the first part of this discussion. The numerical evaluations are shown in Figure 3.5 and Figure 3.6 for $\frac{\tilde{r}_i}{\tilde{R}_i} = \frac{1}{10}$ and $\frac{\tilde{r}_i}{\tilde{R}_i} = \frac{1}{1}$ respectively.

Again, as discussed in previous part, under the same primary network setting, the secondary

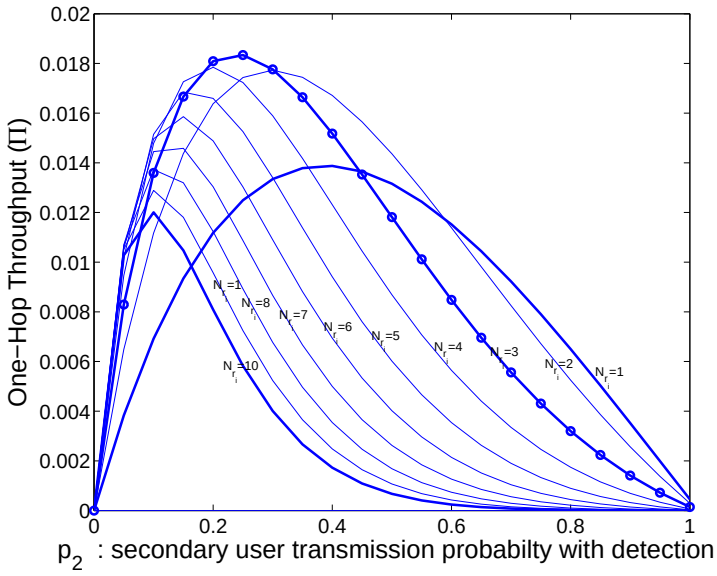


Figure 3.5: Effect of secondary network traffic load, Single-hop network, Scenario I, Primary network radio signal power is significantly larger than secondary network radio signal power

network in Scenario II performs much worse. But other than that, Figure 3.5 and Figure 3.6 are very similar. The performance is optimized under moderate traffic and node density (e.g. $p_2 \approx 0.3$ and $N_{r_i} \approx 3$ in both scenarios) where the tradeoff between offered traffic and collision is optimized. Performance curves in both figures resemble the shapes of similar results for stand alone slotted aloha networks [79]. This indicates that the existence of primary network has not fundamentally changed the behavior of secondary network. Rather, primary network only squeezes the resources available to secondary network. In the white space that left by primary network, secondary network achieves the same performance of an independent stand-alone network.

The results presented in the above sections perfectly match with our intuition on how primary network and secondary network should interact and show the validity and effectiveness of our analytical model at the qualitative level. In the following, we present simulation results to further show that our model is also accurate quantitatively.

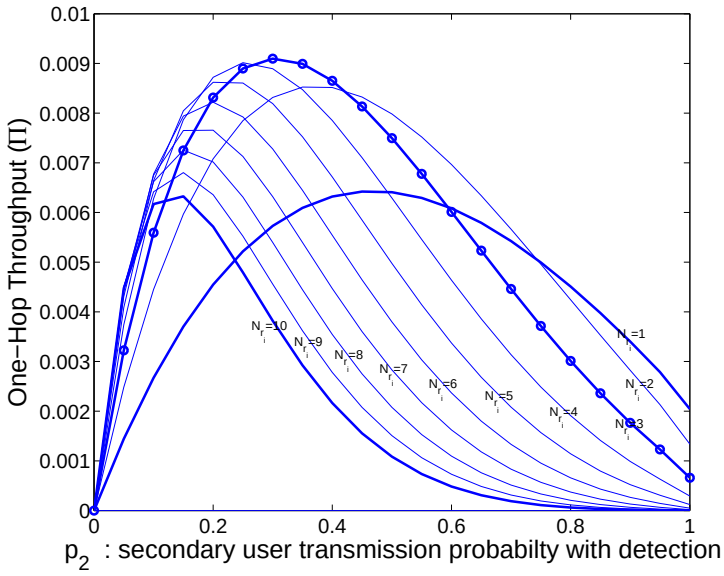


Figure 3.6: Effect of secondary network traffic load, Single-hop network, Scenario II, Primary network radio signal power is only comparable with secondary network radio signal power

3.2.3 Simulation Results

For validation, we simulate a slotted Aloha dynamic spectrum access network in $MATLAB^{©}$ [85]. In each simulation, we setup a 10000m*10000m primary network and an equal-size secondary network on top of the primary network. Both the primary network and secondary network are implemented with slotted aloha protocols. The transmission range R_p and interference range R_i of primary network is set to be 1000m and 2000m respectively. We assume that $R_i = \tilde{R}_i$ and $r_i = \tilde{r}_i$. The parameters of secondary network is set according to the required ratio, e.g. the value of $\frac{\tilde{r}_i}{\tilde{R}_i}$. For each simulation scenario, ten independent runs are performed with each run consisting of 1000 time slots. The final results are the averaged value of the 10 independent runs.

For comparison, we simulate with exactly the same network settings as used in our numerical evaluation for Figure 3.5 and Figure 3.6. The simulated results are shown in Figure 3.7 and Figure 3.8. As shown in the figures, our analytical model shows great accuracy in predicting

the overall performance of a secondary network. The analytical performance curves in Figure 3.5 and 3.6 both shows remarkable resemblance to the simulated curve in Figure 3.7 and 3.8 in both value and shape. It is also obvious that the model is more accurate when $\frac{\tilde{r}_i}{\tilde{R}_i}$ is smaller. For example, the curves in the lower section of Figure 3.8 are more straightened than their counterpart in Figure 3.6. However, even in the worst case of Scenario II, our model still gives very good prediction on optimal throughput and performance changing trends.

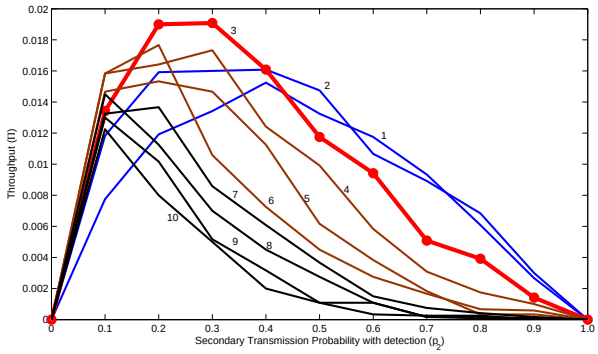


Figure 3.7: Simulation results, Effect of secondary network traffic load, Single-hop network, Scenario I, Primary network radio signal power is significantly larger than secondary network radio signal power

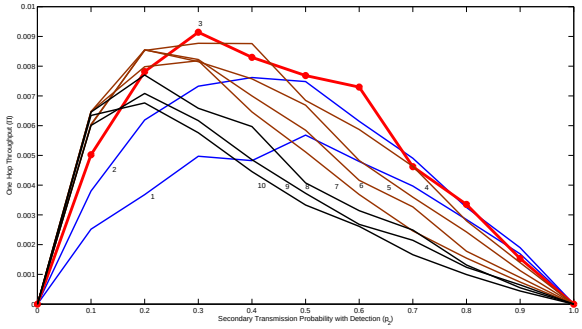


Figure 3.8: Simulation results, Effect of secondary network traffic load, Single-hop network, Scenario II, Primary network radio signal power is only comparable with secondary network radio signal power

3.3 Multihop Performance Model

Multihop performance of a secondary network is measured as the average expected progress per node per slot. The routing strategy used in this chapter is Most Forward Routing (MFR). Note that other theoretical routing strategies like Nearest with Forward Progress (NFP) and Most Forward with Variable Radius (MVR) can also be used with similar derivation. However, a comparative study of these routing strategies is out of scope of this chapter and we leave it for our future study.

In MFR routing, a node always transmits a packet to its most forward transmission neighbor in the direction of the final destination. The packet is transmitted to the least backward transmission neighbor if no transmission neighbor is found in the forward direction. Given a successful transmission, the progress x made by one transmission is the distance between the transmitter and its one-hop receiver (relaying node or destination) projected onto a line drawn towards the final destination.

The expected progress $\mathcal{Z}$ per node per slot can be computed as,

$$\mathcal{Z} = \int_{-r_p}^{r_p} x \cdot \text{prob } \{\text{actual progress equals x}\} dx. \tag{3.10}$$

In (3.10), the actual packet progress equal to x only when the following conditions are fulfilled.

$\tilde{c}_1$ P's next packet is routed towards a receiver with potential packet progress of x

$\tilde{c}_2$ P detects an opportunity

$\tilde{c}_3$ P attempts to transmit with a detection

$\tilde{c}_4$ P successfully transmits a packet to its destination Q given c1, c2 and c3, i.e., Q and Q's secondary interference neighbors do not transmit at the same slot.

The actual packet progress equals the potential progress only if the packet is successfully delivered to its chosen receiver. Otherwise, the actual progress is zero.

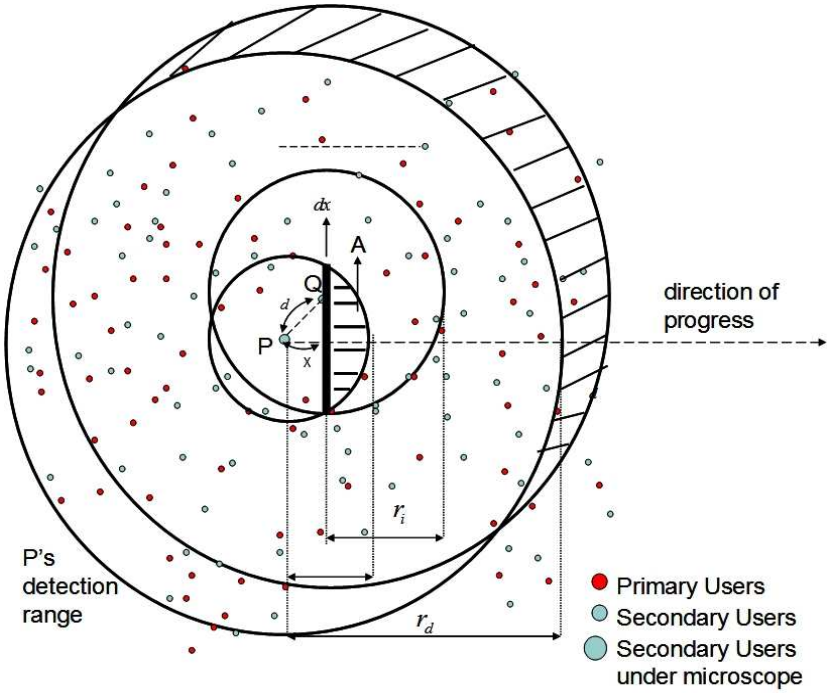


Figure 3.9: Average Expected Progress

3.3.1 Performance Model

$p(\tilde{c}_1)$ is the probability that P's next packet is routed towards a receiver with potential packet progress of x. In MFR routing, a packet is routed to a receiver with potential packet progress less than x if and only if there is no node in the shaded area A as shown in Figure 3.9, i.e.,

$$F(x) = Prob\{z \leq x\} = e^{-\frac{N r_p}{\pi} q(\frac{x}{r_p})}. \tag{3.11}$$

Taking the derivative of $F(x)$, we have

$$\begin{aligned} p(\tilde{c}_1) &= f(x) \\ &= Prob\{x < z < x + dx\} \\ &= -\frac{N_{r_p}}{\pi} e^{-\frac{N r_p}{\pi} q(\frac{x}{r_p})} q'(\frac{x}{r_p}) \frac{1}{r_p} \end{aligned} \tag{3.12}$$

where

$$q'(t) = -2\sqrt{1-t^2} \tag{3.13}$$

$p(\tilde{c}_2)$ is the probability that P detects an opportunity, i.e., there is no primary user transmitting in the detection range r_d of P.

$$\begin{aligned} & p(\tilde{c}_2) \\ =& p(c_2) \\ =& e^{-p_1 N_{r_d}}. \end{aligned} \tag{3.14}$$

$p(\tilde{c}_3)$ is the probability that P attempt to transmit with a detection. $p(\tilde{c}_3)$ simply equals p_2.

$p(\tilde{c}_4)$ is the average successful probability of a transmission from P to Q when potential progress equals x. Note that for each x, the receiver are equally likely to be located anywhere in the strip/line of length $2\sqrt{r_p^2 - x^2}$ as shown in Figure 3.9. For each Q's possible locations y, the probability of a successful transmission from P to Q is determined by the distance from P to Q. Hence, we have

$$\begin{aligned} & p(\tilde{c}_4) \\ =& \int_{-\sqrt{r_p^2-x^2}}^{\sqrt{r_p^2-x^2}} prob\{d(P,Q) = d\}\tilde{\Pi}(d)dy \\ =& \int_{-\sqrt{r_p^2-x^2}}^{\sqrt{r_p^2-x^2}} \frac{1}{2\sqrt{r_p^2 - x^2}}\tilde{\Pi}(\sqrt{x^2+y^2})dy, \end{aligned}$$

where $\tilde{\Pi}(d)$ is the probability of a successful secondary transmission from P to a receiver Q at distance d. From a similar derivation as in one-hop model, $\tilde{\Pi}(d)$ is computed as

$$\begin{aligned} & \tilde{\Pi}(d) \\ =& (1 - p_2 e^{-\lambda p_1[\pi r_d^2 - 2r_d^2 q(\frac{d}{2r_d})]}) \sum_{k=0}^{k=\inf} \frac{\tilde{N}_{r_i}^k}{k!} e^{\tilde{N}_{r_i}} (1-p_2)^{\tilde{k}}, \end{aligned} \tag{3.15}$$

where

$$\tilde{N}_{r_i} = \lambda_2(\pi r_i^2 - A) \tag{3.16}$$

is the average number of secondary neighbors in the interference range of a secondary nodes minus area A. Unlike N_{r_i} in (3.6),here, $\tilde{N}_{r_i}$ is a function of x.

Finally, we can compute the expected progress as

$$
\begin{aligned}
&\mathcal{Z} \\
=&\int_{-r_p}^{r_p} x \cdot \text{prob } \{\text{progress equals x}\} dx \\
=&\int_{-r_p}^{r_p} x \cdot \text{prob } \{\tilde{c}_1\} \cdot \text{prob } \{\tilde{c}_2\} \cdot \text{prob } \{\tilde{c}_3\} \cdot \text{prob } \{\tilde{c}_4\} dx \\
=&\int_{-r_p}^{r_p} -x \frac{N_{r_p}}{\pi} f(x) e^{-p_1 N_{r_d} p_2} \\
&\cdot \int_{-\sqrt{r_p^2-x^2}}^{\sqrt{r_p^2-x^2}} \frac{1}{2\sqrt{r_p^2-x^2}} \tilde{\Pi}(\sqrt{x^2+y^2}) dy dx \\
=&\int_{-r_p}^{r_p} -x \frac{N_{r_p}}{\pi} e^{-\frac{N_{r_p}}{\pi} q(\frac{x}{r_p})} q'(\frac{x}{r_p}) \frac{1}{r_p} e^{-p_1 N_{r_d} p_2} \\
&\cdot \int_{-\sqrt{r_p^2-x^2}}^{\sqrt{r_p^2-x^2}} \frac{1}{2\sqrt{r_p^2-x^2}} \tilde{\Pi}(\sqrt{x^2+y^2}) dy dx \\
=&p_2 e^{-p_1 N_{r_d}} \int_{-r_p}^{r_p} -x \frac{N_{r_p}}{\pi} e^{-\frac{N_{r_p}}{\pi} q(\frac{x}{r_p})} q'(\frac{x}{r_p}) \frac{1}{r_p} \\
&\cdot \int_{-\sqrt{r_p^2-x^2}}^{\sqrt{r_p^2-x^2}} \frac{1}{2\sqrt{r_p^2-x^2}} \tilde{\Pi}(\sqrt{x^2+y^2}) dy dx
\end{aligned} \tag{3.17}
$$

The above equation complete the derivation of expected progress per node per slot. Equation (3.17) is good for computational evaluation as all network performance related parameters in this equation are in explicit expression. However, we hardly get any insight by directly looking into the complicated double integration. In Appendix B, by applying midpoint rule, we show that Equation (3.17) can be simplified to

$$
\begin{aligned}
&\mathcal{Z} \\
=&e^{-p_1 N_{r_d}} \cdot p_2 \cdot (1 - p_2 e^{-\lambda p_1 [\pi r_d^2 - 2 r_d^2 q(\frac{d}{2r_d})]}) e^{-\gamma p_2 N_{r_i}} \\
&\cdot \int_{-r_p}^{r_p} -x f(x) dx,
\end{aligned} \tag{3.18}
$$

where $0 \leq \gamma \leq 1$ is determined by the distribution of both primary and secondary network, e.g., γ is a function of λ_1, λ_2, p_1 and p_2.

Mathematically, equation (3.18) and (3.17) are equivalent. Since equation (3.18) has an undefined function γ, it is not suitable for numerical evaluation. However, (3.18) has a very interesting structure that helps the understanding of the fundamental interaction between primary and secondary networks.

The first part of (3.18), $e^{-p_1 N_{r_d}}$, is interpreted as the probability that a secondary node P finds an opportunity to transmit. It is easy to understand that as the primary network utilization increases, a secondary node's chance to transmit is decreased. Mathematically, this is reflected by $e^{-p_1 N_{r_d}}$ as a decreasing function of p_1 and λ_1.

The second part of (3.18),

$$(1 - p_2 e^{-\lambda p_1 [\pi r_d^2 - 2r_d^2 q(\frac{d}{2r_d})]}) e^{-\gamma p_2 N_{r_i}}, \tag{3.19}$$

can be interpreted as the weighted average of success probability of a secondary transmission given an opportunity detection is claimed at the transmitter. Intuitively, for a given secondary network setting, higher primary network utilization leads to less opportunity detection thus less overall secondary network traffic. For a given secondary transmission,e.g., transmission from P to Q, this means there is less interference from other secondary users. Therefore, the transmission from P to Q sees a higher probability to succeed. This is exactly the information carried by (3.19). When p_2 and λ_2 are fixed, Equation (3.19) is a increasing function of p_1 and λ_1, i.e., as the primary utilization increases, a secondary transmission sees higher chance to get through.

Given a secondary network, the two forces discussed above go in opposite directions when the primary network settings change.

In the next section, we numerically evaluate Equation (3.17) to reveal more insights on this interaction between the two forces.

It is worth to mention that if the primary network is absent or not active at all, Equation (3.17) converges to

$$\mathcal{Z} = p_2 \cdot (1 - p_2) \cdot e^{-p_2 N_{r_i}} \int_{-r_p}^{r_p} -x f(x) dx \tag{3.20}$$

which is equivalent to the classic result derived for expected progress in a stand-alone multihop radio network in [78].

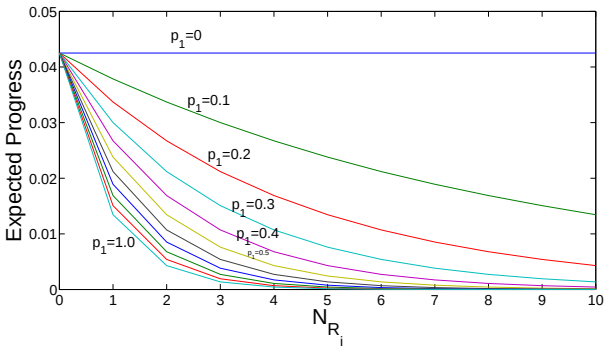


Figure 3.10: Effect of primary network density, Multi-hop network, Scenario I, Primary network radio signal power is significantly larger than secondary network radio signal power

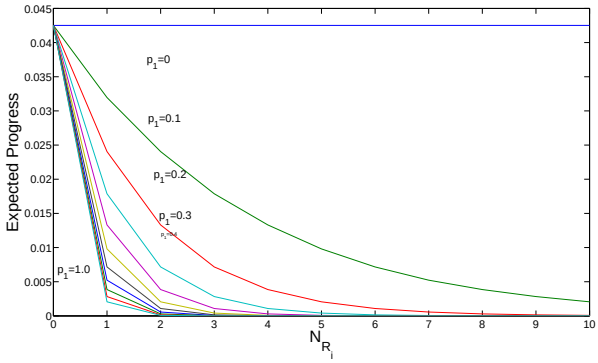


Figure 3.11: Effect of primary network density, Multi-hop network, Scenario I, Primary network radio signal power is only comparable with secondary network radio signal power

3.3.2 Numerical Results

In this section we numerically investigate the impact of primary and secondary network settings on the multihop performance of DSA networks measured as the expected progress per node per time slot. The expected progress is normalized with $\mathcal{Z} \cdot \sqrt{\lambda_2}$ as in [78] to get an dimensionless measure of the expected packet progress. We again examine two representative

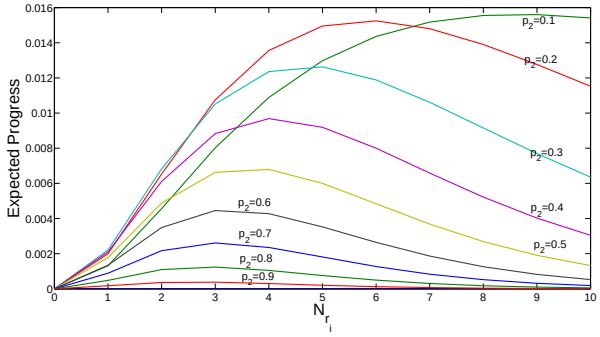


Figure 3.12: Effect of secondary network density, Multi-hop network, Scenario I, Primary network radio signal power is significantly larger than secondary network radio signal power

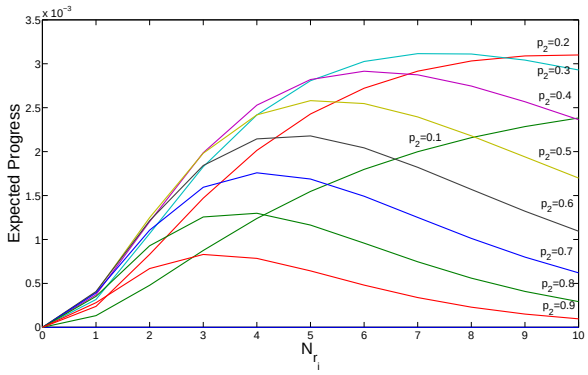


Figure 3.13: Effect of secondary network density, Multi-hop network, Scenario I, Primary network radio signal power is only comparable with secondary network radio signal power

scenarios. In Scenario I, $\frac{\tilde{r}_i}{\tilde{R}_i} = \frac{1}{10}$. In Scenario II, $\frac{\tilde{r}_i}{\tilde{R}_i} = \frac{1}{1}$.

To examine the impact of primary network settings, for each pair of (p_1, λ_1), we numerically obtain the maximum expected progress achievable by any secondary network settings. The results are shown in Figure 3.10 and 3.11. Clearly, the expected progress is an decreasing function of both p_1 and λ_1. When p_1 and λ_1 goes up, the available white space (spatial/temporal) for secondary network goes down and thus the achievable secondary network performance goes down. Mathematically, we see that in both figures, $e^{-p_1 N_{r_d}}$ is the dominant function. It drives the expected progress down as p_1 and λ_1 increases despite the increase in successful rate in secondary transmission. Moreover, the $\mathcal{Z}$ curve is shaped as an

exponential function as a result of such dominance. Also, we see that when $\frac{\tilde{r}_i}{\tilde{R}_i}$ is closer to 1, secondary network performance degrades much more faster as a result of relative denser primary user presence in its interference range in such cases.

To examine the impact of secondary network settings, we fix the primary network setting as in previous section and change p_2 and λ_2 accordingly. The results are shown in Figure 3.12 and Figure 3.13 for $\frac{\tilde{r}_i}{\tilde{R}_i} = \frac{1}{10}$ and $\frac{\tilde{r}_i}{\tilde{R}_i} = \frac{1}{1}$ separately. The curves are not monotonic. In Fig. 3.13, maximum performance is obtained around $p_2 = 0.2$ and $N_{r_i} = 8$. In Fig. 3.12, maximum performance is obtained around $p_2 = 0.1$ and $N_{r_i} = 8.5$. Note these maximum points are different from the optimal points for one-hop performance due to their different performance goal.

3.4 Appendix

Derivation of r_d

We first discuss the case where $\tilde{R}_i \geq \tilde{r}_i$ and $\tilde{R}_i \geq r_i$. Let the distance between the secondary user P and its intended receiver Q be d. We know that $d \leq r_p \leq r_i$ and $d \leq r_p \leq \tilde{r}_i$. Then the minimum detection range to guarantee no primary interference at Q is $\tilde{R}_i + d$ as shown in Figure 3.1. Apparently, if $r_d \geq \tilde{R}_i + d$ then the detection disk of P will always cover the disk of radius $\tilde{R}_i$ centered at Q (called primary to secondary interference disk) and the disk of radius $\tilde{r}_i + R_p$ centered at P (called extended secondary to primary interference disk). Since a detection at P means no active primary user in P's detection disk, there is no active primary user in Q's primary interference disk either, i.e., Q is guaranteed to be interference free. Also since there is no primary transmitter in the extended secondary interference disk, there is zero-probability that any primary receiver to be located in the disk of radius $\tilde{r}_i$ centered at P (called secondary to primary interference disk). Otherwise there will be at least one primary transmitter in the *extended secondary to primary interference disk* which violates the assumption of no active primary transmitter in P's detection disk. On the other

hand, if r_d is less than $\tilde{R}_i + d$, the primary interference disk of Q is not fully covered by P's detection disk. The detection at P provides no information in this uncovered region. There is a non-zero probability that some primary user may actually transmit in this uncovered region, i.e., there is no guarantee at the receiver Q even we have a detection at P. Since $d \in [0, r_p]$, we have in general $r_d = \tilde{R}_i + r_p$ as the unconditional detection range that always guarantee interference-free reception at the intended receiver when there is a detection at the secondary transmitter.

Similarly we can show that when $\tilde{R}_i \leq \tilde{r}_i$ and $\tilde{R}_i \leq r_i$, we have $r_d \geq \tilde{r}_i + R_p$ to be the desired detection range that always guarantee that an actual opportunity exists when a detection is claimed. To summarize, we have $r_d = \max\{r_p + \tilde{R}_i, \tilde{r}_i + R_p\}$.

Derivation of Equation (3.18)

We have

$$\begin{aligned} &\tilde{\Pi}_{min} \cdot \int_{-r_p}^{r_p} x \cdot \text{prob}\ \{\tilde{c}_1\} \cdot \text{prob}\ \{\tilde{c}_2\} \cdot \text{prob}\ \{\tilde{c}_3\} dx \\ &\quad \leq \mathcal{Z} \leq \\ &\tilde{\Pi}_{max} \cdot \int_{-r_p}^{r_p} x \cdot \text{prob}\ \{\tilde{c}_1\} \cdot \text{prob}\ \{\tilde{c}_2\} \cdot \text{prob}\ \{\tilde{c}_3\} dx \end{aligned}$$

where

$$\tilde{\Pi}_{min} = \min_{x,y} \tilde{\Pi}(\sqrt{x^2 + y^2})$$

and

$$\tilde{\Pi}_{max} = \max_{x,y} \tilde{\Pi}(\sqrt{x^2 + y^2})$$

Thus we have

$$\begin{aligned} &\mathcal{Z} \\ =&\tilde{\Pi}_{x_\xi, y_\xi} \cdot \int_{-r_p}^{r_p} x \cdot \text{prob}\ \{\tilde{c}_1\} \cdot \text{prob}\ \{\tilde{c}_2\} \cdot \text{prob}\ \{\tilde{c}_3\} dx, \end{aligned}$$

where $\Pi_{x_\xi, y_\xi} \in [\Pi_{max}, \Pi_{min}]$.

By substitute prob $\{\tilde{c}_1\}$, prob $\{\tilde{c}_2\}$ and prob $\{\tilde{c}_3\}$, we have

$$\begin{aligned}&\mathcal{Z}\\=&\tilde{\Pi}_{x_\xi,y_\xi}\cdot\int_{-r_p}^{r_p}x\cdot\text{prob }\{\tilde{c}_1\}\cdot\text{prob }\{\tilde{c}_2\}\cdot\text{prob }\{\tilde{c}_3\}dx\\=&e^{-p_1N_{r_d}}\cdot p_2\cdot\tilde{\Pi}_{x_\xi,y_\xi}\cdot\int_{-r_p}^{r_p}-xf(x)dx\end{aligned}$$

By ignoring the effect of area A on the interference condition of Q as in [78], we can express $\tilde{\Pi}_{x_\xi,y_\xi}$ as

$$\begin{aligned}&\tilde{\Pi}_{x_\xi,y_\xi}\\=&(1-p_2e^{-\lambda p_1[\pi r_d^2-2r_d^2q(\frac{d}{2r_d})]})e^{-p_2\gamma N_{r_i}},\end{aligned}$$

where $0\leq\gamma\leq1$ is determined by solving the following equation

$$e^{-p_2\gamma N_{r_i}}=\sum_{k=0}^{k=\inf}\frac{N_{r_i}^k}{k!}e^{N_{r_i}}(1-p_2)^{\tilde{k}}.$$

Chapter 4

Simple Linear Optimization Techniques to Improve Spectrum Efficiency in Frequency-Agile Radio Networks

In this chapter, we consider the use of Partially Overlapped Channels in Frequency-Agile radio networks.

4.1 Partially Overlapped Channels

There are two types of relationships among wireless communication channels: orthogonal and partially overlapping. In Orthogonal Channel (OC) only systems, channels have no or negligible spectrum overlap with each other and can be used simultaneously for communication without interfering each other. In Partially-Overlapped-Channel (POC) based systems, however, channels may have overlapped spectrum bands. Communication on one

channel may cause interference for communication on other channels. For example, in IEEE 802.11b/g wireless standard, the maximum number of orthogonal channels is three, i.e., channel 1, 6 and 11. When any other channel is used, it is considered a POC system where channels partially overlap with either one or two of these orthogonal channels.

Traditional multi-channel allocation algorithms [65]-[75] only use orthogonal channels. It is often unavoidable in such systems to assign neighboring links with the same channel due to the limited number of orthogonal channels. The cochannel interference can prevent these links from transmitting simultaneously. While links using POC can still interfere with each other, it is observed that the interference range between links using POC (assuming they are not on exactly the same frequency band) is often much smaller than the typical co-channel interference range. Such reduced interference range between links using POC can enable more simultaneous transmissions and potentially increase network throughput. Indeed, POC based design intentionally employ channels with partially or fully (co-channel) overlapped spectrums to increase channel diversity and potentially to increase overall network throughput by giving nodes with full access to all legitimate channels.

Recent results on POC based design from [51]-[55] have shown that POC based wireless networks outperform non-POC based wireless networks with a significant margin of up to several times. While these works show the great potential of POC networks, they mainly focus on protocol/optimization algorithm design for POC based networks. To the best of our knowledge, a comprehensive network model and an in-depth analysis of the throughput gain of using POC in wireless networks are still lacking. Insights from such models can show how and why POC should benefit wireless networks and can serve as guidelines for future POC network design. Moreover, such models can be used in practice for performance evaluation in POC network planning and optimization. The focus of this chapter is thus to address this technical void.

The main contributions of this chapter are as follows. We first construct a statistical model for general wireless networks. This model is able to quantify the impact of node density and

traffic load on the throughput gain of using POC. The model is also able to accommodate extremely small-area and scarce networks. Second, we propose a joint channel assignment, routing and link scheduling optimization model. This optimization model can be used to study the impact of more specific network settings such as topology shape, traffic pattern and radio-number per node on the performance of POC based wireless networks.

The rest of this chapter is organized as follows. In Section 4.2, we establish the propagation and interference model for POC. In Section 4.3, we present the statistical models to compute network throughput improvement ratio in POC networks. We propose a joint channel assignment, routing and link scheduling optimization model for POC networks and study POC performance in specific network settings in Section 4.4. In Section 4.5, we discuss related works. We conclude this chapter in Section 4.6.

4.2 Interference Model

We model the mutual interference among POCs using a scaling factor ε_{ij} which quantifies the amount of cross-channel interference. ε_{ij} is determined by factors such as propagation loss, channelization scheme, and coding techniques. Given ε_{ij} and using the general path loss model in [90], the received signal power P_r on channel i from a transmitter with transmission power P_t on channel j is:

$$P_r = P_t K \varepsilon_{ij} [\frac{d_0}{d}]^{\gamma}, \tag{4.1}$$

where K is a constant reflecting the combined effect of antenna gain and average channel attenuation, d_0 is a reference distance assumed to be $1-10$m indoors and $10-100$m outdoors, d is the distance between the sender and the receiver, and γ is the path loss exponent. Let CS_{th} denote the carrier-sensing threshold. Using *Eq.* (4.1), the interference range $r(i,j)$ between two nodes that are working respectively on channels i and j is:

$$r(i,j) = d_0(\varepsilon_{ij}\frac{P_t K}{CS_{th}})^{\frac{1}{\gamma}}, \tag{4.2}$$

Table 4.1: Notations for Chapter 4

C	The set of all available channels in the network
M	Number of all channels in set C
τ	Channel separation
τ_{th}	Threshold number of separations needed for two channel to be orthogonal
C_{oc}	The set of orthogonal channels, $C_{OC} \subseteq C$
N	Number of all orthogonal channels in C_{oc}
CS_{th}	Carrier-sensing threshold
$r_(i,j)$	Interference range between nodes on channel i and j
$d(v,u)$	Distance between node u and node v
$S_{int}(v)$	The set of nodes that can produce interference to node v
$p_{suc}^{POC}(v,t)$	The probability that node v successfully transmits a packet at time t in POC networks.
$p_{suc}^{OC}(v,t)$	The probability that node v successfully transmits a packet at time t in networks using only orthogonal channels.
$l(v,t)$	Payload delivered by node v at time slot t
η	Throughput Improvement ratio
ε_{ij}	Cross-channel interference scaling factor between channel i and j
$N(d,\tau)$	Number of nodes covered in interference range $r(\tau)$
A	The size of the area of the network under investigation

Table 4.2: POC testbed measurement results

channel separation τ	0	1	2	3	4	5
case 1 (i= 1)	13.26	9.08	7.59	4.69	3.21	0
case 2 (i= 6)	12.89	9.21	6.98	5.15	3.84	0

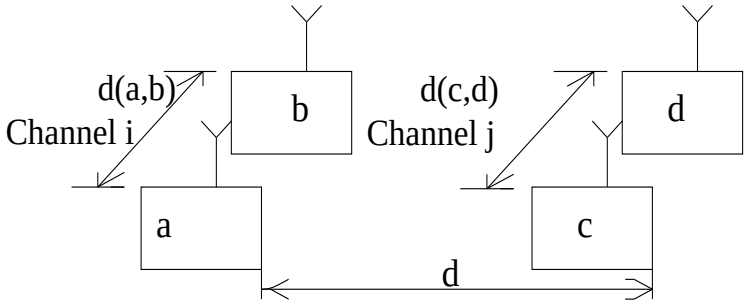
The theoretical value of ε_{ij} is determined by calculating the convolution of the power spectrum densities (PSDs) of the two interfering channels. If two POCs are only different in their center frequencies, ε_{ij} is similar to the I factor introduced in [52] and can be determined by the differences in their channel numbers. This difference in channel numbers is called *channel separation* and is denoted as $\tau = |i - j|$.

Given ε_{ij}, we can obtain $r(i, j)$ using *Eq.* (4.2). Using the concept of channel separation, $r(i, j)$ can be expressed simply as $r(|i - j|)$ or $r(\tau)$. The maximum interference range, which is the co-channel interference range, is represented by $r(0)$.

In practice, $r(i, j)$ is more often obtained through field measurements. This approach can increase model accuracy and thus is adopted in this chapter. In our field measurement, we setup two pairs of communicating nodes equipped with commodity 802.11b/g compatible radios to transmit on channel i and channel j respectively as shown in Figure 4.1. Then, we gradually increase the distance d between the two communicating pairs and record the interference range, which is the maximum distance that these two can affect each other's communications. To reduce measuring error, we did several groups of experiments and took the average. The results are summarized in Table 4.2.

4.3 Network Model

In this section, we present a statistical model to compute throughput improvement of using POCs compared to using only orthogonal channels in wireless networks. The cases of extremely small-area networks and scarce networks are handled separately in the next section.



Measured POC Interference Range

Channel Separation τ	0	1	2	3	4	5
case 1 (i=1)	13.26	9.08	7.59	4.69	3.21	0
case 2 (i=6)	12.89	9.21	6.98	5.15	3.84	0

Figure 4.1: Illustration of testbed setting. Node pair (a,b) and (c,d) are placed so that d(a,b)$\ll$ d, d(c,d)$\ll$d. Node a transmits to node b on channel i. Node c transmits to node d on channel j

4.3.1 General Model

We consider a set of transmitting nodes V distributed in an area A. The set of receiving nodes is denoted as $\hat{V}$. The receiving nodes are uniformly distributed in the area A so that every transmitting node always has at least one node in $\hat{V}$ within its transmission range. We assume a simple Listen-Before-Transmit (LBT) type medium access scheme. We set the carrier sensing threshold so that the transmitting node can sense the potential interfering nodes in the neighborhood of its intended receiver and make sure its transmission to be collision-free. We use LBT because it is the basis for many modern non-cooperative spectrum access methods. A framework for modeling network with LBT can provide fundamental insights into the behavior of non-cooperative spectrum access based networks in general [77][76]. A typical example of such system is the off-the-shelf WiFi networks with the RTS-CTS feature turned off (which is common in practice). We also assume that all nodes are equipped with radios of similar settings, such as transmission power P_t and carrier sensing

threshold CS_{th}. According to the propagation model in Section II, this assumption implies that all radios have the same set of interference ranges. Let $S_{int}(v)$ denote the set of all nodes that can produce interference to a receiving node $v \in V$. Obviously, $S_{int}(v)$ consists of all the nodes covered in v's largest interference range $r(0)$. The distance between any two nodes v and u is denoted as $d(v,u)$. The set of available channels is denoted as $C = \{1, 2, ..., M\}$. Among these M channels, N of them are orthogonal and are denoted by the set $C_{OC} \subseteq C$. Assuming the minimum channel separation between two orthogonal channels is τ_{th}, we have $N = \lceil \frac{M}{\tau_{th}} \rceil$. In the IEEE 802.11b standard, M=11, N=3, $\tau_{th} = 5$.

We denote by $p_{tr}(v)$ the probability that node $v \in V$ transmits at a time slot t. Suppose that a randomly selected node v picks channel i as its transmitting channel and one of its neighboring nodes $u \in S_{int}(v)$ picks channel j as its transmitting channel. Under the LBT MAC assumption and the carrier sensing setup, whether node u interferes with node $\hat{v}$, i.e., prevents v from transmitting to $\hat{v}$ due to carrier sensing, is determined by their distance $d(u,v)$ and channel separation $|i-j|$. More specifically, for node u's transmission on channel j to interfere node v's transmission on channel i, node u must be located in node v's interference range of radius $r(|i-j|)$ as shown in Figure 4.2.

Let $n(|i-j|)$ denote the number of nodes that are covered within radius $r(|i-j|)$. The probability that node u's transmission on channel j interferes with node v's transmission on channel i at time slot t is $(\frac{n(|i-j|)}{n(0)})$. Considering that all nodes randomly choose one of the M POCs with equal probability for their transmissions, the probability that node u chooses channel j is $\frac{1}{M}$. Hence, considering all M channels, the probability that a neighboring node interferes with node v at time slot t is

$$\sum_{j=1..M} \frac{p_{tr}(u)}{M} (\frac{n(|i-j|)}{n(0)}). \tag{4.3}$$

Let $P_{suc}^{POC}(v,t)$ denote the probability that node v successfully transmits a packet at time slot t in a POC based design. Since node v's transmission is successful only when none of

its neighbors in $S_{int}(v)$ interfere with node v, we have

$$p_{suc}^{POC}(v,t) = p_{tr}(v,t) \prod_{u \in S_{int}(v)} (1 - \sum_{j=1..M} \frac{p_{tr}(u)}{M} \frac{n(|i-j|)}{n(0)}). \quad (4.4)$$

Now consider the case where only the N orthogonal channels are used in the network. In this case, node $u \in S_{int}(v)$ interferes with node v only when node u transmits on the same channel as node v. Hence, the probability that a node $u \in S_{int}(v)$ interferes with node v is $\frac{p_{tr}(u)}{N}$. The probability of a successful transmission from node v under this orthogonal channel case, denoted as $p_{suc}^{OC}(v,t)$, is:

$$p_{suc}^{OC}(v,t) = p_{tr}(v,t) \prod_{u \in S_{int}(v)} (1 - \frac{p_{tr}(u)}{N}). \quad (4.5)$$

It is straightforward that, given Equation (4.4) and (4.5), the total expected number of simultaneous transmissions in the entire network is the summation of $p_{suc}^{POC}(v,t)$ and $p_{suc}^{OC}(v,t)$ over all nodes in time slot t for POC based design and non-POC based design respectively. Mathematically, they are expressed as $\sum_{v \in V} p_{suc}^{POC}(v,t)$ and $\sum_{v \in V} p_{suc}^{OC}(v,t)$. Assuming each successful transmission from node v at time slot t delivers $l(v,t)$ payload, the improvement of maximum throughput by using POC over a period of time T can then be expressed as η, as follows:

$$\eta = \frac{\sum_{t \in T} \sum_{v \in V} l(v,t) p_{suc}^{POC}(v,t)}{\sum_{t \in T} \sum_{v \in V} l(v,t) p_{suc}^{OC}(v,t)}. \quad (4.6)$$

Equation (4.6) provides us a formula to calculate the throughput improvement by using POCs compared to designs where only orthogonal channels are used. The model is robust and versatile in the sense that it takes into account all major factors that affect the one-hop network throughput, such as traffic load, node density, and spectrum segmentation, etc.. The use of this general formula is more clearly shown in the following sections where we apply this model in different network scenarios.

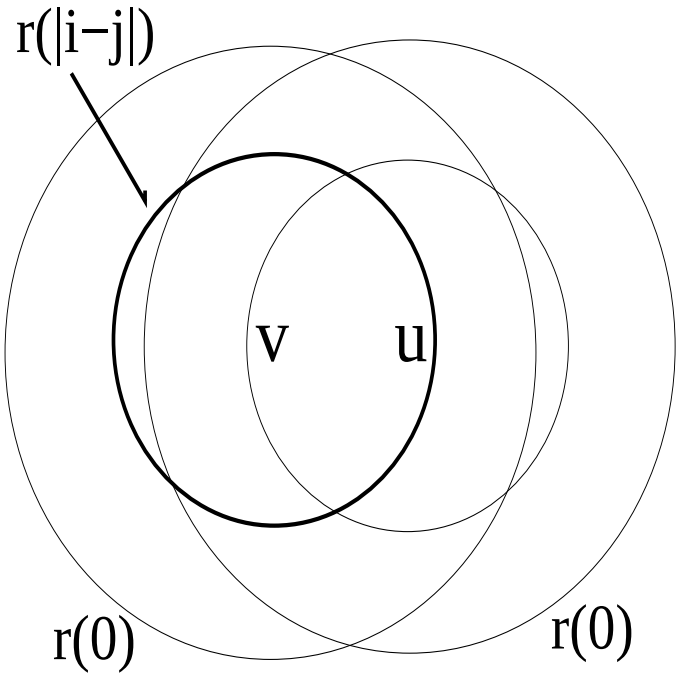


Figure 4.2: Illustration of the effect of channel selection on the interference relationship in POC networks: Node u transmitting on channel j needs to be in $r(|i-j|)$ of v to interfere with $\hat{v}$'s reception on channel i

4.3.2 Large-Scale Networks

By large-scale, we assume that the co-channel interference range $r(0)$ is significantly smaller than the network dimension. More precisely, $r(0) << \sqrt{A}$, where A denotes the area of the network under investigation. With this assumption, we can omit the impact of the size of the distribution area A and its boundary effect in our analysis without harming the validity of the analytical results. Many practical networks, such as large scale sensor networks and mesh networks, loosely bear this characteristic.

Without loss of generality, we assume that all nodes use similar radio settings and traffic patterns. Communication and interference ranges are the same for all nodes and do not change in the period of observation. Each node $v \in V$ transmits packet at any time slot t with constant probability p_{tr} and each transmission carries the same amount of payload l.

In such case, *Eq.* (4.6) can be reduced to a much simpler form as

$$\eta = \frac{p_{suc}^{POC}}{p_{suc}^{OC}} = \frac{(1 - \sum_{j=1..M} \frac{p_{tr}}{M} \cdot \frac{n(|i-j|)}{n(0)})^{n(0)}}{(1 - \frac{p_{tr}}{N})^{n(0)}}. \tag{4.7}$$

Eq. (4.7) indicates that instead of directly comparing the maximum throughput, we can simply compare the p_{suc}^{POC} and p_{suc}^{OC} to obtain the improvement ratio.

The assumption presented in this section and *Eq.* (4.7) can easily be extended to describe any specific traffic patterns and radio characteristics. For ease of presentation, we will use the reduced form of throughout in the remaining part of this chapter. In the remainder of this section, we derive the expression of η in string, grid and random network topologies.

Performance Models

String Topology

We assume nodes are uniformly distributed on a string with a unit distance of d between each pair of neighboring nodes. From elementary geometry, the number of nodes on the string covered by the circle centered at node v with radius $r(\tau)$ is

$$N(d, \tau) = 2\lfloor \frac{r(\tau)}{d} \rfloor + 1. \tag{4.8}$$

Consider a simple example illustrated in Figure 4.3. If r(0)=13m and d=5m, we get N(4, 0)=5.

Based on *Eq.* (4.7) and (4.8), the improvement ratio then can be rewritten as

$$\eta = \frac{p_{suc}^{POC}}{p_{suc}^{OC}} = (\frac{1 - \sum_{j=1..M} \frac{p_{tr}}{M} \cdot (\frac{2\lfloor \frac{r(|i-j|)}{d} \rfloor + 1}{2\lfloor \frac{r(0)}{d} \rfloor + 1})}{1 - \frac{p_{tr}}{N}})^{N(d,0)}. \tag{4.9}$$

Grid Topology

Applying results of the Gauss's Circle Problem [84] in a grid of area A with elementary grid length $d \ll \sqrt{A}$, the number of nodes covered by the interference range $r(\tau)$ centered at a

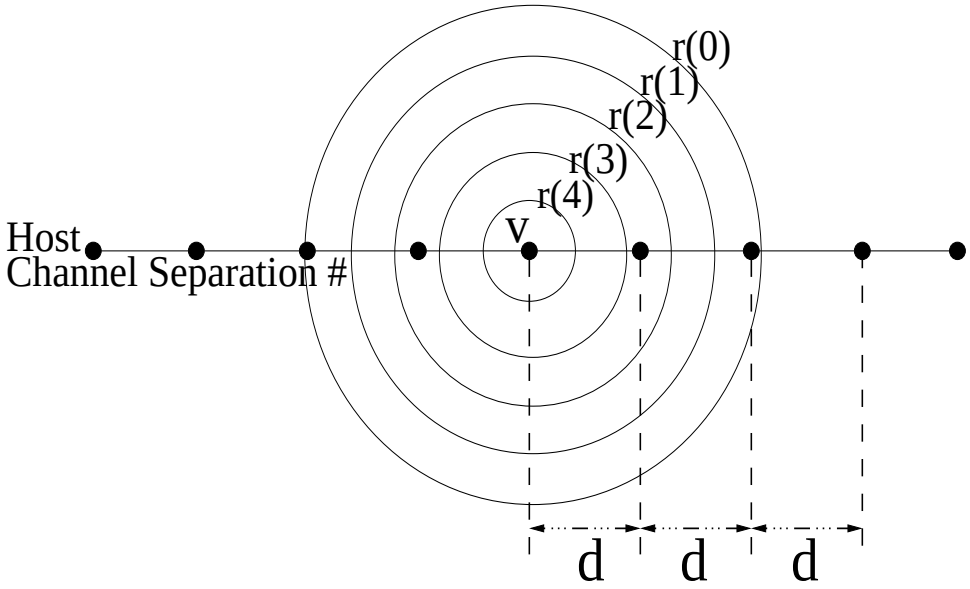


Figure 4.3: Geometric relationship between interference range and neighbor distance d. The circles represent the interference ranges corresponding to different channel separations.

node v is:

$$N(d,\tau) = 1 + 4\lfloor\frac{r(\tau)}{d}\rfloor + 4\sum_{i=1}^{\lfloor\frac{r(\tau)}{d}\rfloor}\lfloor\sqrt{(\frac{r(\tau)}{d})^2 - l^2}\rfloor. \tag{4.10}$$

Combining Equations (4.7) and (4.10), we get:

$$\eta = \left(\frac{1 - \sum_{j=1..M}\frac{p_{tr}}{M}\cdot\left(\frac{1+4\sum_{k=1}^{(\frac{d}{r(\tau)})^2}(-1)^{k-1}\lfloor(\frac{\frac{d}{r(\tau)})^2}{2k-1}\rfloor}{N(d,0)}\right)}{1-\frac{p_{tr}}{N}}\right)^{N(d,0)}. \tag{4.11}$$

Randomly distributed topology

Let us relax the constraints on node placement and assume that nodes are placed uniformly at random on a two-dimensional infinite plane with node density ρ. The number of interfering neighbors covered in the interference range $r(\tau)$ is $n(\tau) = \rho\pi(r(\tau))^2$. Therefore, based on *Eq.* (4.7),

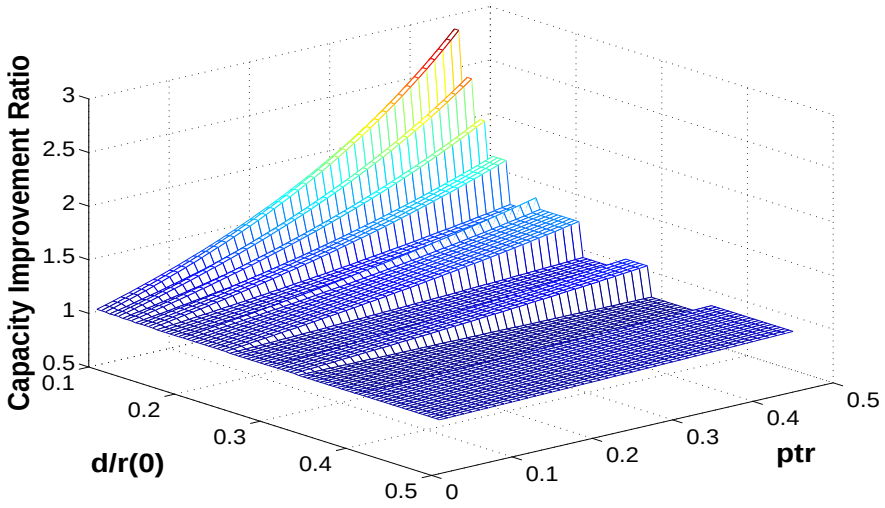


Figure 4.4: Effect of network density and traffic load on network capacity improvement ratio in string network

$$\eta = \left(\frac{1 - \sum_{j=1..M} \frac{p_{tr}}{M} \cdot \left(\frac{r_{cs}(|i-j|)}{r_{cs}(0)}\right)^2}{1 - \frac{p_{tr}}{N}}\right)^{\rho\pi(r_{cs}(0))^2}. \tag{4.12}$$

Note that transmission probability p_{tr} links to traffic load of the network, ρ and d link to node density. In Section V, we evaluate these models and show how the throughput gain η changes with different node density and traffic load.

Numerical Results

We put $M = 11$, $N = 3$ and the measured value in Table 4.2 into Equations (4.9),(4.11) and (4.12). For example, in *Eq.* (4.12), with the aforementioned input value, we get

$$\eta = \left(\frac{1 - \frac{2.8 p_{tr}}{11}}{1 - \frac{p_{tr}}{3}}\right)^{\rho\pi(r_{cs}(0))^2}. \tag{4.13}$$

We then evaluate this equation against p_{tr} and node density ρ. The results is shown in Figure 4.6. Since $\frac{1-\frac{2.8P_{tr}}{11}}{1-\frac{P_{tr}}{3}} > 1$ and $\rho\pi(r_{cs}(0))^2 > 1$, increasing either ρ or p_{tr} leads to an increase

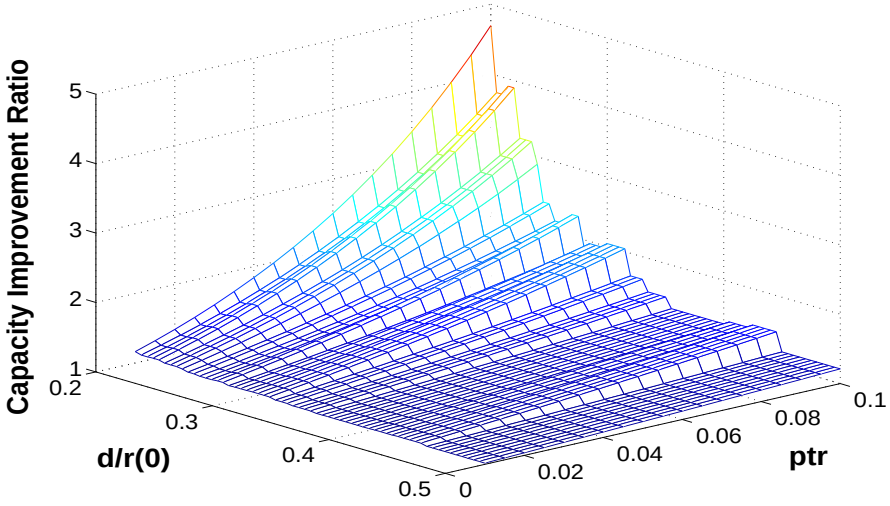


Figure 4.5: Effect of network density and traffic load on network capacity improvement ratio in grid network

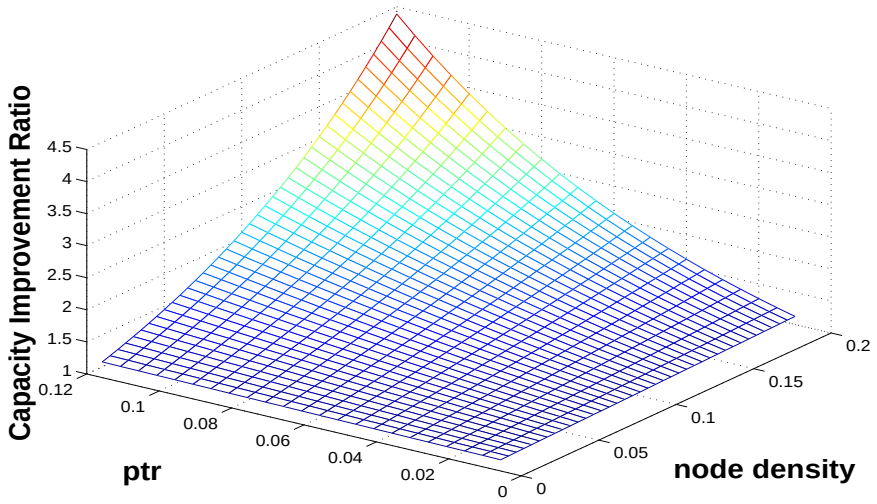


Figure 4.6: Effect of network density and traffic load on network capacity improvement ratio in random network

in η. Since p_{tr} links to traffic load and ρ links to node density, the results show that POCs works better in denser networks with heavier traffic load. Similarly, based on Equations

(4.9), (4.11), the values of η corresponding to different p_{tr} and node density in String and Grid topologies are depicted in Figures 4.4 and 4.5. Since smaller d corresponds to larger density, Figures 4.4 and 4.5 show clearly how larger density and higher traffic load favor the use of POCs. The discontinuity in these two figures is attributed to the discreteness of the number of nodes that are covered in different interference ranges.

4.3.3 Small Networks and Scarce Networks

Intuitively, if the network is too scarce, nodes are out of interference range of each other and can transmit concurrently on whatever channels they use. Also note that only orthogonal channels can transmit simultaneously on the same link. This limits the maximum number of simultaneous transmissions on one link to be $\lceil \frac{M}{\tau_{th}} \rceil$, i.e., the number of available orthogonal channels. Since the use of POC does not increase the number of orthogonal channels, it does not increase network throughput in the scarce case. Thus, $\eta \approx 1$. The scarce case is well recorded in Figures 4, 5 and 6. As shown in these figures, when node density goes really low, η gradually drops to 1.

On the other hand, if the network area A is too small, to the degree that the distance between any pair of nodes is smaller than $\min\{r(\tau)\}$, none of the partially overlapping channels can transmit simultaneously. In such case, no benefit can be obtained from using POC either. Therefore, $\eta \approx 1$.

Different from the scarce case, the assumption that $r(0) \ll \sqrt{A}$ does not necessarily hold true in extremely small-area networks. The boundary effect can no longer be ignored. Nodes closer to the boundary may experience less interference because part of their interference ranges may cover areas outside of A where there is no node. As a result, previous models for large POC networks cannot be used. Instead, we use the following model for extremely small-area networks.

Given a set of node V, we constrain the node to be uniformly distributed in a circle of radius

r centered at a node $v \in V$. If network area is smaller than any $\pi \cdot r(\tau)^2$, nodes fall in the interference range $r(\tau)$ with probability 1. Otherwise, this probability is $(\frac{r(\tau)}{r})^2$ where $r = \sqrt{A}$ is an indicator that are positively proportional to the size of the network area. Considering the fact that a node randomly chooses any particular channel j with probability $\frac{1}{M}$, the probability that a node u interferes with transmitting node v is

$$\sum_{j=1..M} \frac{p_{tr}}{M} \cdot (\min\{\frac{r_{cs}(|i-j|)}{r}, 1\})^2, \tag{4.14}$$

where p_{tr} is the probability that a node v tries to transmit a packet in a random time slot. Following a similar analysis in this section, the probability that node v successfully captures the channel for its immediate transmission is

$$p_{suc}^{POC} = p_{tr} \cdot (1 - \sum_{j=1..M} \frac{p_{tr}}{M} \cdot (\min\{\frac{r(|i-j|)}{r}, 1\})^2)^n. \tag{4.15}$$

For the orthogonal case, we have

$$p_{suc}^{OC} = p_{tr} \cdot (1 - \frac{p_{tr}}{N} \cdot \min\{(\frac{r(0)}{r})^2, 1\})^n. \tag{4.16}$$

Hence, the throughput improvement ratio is

$$\eta = (\frac{1 - \sum_{j=1..M} \frac{p_{tr}}{M} \cdot (\min\{\frac{r(|i-j|)}{r}, 1\})^2}{1 - \frac{p_{tr}}{N} \cdot \min\{(\frac{r(0)}{r})^2, 1\}})^n. \tag{4.17}$$

By evaluating *Eq.* (4.17), we get the following results depicted in Figure 4.8. The maximum improvement is obtained when $\frac{r}{r(0)} = 1$. As expected, when the the network is really scarce, $\frac{r}{r(0)}$ turns to be very large and there is almost no throughput improvement. The interesting part is that when $\frac{r}{r(0)}$ drops below 1, the improvement ratio decreases as $\frac{r}{r(0)}$ decreases. When $\frac{r}{r(0)}$ drops below some point around 0, the improvement ratio can even drop below 1. This is because communications using POCs are significantly constrained in a small-area network. The benefits of using POCs do not add much to channel diversity but only add to an increased level of interference. Consider an $IEEE802.11b/g$ based network where all nodes are within the smallest interference range $r(\tau_{th})$ of a specific node v. The network is so small that no parallel communication using partially overlapping channel is possible

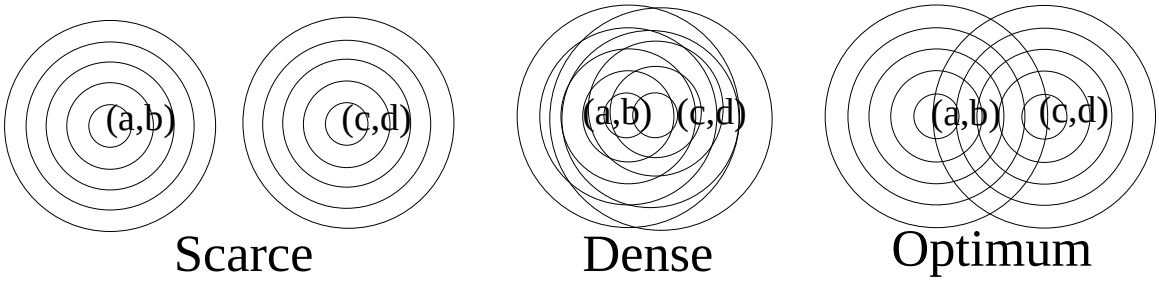


Figure 4.7: POCs cannot improve network throughput if the network is too scarce or too small. Optimum point/range is achieved when r/r(0)=1

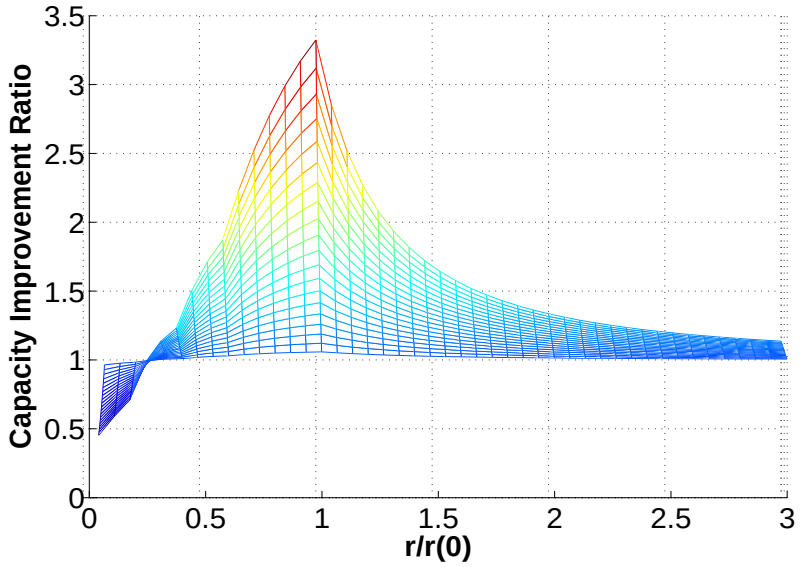


Figure 4.8: r is the radius of the constraining circle, r(0) is the co-channel interference range. Maximum improvement is obtained when r/r(0)=1.

without collision. If we only use orthogonal channels 1, 6 and 11, the probability that a nodes interferes with node v is 1/3. However, if POCs are used, since nodes pick any channel with equal probability, this probability is at least $\frac{11-3}{11} = \frac{8}{11}$, which is much higher. Hence, in practice,for small-scale networks, POCs should be avoided.

4.4 Joint Flow Routing and Link Scheduling Optimization

In this section, we first present a joint channel assignment, flow routing and link scheduling optimization model for POC networks. We then numerically evaluate our optimization model in a few representative network scenarios to obtain insights about throughput gain of POC networks as well as to validate our findings in previous sections.

4.4.1 Network Model and Constraints

We consider a network with a node set V and a flow set Q. Each data flow $q \in Q$ is represented by a source-destination pair $(src(q), dst(q))$. Each pair has a traffic load of $r(q)$ over time period T. The set of all nodes that can directly communicate with a node v is denoted by $S_c(v)$. The set of nodes that can interfere with node v's reception if they transmit on a channel within channel distance τ of the channel used by v is denoted as $S_{int}^{r(\tau)}(v)$. Other notations are the same as in Section III.

We use a binary variable $X_i^t(v, u, q)$ to indicate the spectrum usage of a wireless link from node v to node u as

$$X_i^t(v,u,q) = \begin{cases} 1 & \text{if } v \text{ transmits to } u \text{ for flow } q \\ & \text{on channel i in slot t} \\ 0 & \text{otherwise} \end{cases}$$

The network constraints are summarized as follows.

Load Balancing Constraint: We assume that in each time slot t, each wireless link can deliver one unit of data traffic. For each node, let $g(v, q)$ represent the number of traffic unit generated by node v for flow q. A node can be the sources or destinations of multiple data flows, or it can be both source and destination for different flows simultaneously. For nodes

that only relay traffic,

$$g(v,q) = 0, \forall q \in Q, v \in V.$$

The load balancing constraint thus can be represented as

$$\sum_{u \in S_c(v)} \sum_{i=1,...,M} (X_i^t(v,u,q) + X_i^t(u,v,q)) + g(v,q) = 0, \quad \forall q \in Q, v \in V \tag{4.18}$$

Orthogonality Constraints: A critical observation is that two partially overlapped channels on the same node should not be allowed to transmit/receive simultaneously due to strong self-interference. This observation results in the following three constraints.

Firstly, an active channel i on a node v prevents any partially overlapped channel to be used simultaneously on node v, i.e.,

$$\sum_{u \in S_c(v)} \sum_{q \in Q} \sum_{j \in POC(i)} (X_j^t(v,u,q) + X_j^t(u,v,q)) \leq 1, i = 1, ..., M, \forall v \in V, \forall t \in T. \tag{4.19}$$

where

$$POC(i) = \{\max\{1, i - \tau_{th} + 1\}, ..., \min\{M, i + \tau_{th} - 1\}\}$$

denotes the set of channels overlapping with channel i. τ_{th} is the minimum channel distance between two orthogonal channels.

Secondly, the maximum number of channels that can work simultaneously on node v should be no larger than the maximum number of non-overlapping channels available to node v. Mathematically, this is

$$\sum_{u \in S_c(v)} \sum_{q \in Q} \sum_{i=1,...,M} (X_i^t(v,u,q) + X_i^t(u,v,q)) \leq N, \forall v \in V, \forall t \in T. \tag{4.20}$$

Third, the number of active channels at a node v should not be larger than the number of available channels,

$$\sum_{u \in S_c(v)} \sum_{q \in Q} \sum_{i=1,...,M} (X_j^t(v,u,q) + X_j^t(u,v,q)) \leq M, \forall v \in V, \forall t \in T. \tag{4.21}$$

In POC based design, since $M \geq N$ is always true, Constraint (4.21) is inherently contained by Constraint (4.20). Furthermore `Lemma 1` shows that Constraint (4.20) is itself contained by Constraint (4.19) and hence is not included in the final formulation.

`Lemma 1:` For POCs based link-channel scheduling problem stated above, Constraint (4.19) is a sufficient condition for Constraint (4.20)

`Proof:` The proof of `lemma 1` is presented in the appendix.

Radio Constraint: Radio constraint states that no node can be assigned more simultaneous transmissions or receptions than its maximum number of radios at any time slot. This leads to the following constraint

$$\sum_{u \in S_c(v)} \sum_{q \in Q} \sum_{i=1,...,M} (X_i^t(v,u,q) + X_i^t(u,v,q)) \leq \vartheta(v), \quad \forall v \in V, \forall t \in T, \tag{4.22}$$

where $\vartheta(v)$ is the number of radios available on node v.

Interference Constraint: Under protocol interference model, a successful transmission to a node v on channel i requires all nodes in $S_{int}^{r(0)}(v)$ to keep silent on channel i, all nodes in $S_{int}^{r(1)}(v)$ to keep silent on channel $i-1$ and $i+1$, and so forth. This can be expressed as the following constraint

$$\sum_{q \in Q} \sum_{a \in S_{int}^{r(\tau)}(u)} \sum_{b \in S_c(a)} \sum^{j \in \{\{i-\tau, i+\tau\} \bigcap \{1,...,M\}\}} X_j^t(a,b,q) + \sum_{q \in Q} \sum_{v \in S_c(u)} \sum_{j \in POC(i)} X_j^t(v,u,q) \leq 1, \forall v \in V, i = 1, ..., M. \tag{4.23}$$

4.4.2 Optimization Model

Define λ as the fraction of flow demand that could be delivered over the time period T for any traffic load $r(q)$. The network throughput maximization problem can then be formulated as

$$\max \lambda$$

$$s.t. \quad \sum_{u \in S_c(v)} \sum_{i=1,...,M} (X_i^t(v,u,q) + X_i^t(u,v,q)) + \lambda g(v,q) = 0, \forall q \in Q, \forall v \in V.$$

$$\sum_{u \in S_c(v)} \sum_{q \in Q} \sum_{j \in POC(i)} (X_j^t(v,u,q) + X_j^t(u,v,q)) \leq 1, \forall t \in T, \forall v \in V, i = 1, ..., M.$$

$$\sum_{u \in S_c(v)} \sum_{q \in Q} \sum_{i=1,...,M} (X_i^t(v,u,q) + X_i^t(u,v,q)) \leq \vartheta(v), \quad \forall v \in V, \forall t \in T$$

$$\sum_{q \in Q} \sum_{a \in S_{int}^{r(\tau)}(u)} \sum_{b \in S_c(a)} \sum^{j \in \{\{i-\tau, i+\tau\} \bigcap \{1,...,M\}\}} X_j^t(a,b,q) + \sum_{q \in Q} \sum_{v \in S_c(u)} \sum_{j \in POC(i)} X_j^t(v,u,q) \leq 1, i = 1, ..., M, \forall v \in V.$$

This formulation simply computes the maximum network throughput over a period T.

We then relax the ILP (Integer Linear Programming) to a Linear Programming (LP) problem as in [24] [25]. We denote the fraction of time that v is active on channel i during a given time period T by

$$x_i^T(v,u,q) = \sum_{t \in T} X_i^t(v,u,q)/T, \quad 0 \leq x_i^T(v) \leq 1.$$

Note that If T is large enough, $x_i^T(v)$ can be viewed as a continuous variable. By summing up both sides of *Eq.* (4.19) (4.22) (4.23) over all time slots $t \in T$ and then dividing the summation by T, we get LP relaxation of the original ILP problem based on $x_i^T(v,u,q)$.

$$max\lambda$$

Subject to

$$\sum_{u\in S_c(v)}\sum_{i=1,...,M}(x_i^T(v,u,q)+x_i^T(u,v,q))+\lambda g(v,q)=0,\forall q\in Q,\forall v\in V$$

$$\sum_{u\in S_c(v)}\sum_{q\in Q}\sum_{j\in POC(i)}(x_j^T(v,u,q)+x_j^T(u,v,q))\leq 1,\forall v\in V,i=1,...,M$$

$$\sum_{u\in S_c(v)}\sum_{q\in Q}\sum_{i=1,...,M}(x_i^T(v,u,q)+x_i^T(u,v,q))\leq \vartheta(v),\forall v$$

$$\sum_{q\in Q}\sum_{a\in S_{int}^{r(\tau)}(u)}\sum_{b\in S_c(a)}\sum^{j\in\{\{i-\tau,i+\tau\}\bigcap\{1,...,M\}\}}x_j^t(a,b,q)$$
$$+\sum_{q\in Q}\sum_{v\in S_c(u)}\sum_{j\in POC(i)}x_j^t(v,u,q)\leq 1,\forall v\in V,i=1,...,M.$$

$$x_i^T(v,u,q)\geq 0,i=1,...,M,\forall u,\forall v\in V,\forall q\in Q$$

The maximum throughput using only orthogonal channels can be obtained by simply adding the following constraint to the above formulation,

$$x_i^T(v)=0,\forall i\notin C_{OC} \tag{4.24}$$

where C_{OC} is the set of orthogonal channels.

Let c_{poc} and c_{oc} be the optimal solutions for POC based and OC-based formulations. The throughput improvement ratio is then

$$\eta=\frac{c_{poc}}{c_{oc}} \tag{4.25}$$

The solution of the relaxed formulation can be viewed as the long-term average percentage of traffic that can be delivered over a period T for the traffic load Q. This solution serves well as a performance measure for wireless network in term of throughput. Note that this model can also be extended to capture more realistic factors, such as fairness, varying channel states, power control, different modulation schemes and error correction coding. To consider different fairness criteria, appropriate utility functions can be used as the objective function in the formulation. To consider power control, modulation schemes and error correction coding, we can differentiate wireless links by their throughput and differentiate interference ranges by their interference power level according to assigned transmission power and coding/modulation schemes.

4.4.3 Numerical Results

In the following, we numerically evaluate the optimization model in some representative network scenarios and discuss the impact of per-node radio number, flow patterns and average node distance on the throughput gain. The network topologies and traffic patterns discussed in this section are shown in Figures 4.9-4.13. We assume the spectrum to be divided into 11 channels with $N = 3$ orthogonal channels (e.g. $1, 6, 11$). Each wireless link can transmit one unit of traffic over a period of time T if fully utilized. Other settings are similar as in Section III.

General Ad Hoc Network

In the network as shown in Figure 4.9, there are 16 nodes and 4 data flows. Each node is $50m$ away from its closest neighbors in the grid. Under the protocol interference model, the transmission range is set as $55m$ and the co-channel interference range is set as $100m$. The POC scaling factor ϵ is obtained from *Table II* from the testbed measurement.

Figure 4.14 shows the impact of per-node radio number on the throughput gain of using POC.

We can see from the plot that as the number of radio increases, the data delivery ratios for both POC and non-POC based networks increase. However, since the data delivery ratio for POC network grows faster, the improvement ratio η increases. The observation is that when the number of radio-per-node is small, i.e., smaller than the number of orthogonal channels, the major constraint on network throughput is the limited number of radios, not the number of channels. Hence, there is less throughput gain by using POC.

Figure 4.15 shows that as the traffic load increases in the network, the throughput improvement ratio η increases. This result is similar to our theoretical prediction in Section III. It is straightforward that when the traffic is low, the orthogonal channels are enough to deliver all the traffic. Only when the traffic is heavy and all the orthogonal channels are used, partially overlapped channels become helpful.

In the network as shown in Figure 4.10, we keep the radio-per-node as three and keep the traffic rate as 0.8 per node. The destinations of flows are set so that the paths of flows are parallel. The results are shown in Figure4.16. Compared with Figure 4.15, Figure 4.16 shows that POC based network have a much higher improvement ratio in the network with mesh-style traffic pattern than in the network with parallel traffic pattern. This is expected because a complex traffic pattern as shown in Figure 4.9 tends to have more overlapped route among flows and thus more contention for channels among links. The increased contention gives advantage to POC.

WLAN with multiple access points

In the network as shown in Figure 4.11, we set to mimic a WLAN with access points $1, 2, 3, 4$ and randomly distributed client nodes $5, 6, 7, 8$. There are 4 flows from the 4 access points to their corresponding client nodes respectively. We keep the traffic rate at three units during period T for each flow and let each node have three radios. Under the protocol interference model, the transmission range is set as $55m$ and the co-channel interference range is set as $100m$. The POC scaling factor ϵ is the same as obtained from the testbed measurement.

Figure 4.17 shows that as the distance between APs increase from $40m$ to $120m$, the throughput improvement ratio first increases from 1 to 1.28 and then quickly drops to 1. This curve has a similar shape as predicted in the statistical model as shown in Figure 4.8 and can be viewed as a snapshot of the scenario predicted in the statistical model. Note that the peak is obtained between $101m$ and $109m$. This is also well predicted by Figure 4.8 where the peak throughput gain is obtain when $\frac{r(0)}{d}$ is slightly above 1.

WLAN with a single access point

Figure 4.12 shows a WLAN with a single access point (i.e. node 1) as in most family homes. We set the communication range to be $55m$ and the interference range to be $100m$. We also set the distance from each client to access point to be around $40m$. In this setting, any two nodes can interfere with each other when they communicate on the same channel. Other settings are the same as in the multiple access points scenario as discussed in the previous section. The results are shown in Figure 4.18. The results show that as traffic load increases, the traffic delivery ratio of both POC and non-POC based network decreases. The interesting observation is that POC does not help in all cases. A similar results would be obtained for a sensor network scenario (Figure 4.13)as shown in Figure 4.19. This is surprising but also quite easy to understand. According to the orthogonality constraint in *Eq.* (4.19), the link throughput cannot be increased by using POC. The maximum throughput of a single link is only determined by the number of radio-per-node and the number of available orthogonal channels. In both of the above network scenarios, there is a single bottleneck link. In the WLAN scenario, the bottleneck is the access point. In the sensor network scenario, the bottleneck is the sink. POC cannot increase the throughput of these bottleneck links. Hence, they limit the usefulness of POC.

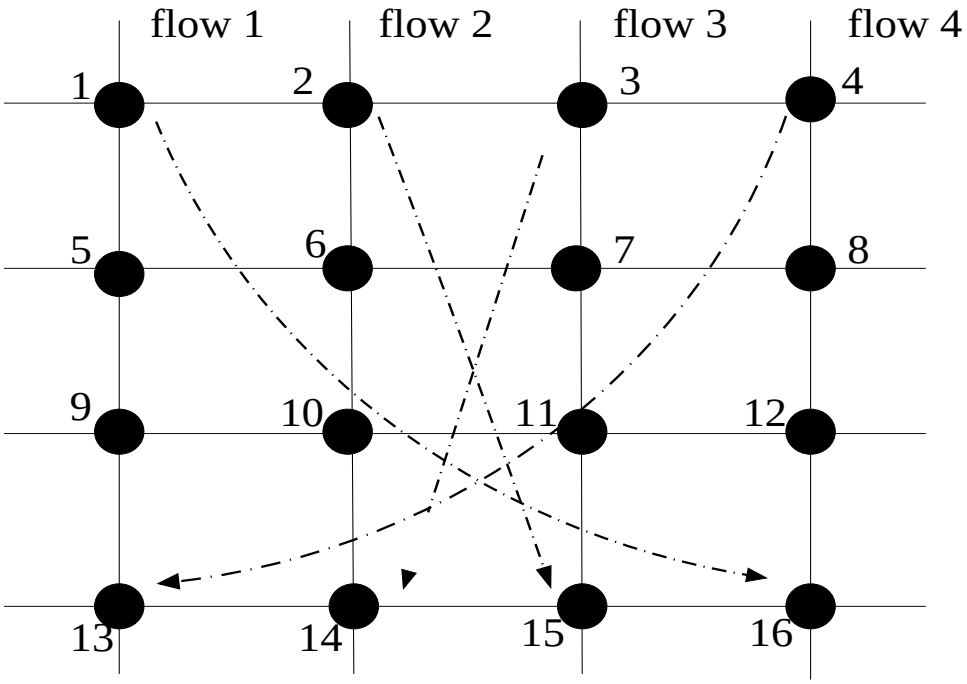


Figure 4.9: Simulated network topology and traffic: Mesh network with heavy traffic

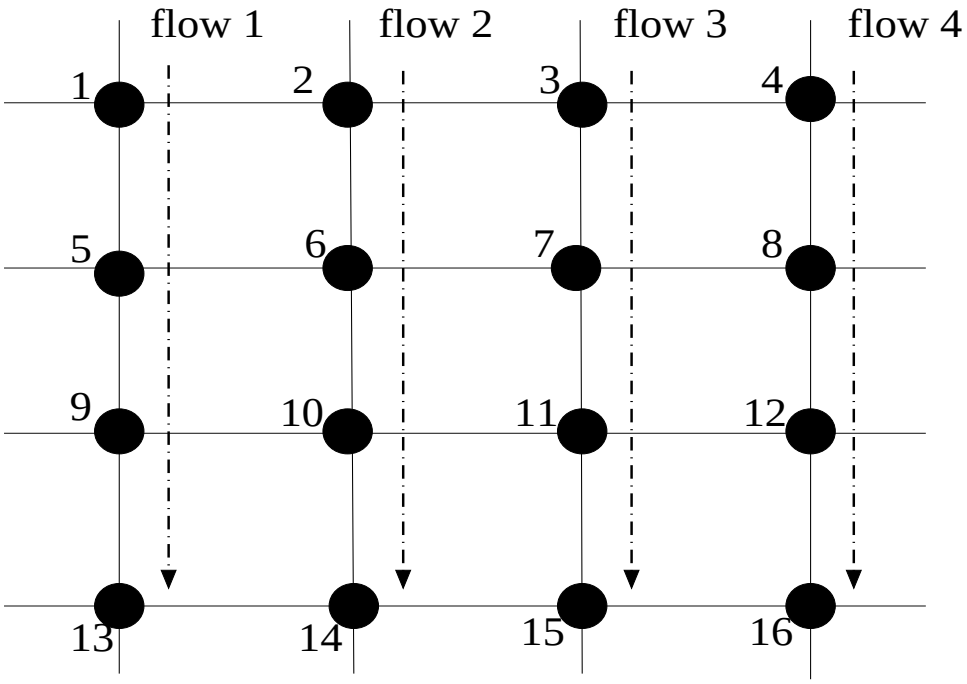


Figure 4.10: Simulated network topology and traffic: Mesh network with parallel traffic

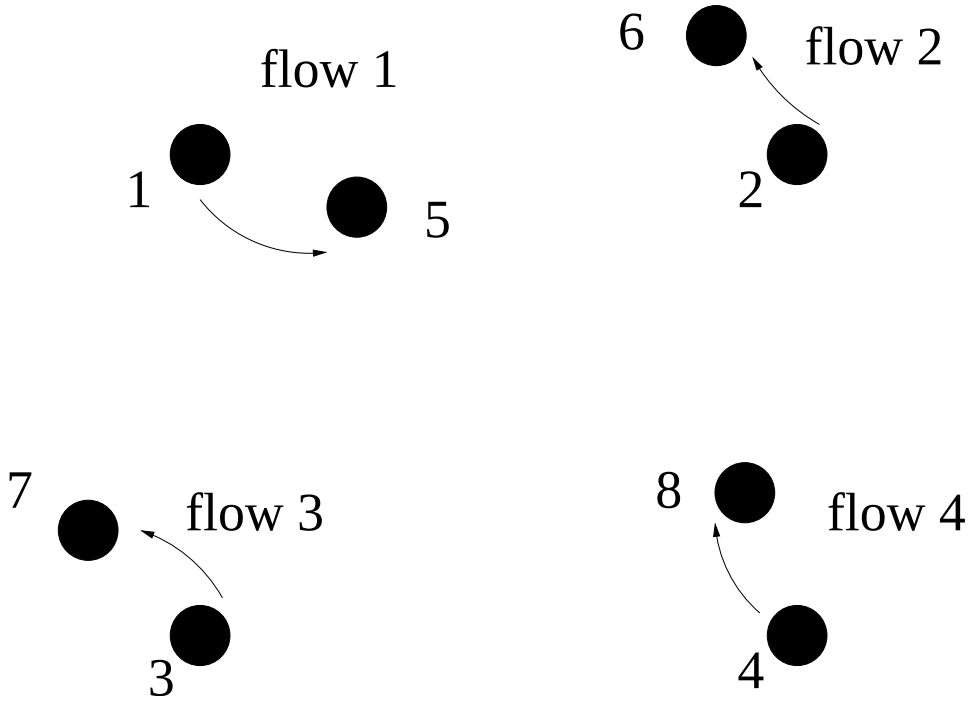


Figure 4.11: Simulated network topology and traffic: WLAN with multiple APs

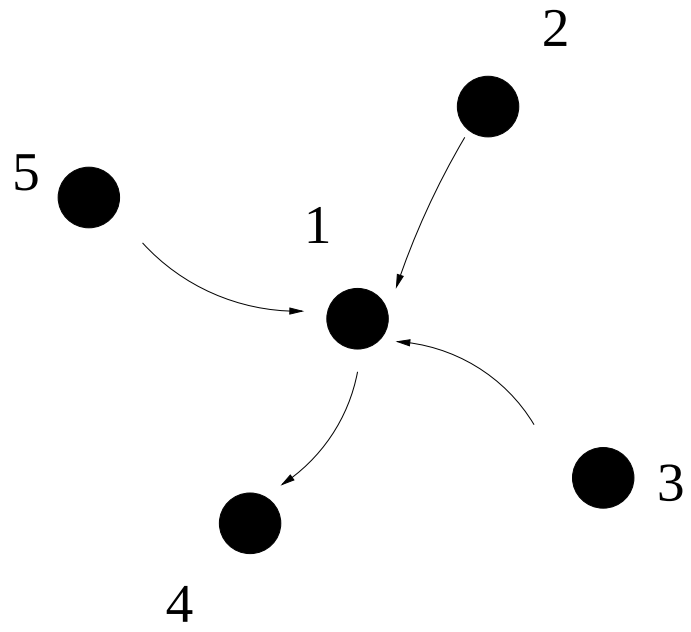


Figure 4.12: Simulated network topology and traffic: WLAN with a single AP

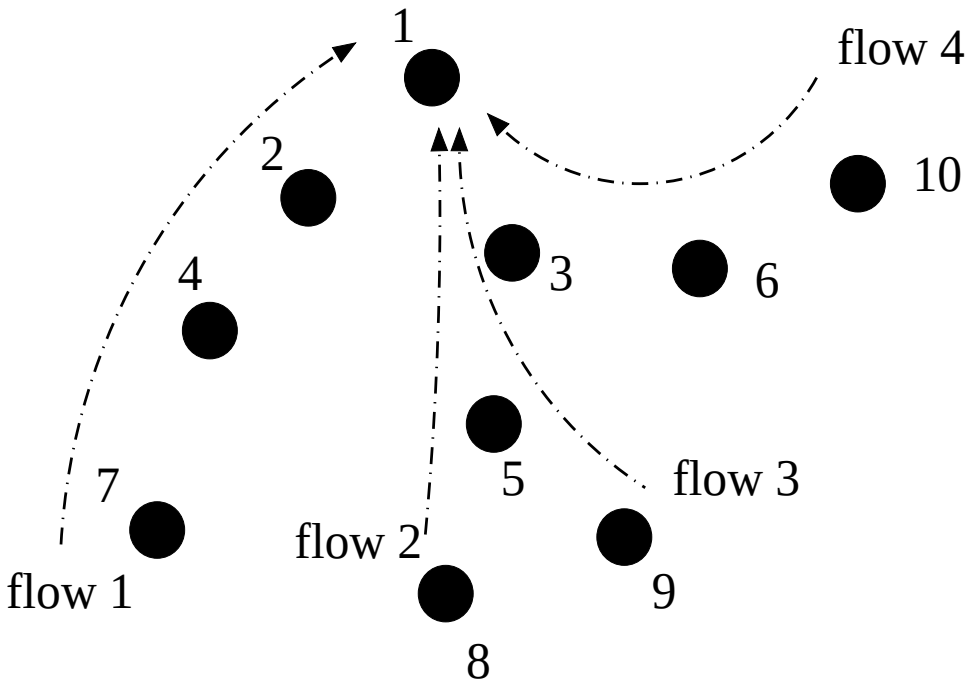


Figure 4.13: Simulated network topology and traffic: Sensor network

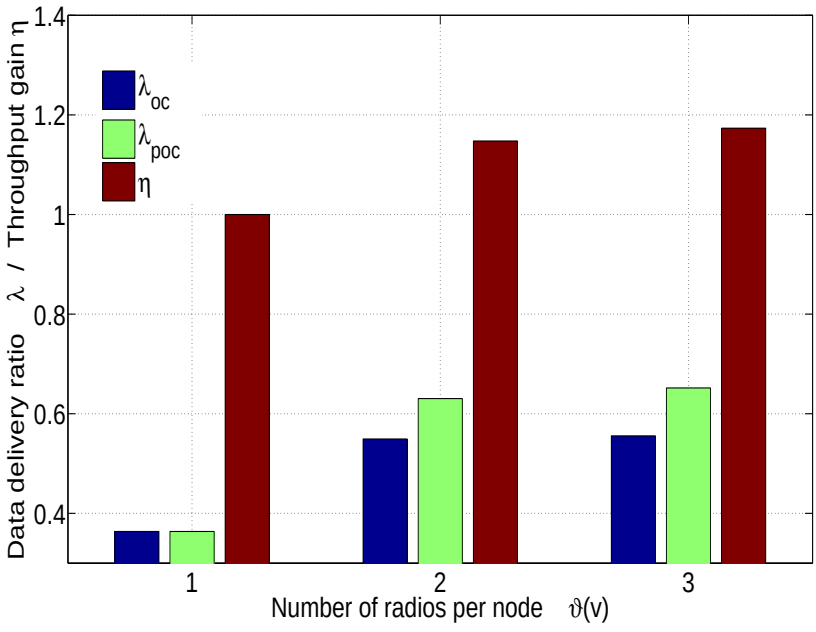


Figure 4.14: Impact of radio number per node on data delivery ratio and throughput gain

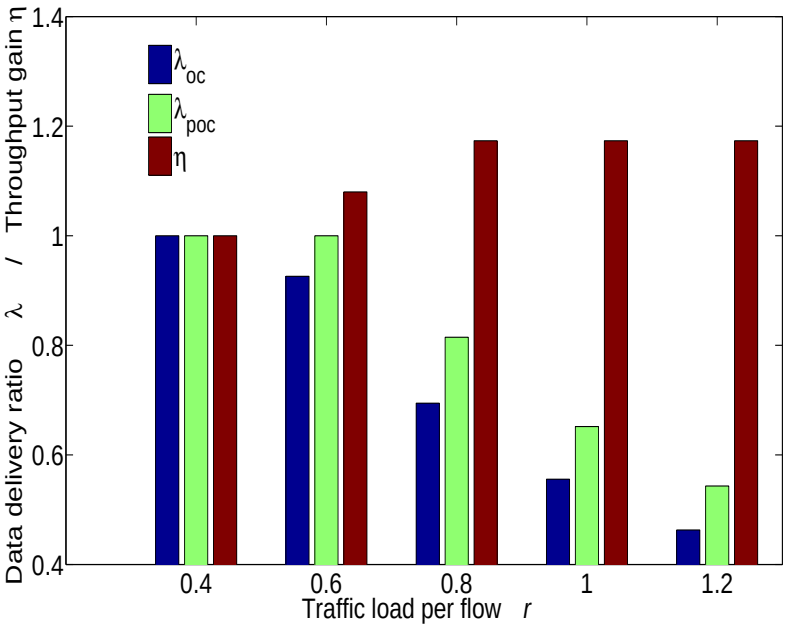


Figure 4.15: Impact of traffic load on data delivery ratio and throughput gain

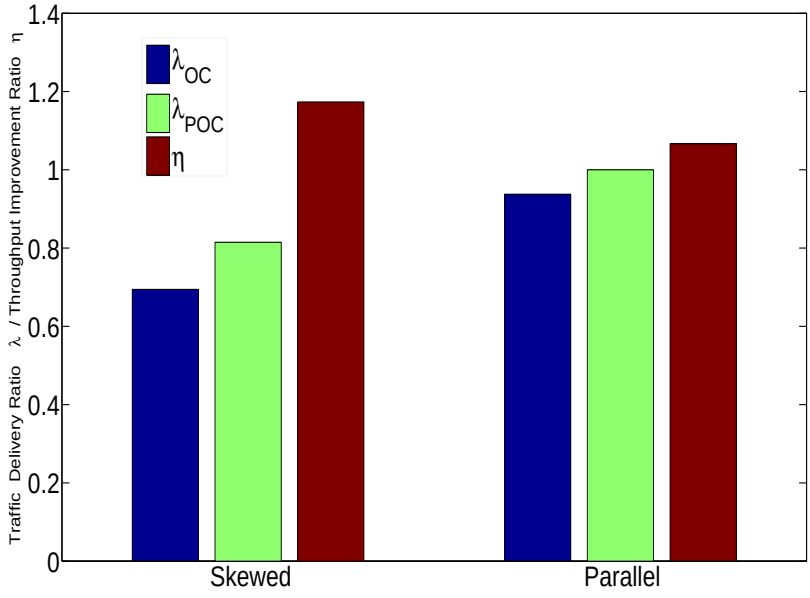


Figure 4.16: Impact of traffic pattern on data delivery ratio and throughput gain

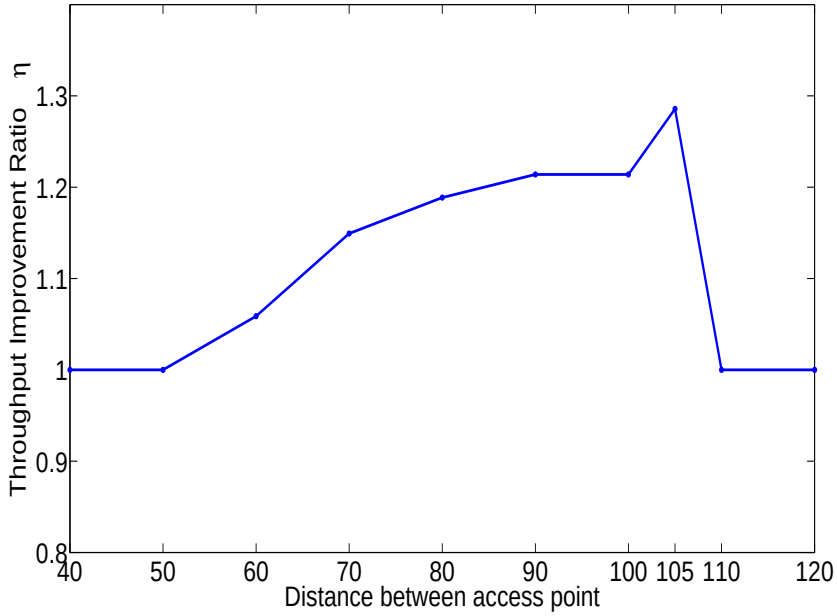


Figure 4.17: Impact of distance among access point on data delivery ratio and throughput gain

4.5 Related Work

In [52][51], the I-factor is defined as the multiplication of the frequency response of the receiver's band-pass filter to the incoming signal. The I-factor captures the percentage of energy leaked into adjacent spectrum band by an active transmitting channel. The I-factor serves as the theoretical foundation for their proposed channel assignment scheme for wireless mesh networks. The theoretical and simulation results showcase the potential of POC networks. The authors, however, do not address the issue of self-interference which prevent simultaneous usage of partially overlapped channels on the same node. In [54], the authors propose a joint channel allocation and link scheduling algorithm for POC based wireless networks using Genetic Algorithm (GA). Based on their simulation results, the authors hint that POC is more effective in dense networks and symmetric networks. The authors, however, do not further study the actual cause of their finding, hence providing

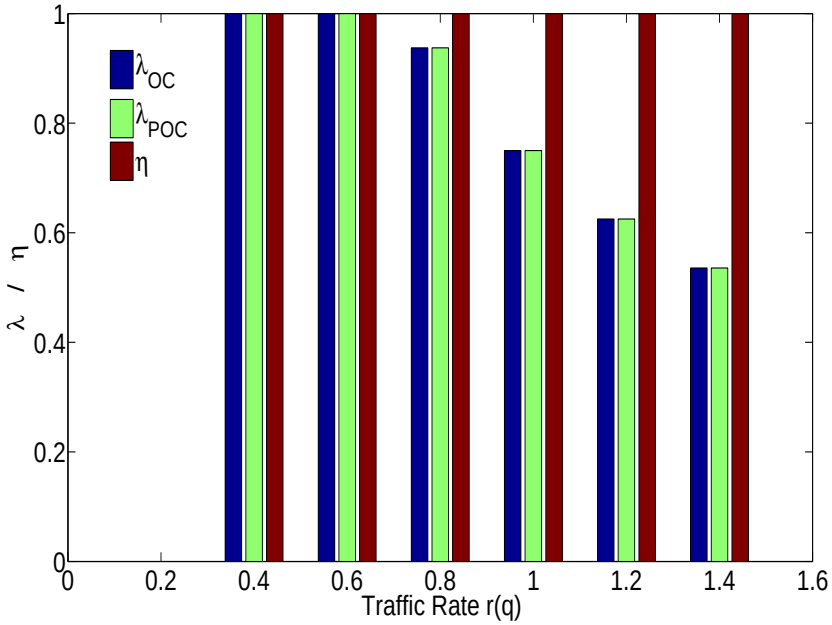


Figure 4.18: Impact of traffic load on data delivery ratio and throughput gain in WLAN with a single access point

limited insights for future algorithm design and network planning for POC based networks. Moreover, the efficiency of their GA algorithm is not discussed. In [55], Rad, et al. formulate POC allocation as a linear mixed integer programming problem. Mixed Integer Linear Programming (MILP) problem is known [89] in general to be NP-hard. As the network grows large, the MILP based solution can have scalability issues. Moreover, the authors suggest the use of commercial solvers such as CPLEX [83]. The cost and computational power required by such software might cause problems in practical scenarios.

4.6 Conclusion

In this chapter, we introduce a simple and robust model to compute aggregate one-hop network throughput for wireless networks using partially overlapped channels. We present a quantitative study on the relations between network performance and critical network

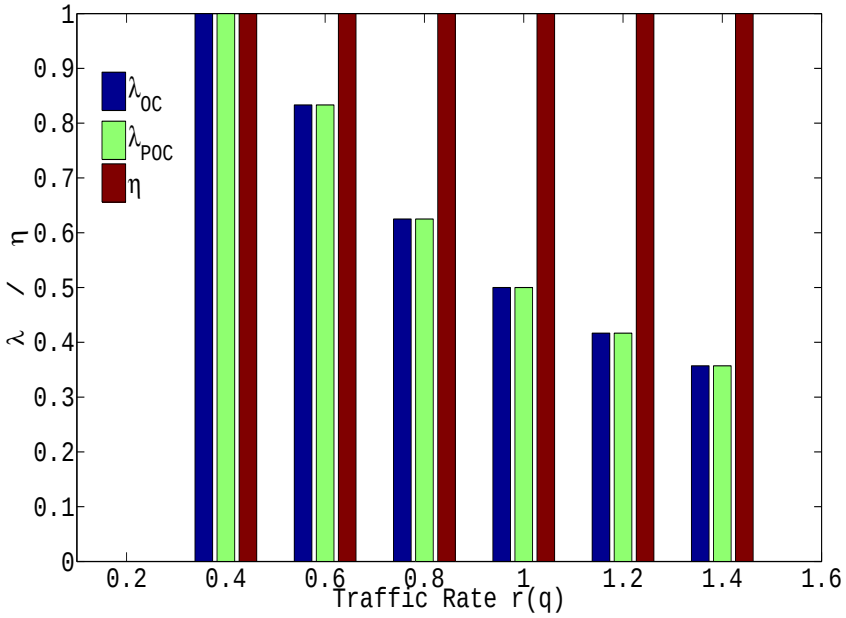


Figure 4.19: Impact of traffic load on data delivery ratio and throughput gain in sensor network

settings such as node density, network topology and traffic load. Both our mathematical induction and numerical results show that, in an uniformly distributed network, the use of partially overlapped channels significantly improves network performance. Additionally, our analytical results show that denser networks with higher traffic load favor POC based designs. Our research also reveals that throughput improvement ratio is not a monotonic function of node density. In extreme cases, such as when distribution area is at the smallest interference range level, the marginal gain of throughput improvement by using POC can be diminished or even reversed.

Based on the statistical analysis and computational results presented in this chapter, we recommend the following guidelines on the usage of POC in wireless networks regarding the throughput improvement ratio η:

- η generally increases with network density and traffic load.

- η does not increase in scarce network and the network can be decoupled into a set of sub-networks that are not interfering with each other. In each of these sub-networks, the rule in general cases applies.
- η achieves maximum when average $r(0)/d$ slightly above 1.
- η may decrease in networks with extremely small geographic size such that $r(0)/d \ll 1$. POC based design should be avoided in this case.

4.7 Appendix

Lemma 1: For POC based link-channel scheduling problem, Constraint (4.19) is a sufficient condition for Constraint (4.20)

Proof: Since $N = \lceil \frac{M}{\tau_{th}} \rceil$, the remainder of M by N can be computed as $\kappa = M - (N-1) \cdot \tau_{th}$. The M channels thus can be indexed as $I = [1, ..., \tau_{th}, \tau_{th} + 1, 2\tau_{th}, ..., (N-1)\tau_{th}, (N-1)\tau_{th} + 1, ..., (N-1)\tau_{th} + \kappa]$. With this index method, the M channels are readily divided into N groups each contains at most τ_{th} channels. Namely, they are $I_1 = [1, ..., \tau_{th}], I_2 = [\tau_{th} + 1, ..., 2\tau_{th}], ...$ If $\kappa \neq 0$ the N_{th} channels group contains less than τ_{th} channels as $[(N-1)\tau_{th}, ..., (N-1)\tau_{th} + \kappa]$. For each of these channel groups, we have $I[i] \subseteq POC(i)$ which leads to the following inequality.

$$\begin{aligned} & \sum_{u \in S_c(v)} \sum_{q \in Q} \sum_{j \in I_i} (X_j^t(v,u,q) + X_j^t(u,v,q)) \\ \leq & \sum_{u \in S_c(v)} \sum_{q \in Q} \sum_{j \in POC(i)} (X_j^t(v,u,q) + X_j^t(u,v,q)) \\ \leq & \quad 1 \end{aligned}$$

Sum up for all N channels, we have

$$
\begin{aligned}
& \sum_{u \in S_c(v)} \sum_{q \in Q} \sum_{i=1,\ldots,M} (X_i^t(v,u,q) + X_i^t(u,v,q)) \\
\leq & \sum_{i \in [1,N]} \sum_{u \in S_c(v)} \sum_{q \in Q} \sum_{j \in I_i} (X_j^t(v,u,q) + X_j^t(u,v,q)) \\
\leq & \sum_{i \in [1,N]} \sum_{u \in S_c(v)} \sum_{q \in Q} \sum_{j \in POC(i)} (X_j^t(v,u,q) + X_j^t(u,v,q)) \\
\leq & \quad N
\end{aligned}
$$

Q.E.D.

Chapter 5

Joint Spectrum Sharing and Rate Control in Frequency-Agile Radio Networks

In this chapter, we present the distributed algorithm JSSRC for joint spectrum sharing and end-to-end rate control for multihop frequency-agile networks. By taking advantage of cognitive radio's capability to dynamically reconfigure central frequency and channel width, JSSRC with appropriate link layer scheduling is able to achieve optimal spectrum sharing that maximizes the aggregate network utility of a frequency-agile radio network.

5.1 Introduction

Recent advances in frequency-agile radio technologies (e.g. software defined radio and cognitive radios) enable more flexible spectrum access through spectrum sensing and dynamic tuning of the central frequency of wireless channels. Recent development further shows that the amount/width of spectrum band of radio channels can also be configured dynamically.

For examples, in WiMax networks [80], users are allowed to use channel bandwidths that are multiples of 1.25MHz, 1.5MHz and 1.75MHz. In the IEEE 802.11n standard [81], channel bonding allows radios to form wider channel by bonding continuous smaller channels. In a recently published paper [66], a prototype radio that is capable of transmitting in four channel width of 5, 10, 20, and 40 MHz is developed from commodity Atheros-based Network Interface Card. In Virginia Tech, a prototype Multiband/Multimode Radio (MMR) is developed for public safety applications [58].

Despite the flexibility enabled by modern radios, most existing spectrum allocation works are still based on fixed channelization assumption where the center frequency and the bandwidth of wireless channels are predefined and programmed into radio firmware. This assumption of fixed channelization stems from traditional radio designs, where a radio is only designed for a specific communication system, such as a cellular system, a GPS system, a TV broadcasting system or a WLAN. Devices used in these systems work in predefined and stable spectrum ranges. Since the boundary of the available spectrum range is known to both system designers and device manufacturers, fixed channelization of this predefined spectrum range is appropriate for these systems. A pre-calculated channelization of the available spectrum and a carefully designed channel allocation scheme, e.g. [24][25], can hence be used to optimize the network performance. However, in dynamic spectrum access networks, the boundary and characteristics of the available spectrum are known to neither system designers nor device manufacturers. Depending on the activities of primary users, a spectrum whitespace may be located in any spectrum range with any size. Therefore, a spectrum access scheme based on predefined channelization is often not flexible enough and leads to low spectrum efficiency [11]. The focus of this chapter, hence, is to address this technical void by designing a practical channel spectrum adaptation algorithm to exploit the vastly improvable spectrum utilization by reconfiguring both channel spectrum width and central frequency.

Different from classic frequency scheduling in multichannel systems where only the channel number/central frequency of a radio is determined, spectrum sharing among frequency-agile radios requires reconfiguring both their channel spectrum width and central frequency. The

challenge of our work comes from the fact that optimal channel allocation for general ad-hoc wireless networks is itself a difficult (generally NP-hard) problem widely researched in recent years. The joint allocation of channel width and frequency only adds to the complexity. Hence, to obtain a computationally tractable solution, we resort to a divide-and-conquer approach. We divide the spectrum sharing problem into two subproblems: channel width allocation and link layer frequency scheduling. We model the channel width allocation (Phase I) as a joint optimization problem and derive a distributed algorithm (JSSRC) through dual decomposition. For the link layer frequency scheduling subproblem (Phase II), we propose an innovative timing window based spectrum access scheme (TWSR) that approximate the optimal spectrum width obtained in JSSRC.

The two algorithms JSSRC and TWSR together provide a systematic treatment for the inherently cross-layer problem of spectrum sharing in wireless networks and comprise the main contributions of this work. Moreover, the derivation of the algorithms reveals some of the most fundamental parameters that are critical to network performance and spectrum efficiency, e.g., local buffer saturation level, wireless link capacity saturation level, number of contenders for the same spectrum band and number of links a specific link can interfere.

The rest of this chapter is organized as follows. In Section 5.2, we describe the system model. In Section 5.3, we introduce the network utility maximization problem. In Section 5.4, the dual decomposition based framework (phase I algorithm) is detailed. We propose and provide an extensive study of the JSSRC Phase II algorithm TWSR in Section 5.5. Simulation results are presented and discussed in Section 5.6. Finally, we give concluding remarks in Section VII.

5.2 System Model

We consider a general wireless network modeled as a directed graph $G(V, E)$, where V and E denote the set of nodes and the set of directed edges respectively. For two nodes $i, j \in V$,

an edge/link l from i to j exists only if node j can successfully receive and decode signals from i, i.e, j is in the transmission range of node i. We use protocol interference model in our work. Under this model, link $\hat{l}$ can cause interference to link l if and only if the receiver of link l is in the interference range of the transmitter of link $\hat{l}$. The set of all links that can cause interference to link l is denoted as $E^I(l)$.

There are a set S of active flow sessions in the network. Each flow $s \in S$ has a session rate of y_s. The set of links on the path of a session $s \in S$ is denoted as $L(s)$ and the set of sessions that are using link l is denoted as $S(l) = \{s|l \in L(s)\}$. The set of all active links, denoted as L, can be represented by the union of L(s), i.e. $L = \bigcup_{s \in S} L(s)$. A link l is active if and only if $l \in L(s)$ for some s, i.e., l is on the path of a session s. L is a subset of E.

We define spectrum sharing policy as a set of quantities $\{(b_l, w_l)\}_{l \in \mathbf{L}}$ which specifies the width b_l and central frequency ω_l of spectrum bands of each link l. In a feasible sharing policy, there is no spectrum overlap between any two active links that can produce interference to each other. The upper bound of the amount of available spectrum that can be allocated to a single link is $b_l^{ava} \in [0, \infty)$. The upper and lower bound for the central frequency of each link l is denoted as ω_l^{max} and ω_l^{min}. In DSA networks, $\mathbf{b}^{ava} = [b_1^{ava}, b_2^{ava}, ..., b_{|\mathcal{L}|}^{ava}]$ corresponds to the spans of the white spaces that are available to secondary users. ω_l^{max} and ω_l^{min} set the boundaries of the white space that are available to link l.

We assume that the transmission power of each radio is determined before the execution of JSSRC algorithm. This means that transmission power remains relatively constant during each execution cycle of JSSRC algorithm and changes at a slower rate than the convergence time of the JSSRC. The direct impact of these assumptions combined with the non-overlapping spectrum allocation policy is that we can treat the signal-to-noise ratio SNR_l at the receiver of a link l as a constant during the execution time of JSSRC. Consequently, the capacity of a link l becomes only the function of link l's spectrum width b_l, i.e., $c_l = b_l \log(1 + SNR_l)$, where SNR_l is the signal-to-noise ratio (SNR) at the receiver of link l. We discuss how to relax this assumption in Section VIII.

Table 5.1: Notations for Chapter 5

y_s	data rate of flow session s
$\mathbf{y}$	vector form of flow sessions
c_l	capacity of link l
$\mathbf{c}$	vector form of link capacities
b_l	spectrum allocated to link l
$\mathbf{b}$	vector form of allocated spectrum
b_l^{ava}	maximum spectrum available to link l
$\mathbf{b}^{ava}$	vector form of available spectrum
L	set of unidirectional communication links in the network
S	set of flow sessions
$L(s)$	set of links in the path of source session s
$S(l)$	set of flow sessions use link l
$E^I(l)$	set of links that can interfere link l
$\mathbf{R}$	routing matrix
$\mathbf{A}$	spectrum contention matrix
$\mathbf{N}$	SNR matrix

5.3 Problem Statements of Phase I algorithm

The objective of Phase I algorithm (JSSRC) is to distributively compute optimal spectrum width allocation for each link.

If we associate each session $s \in S$ with a utility function $U_s(y_s)$ [1], the objective of JSSRC can then be summarized as a maximization problem as

$$\max \quad \sum_{s \in S} U_s(y_s), \tag{5.1}$$

subject to

$$\sum_{s \in S(l)} y_s \leq c_l \quad \forall l \in L \tag{5.2}$$

$$c_l = b_l \log(1 + SNR_l) \quad \forall l \in L \tag{5.3}$$

$$b_l + \sum_{\hat{l} \in E^I(l)} b_{\hat{l}} \leq b_l^{ava} \quad \forall l \in L \tag{5.4}$$

$$b_l \geq 0, c_l \geq 0, y_s \geq 0, \quad \forall l \in L, s \in S \tag{5.5}$$

In this formulation, Eq. (5.2) ensures that the aggregate rate of sessions that traverse link l does not exceed l's link capacity $c(l)$. Eq. (5.3) is the capacity constraint derived from information theory. Eq. (5.4) ensures that any allocation obtained from the formulation is feasible, i.e., there is no spectrum overlap between two active links that are in each other's interference range.

For ease of presentation, we derive a vector form of the primal problem in the following. Let $\mathbf{b} = [b_1, ..., b_{|L|}]^T$ and $\mathbf{c} = [c_1, ..., c_{|L|}]^T$ denote the spectrum width allocation vector and the link capacity vector respectively. Let $\mathbf{N}$ be a $|L| \cdot |L|$ diagonal matrix where $n_{ll} = \log(1 + SNR_l)$ and $n_{lk} = 0$ for $l \neq k$. Let $\mathbf{y} = [y_1, ..., y_{|S|}]^T$ denote the vector of session rates.

[1] For convergence of JSSRC, we assume U(.) to be twice continuously differentiable, increasing and strictly concave function of y_s.

Define spectrum contention matrix $\mathbf{A}^{|\mathbf{L}|\cdot|\mathbf{L}|}$ as

$$A_{l\hat{l}} = \begin{cases} 1 & \hat{l} = l \text{ or } \hat{l} \in E^I(l), \text{ i.e.,} \\ & \hat{l} \text{ can produce interference to link } l\text{'s receiver;} \\ 0 & \text{otherwise.} \end{cases}$$

Define routing matrix $\mathbf{R}$ as

$$R_{ls} = \begin{cases} 1 & \text{link } l \in L \text{ is on the path of session } s \in S; \\ 0 & \text{otherwise.} \end{cases}$$

We can obtain an equivalent but more compact form of the primal problem:

$$\max_{\mathbf{y},\mathbf{b}} \quad \mathbf{U(y)} = \sum_s U(y_s) \tag{5.6}$$

subject to

$$\mathbf{Ry} \leq \mathbf{c} \tag{5.7}$$

$$\mathbf{c} = \mathbf{Nb} \tag{5.8}$$

$$\mathbf{Ab} \leq \mathbf{b}^{ava} \tag{5.9}$$

$$\mathbf{y,\ b,\ c} \geq \mathbf{0} \tag{5.10}$$

We use both the original and the compact formulations interchangeably whenever there is no confusion.

5.4 The dual framework of Phase I algorithm

Note that in the primal formulation (5.1-5.5), the spectrum width allocation and the session rate control are coupled. Moreover, all session rates are coupled in the first constraint Eq. (5.2). Solving this primal optimization problem requires global information and centralized control of the transport and link layers of all links in a network. The difficulty in gathering all the information makes such centralized approach impractical in a distributed multihop

radio network. Therefore, in this section, we present a dual based approach to solve the problem in a distributed manner.

Consider the dual of the primal problem (5.1)

$$min_{\mathbf{u}\geq 0} \quad D(\mathbf{u}) \tag{5.11}$$

with partial dual function

$$D(\mathbf{u}) = \max_{\mathbf{b},\mathbf{y}\geq \mathbf{0}} \quad \mathbf{U(y)} - \mathbf{u}^T(\mathbf{Ry} - \mathbf{c}) \tag{5.12}$$

$$\text{subject to} \qquad \mathbf{Ab} \leq \mathbf{b}^{ava} \tag{5.13}$$

$$\mathbf{c} = \mathbf{Nb}. \tag{5.14}$$

Here, we relax the constraint $\mathbf{Ry} \leq \mathbf{c}$ by introducing Lagrange multiplier $\mathbf{u} = [u_1, ..., u_{|L|}]^T$. The maximization problem in the partial dual function can be further decomposed into the following two subproblems

$$D_1(\mathbf{u}) = \max_{\mathbf{y}\geq 0} \quad \mathbf{U(y)} - \mathbf{u}^T\mathbf{Ry}$$

and

$$D_2(\mathbf{u}) = \max_{\mathbf{b}\geq 0} \quad \mathbf{u}^T\mathbf{c}$$

$$\text{subject to} \quad \mathbf{Ab} \leq \mathbf{b}^{ava}$$

$$\mathbf{c} = \mathbf{Nb}$$

This decomposition successfully makes the joint spectrum sharing and rate control problem into three separate optimization problems: D_1, D_2 and the master dual problem at Eq. (5.11). These three optimization problems are correlated through the Lagrangian multipler $\mathbf{u}$. $D_1(\mathbf{u})$ is a standard rate control problem which can be solved at the transport layer. $D_2(u)$ is a link layer resource allocation problem that maximizes the weighted sum of link capacities with the Lagrangian multiplier as the weights. The master dual problem is an optimization problem regarding the Lagrangian multiplier $\mathbf{u}$. Each of these three problems can be solved using distributed iterative algorithms.

5.4.1 Solving the subproblem D_1

To solve D_1, note that D_1 can be rewritten as

$$D_1(\mathbf{u}) = \sum_{s \in S} \left[\max_{y_s \geq 0} \quad U_s(y_s) - \sum_{l \in L(s)} u_l y_s \right]$$

From the above equation, it can be seen that the optimal solution to D_1 can be further decomposed into many small maximization problems that can be executed distributively at each source requiring only local information, i.e., source node of a session only needs to know the summation of u_l along the session's path to calculate the optimal rate for this session. For each source, the unique maximizer is computed as

$$y_s = arg \max_{y_s > 0} \{U_s(y_s) - \sum_{l \in L(s)} u_l y_s\}, \quad \forall s \in S. \tag{5.15}$$

5.4.2 Solving the subproblem D_2

The spectrum width allocation sub-problem D_2 is a linear programming problem that determines the optimal amount of spectrum width allocated to each link. By substituting link capacity $\mathbf{c}$ with allocated spectrum width $\mathbf{b}$, the objective function of D_2 is just a weighted sum of the elements of the spectrum width allocation vector $\mathbf{b}$, which is a linear function. The constraints are also linear functions. While solving such a linear programming problem requires only moderate computational power (polynomial time), directly calculating the spectrum width allocation requires global information about the lagrange multiplier u, which is hard to implement in large networks. Hence, we design an efficient and distributed algorithm requiring only local information to solve D_2 as follows.

Again, we use dual decomposition and subgradient method to obtain a distributed algorithm to solve problem D_2. Consider the dual of D_2

$$min_{\mathbf{v} > 0} \quad L(\mathbf{v}) \tag{5.16}$$

with lagrangian dual function

$$L(\mathbf{v}) = max_{\mathbf{b}\geq\mathbf{0}} \quad \mathbf{u}^T\mathbf{N}\mathbf{b} - \mathbf{v}^T(\mathbf{A}\mathbf{b} - \mathbf{b}^{ava}). \tag{5.17}$$

It is worth noting that by solving (6.17), the primal optimal variables are not immediately available. This is because the dual problem may have multiple optimal points with the same dual objective. To remove this complication, we add a small regularization term in D_2, i.e., a small quadratic term $\sigma\mathbf{b}^T\mathbf{b}$, and maximize

$$\mathbf{u}^T\mathbf{N}\mathbf{b} - \sigma\mathbf{b}^T\mathbf{b}.$$

Here, σ is a very small positive constant. Subsequently, the dual problem of D_2 changes to

$$min_{\mathbf{v}>0} \quad L(\mathbf{v})$$

with Lagrangian dual function

$$L(\mathbf{v}) = max_{\mathbf{b}\geq\mathbf{0}} \quad \mathbf{u}^T\mathbf{N}\mathbf{b} - \sigma\mathbf{b}^T\mathbf{b} - \mathbf{v}^T(\mathbf{A}\mathbf{b} - \mathbf{b}^{ava}) \tag{5.18}$$

Using the gradient projection method, we can derive the iterative algorithm to solve the regularized dual problem of D_2 as follows:

$$\mathbf{b}(m+1) = [\frac{1}{2\sigma}(\mathbf{N}\mathbf{u} - \mathbf{A}^T\mathbf{v}(m))]^+ \tag{5.19}$$

$$\mathbf{v}(m+1) = [\mathbf{v}(m) + \beta(\mathbf{A}\mathbf{b}(m+1) - \mathbf{b}^{ava})]^+ \tag{5.20}$$

where β is a positive stepsize.

For each link l, Eq. (6.24) and Eq. (6.26) reduce to

$$b_l(m+1) = \left[\frac{1}{2\sigma}\left(\log(1+SNR_l)u_l - \sum_{\hat{l}:l\in\mathcal{C}(\hat{l})}(v_{\hat{l}}(m)A_{\hat{l}l})\right)\right]^+, \tag{5.21}$$

$$v_l(m+1) = \left[v_l(m) + \beta\left(\sum_{\hat{l}\in\mathcal{C}(l)}(A_{l\hat{l}}b_{\hat{l}}(m+1)) - b_l^{ava}\right)\right]^+, \tag{5.22}$$

The above iterative update algorithms for b_l and v_l only requires local information, e.g. $v_{\hat{l}}$ and $b_{\hat{l}}$ of links that can generate interference to link l and links that link l can interfere if transmit simultaneously using the same spectrum block.

The convergence of the above iterative algorithm to the optimal solution of D_2 is well established [88]. Define $\bar{a}_1 = max_{l \in L} |E^I(l)|$. Define $\bar{a}_2$ as the largest number of constraints (inequalities) in (5.4) that any b_l may be involved in.

The range of the stepsize with which the algorithm converges can be defined as in [39]:

$$0 < \beta < \frac{4\sigma}{\bar{a}_1 \cdot \bar{a}_2}$$

5.4.3 Solving the master dual problem

To solve the master dual problem in Eq. (5.11), we use an iterative algorithm based on the subgradient projection method.

Let $\mathbf{y}(\mathbf{u}(n))$ and $\mathbf{c}(\mathbf{u}(n))$ denote the maximizers of D_1 and D_2 given the value of $\mathbf{u}$ at the n_{th} iteration, i.e.,

$$\mathbf{y}(\mathbf{u}(n)) = \arg \max_{\mathbf{y} \geq \mathbf{0}} \mathbf{U(y)} - \mathbf{u}^T \mathbf{R} \mathbf{y} \tag{5.23}$$

and

$$\mathbf{c}(\mathbf{u}(n)) = \arg \max_{\mathbf{b} \geq \mathbf{0}} \mathbf{u}^T \mathbf{c} \quad \text{subject to} \quad A\mathbf{b} \leq \mathbf{b}^{ava} \tag{5.24}$$

Given $\mathbf{y}(\mathbf{u}(n))$ and $\mathbf{c}(\mathbf{u}(n))$, we can then use the subgradient method to solve the master dual problem in Eq. (5.11) by iteratively updating the Lagrangian multiplier $\mathbf{u}$ as follows:

$$\mathbf{u}(n+1) = [\mathbf{u}(n) + \gamma(\mathbf{R}\mathbf{y}(\mathbf{u}(n)) - \mathbf{c}(\mathbf{u}(n)))]^+ \tag{5.25}$$

where γ is a positive scalar stepsize, $(\mathbf{R}\mathbf{y}(\mathbf{u}(n)) - \mathbf{c}(\mathbf{u}(n))$ is the subgradient of the dual function (Eq. (5.12)) at the n_{th} iteration, and '+' denotes the projection to the set $\mathcal{R}^+$ of non-negative real numbers.

For each link l, the above equation reduces to

$$u_l(n+1) = \Big[u_l(n) + \gamma\Big(\sum_{s:l\in L(s)} y_s(\mathbf{u}(n)) - c_l(\mathbf{u}(n))\Big)\Big]^+.$$

Essentially, a link l updates its local u_l based on the amount of traffic that goes through it and its own capacity.

The system of Eq. (6.6) , (6.24) , (6.26) and (6.30) forms the mathematical base to design JSSRC Phase I algorithm for solving the dual problem (5.11).

The pseudo code for JSSRC Phase I algorithm is provided in Algorithm 1.

Algorithm 1: JSSRC Phase I

Input : $\mathcal{G}(\mathcal{V},\mathcal{E}), \mathcal{A}, \mathcal{R}$

Output : **b,y**

begin
  Initialize (**b,y,u**)
  while $||\mathbf{u(n+1)} - \mathbf{u(n)}||_\mathbf{2} > \delta_1$ **do**
    while $||\mathbf{y(m+1)} - \mathbf{y(m)}|| > \delta_\mathbf{2}$ **do**
      Source rate update:
      $y_s \leftarrow \Big[y_s + \alpha\Big(U_s'(y_s) - \sum_{l\in L(s)} u_l\Big)\Big]^+$
    end
    while $||\mathbf{b(m+1)} - \mathbf{b(m)}||_\mathbf{2} < \delta_\mathbf{3}$ **do**
      Update spectrum width and v_l as:
      $b_l \leftarrow \Big[\frac{1}{2\sigma}\Big(\log(1+SNR_l)u_l - \sum_{\hat{l}:l\in\mathcal{C}(\hat{l})} v_{\hat{l}} A_{\hat{l}l}\Big)\Big]^+,$
      $v_l \leftarrow \Big[v_l + \beta\Big(\sum_{\hat{l}\in\mathcal{C}(l)} A_{l\hat{l}} b_{\hat{l}} - b_l^{ava}\Big)\Big]^+.$
    end
    Update u_l as:
    $u_l \leftarrow \Big[u_l + \gamma\Big(\sum_{s:l\in L(s)} y_s - c_l\Big)\Big]^+$
  end
  output $(\bar{\mathbf{b}}, \bar{\mathbf{y}})$
end

$*\delta_1, \delta_2, \delta_3$ are corresponding termination thresholds

5.4.4 Convergence and Optimality

If we assume the norm of the subgradients of the dual function is bounded, i.e., there exist a constant G such that $||\sum_{s:l\in L(s)}(y_s(\mathbf{u}(n)) - c_l(\mathbf{u}(n)))|| \leq G$, $\forall n, l$. the convergence of the algorithm is guaranteed by Proposition 1. Since the primal problem is strict convex, the obtained fixed point is also the optimal solution. The proof of Proposition 1 is provided in the Appendix.

Proposition 1: Let $\mathbf{u}^*$ denote an optimal spectrum contention price, the iterative solution system of Eq.(6.6), (6.24), (6.26) and (6.30) converges statistically to within $\frac{\gamma G^2}{2}$ of the optimal value.

5.4.5 Understanding JSSRC Phase I algorithm

Before exploring any further, let us take a closer look at Phase I algorithm to gain some insight on its physical meaning.

By treating u_l as link price, the subproblem $D_2(\mathbf{u})$ can be interpreted as maximizing the aggregated link layer utility through distributed allocation of spectrum width at each link. Links with higher prices u_l are likely to be allocated with more spectrum as they contribute more to the aggregate utility. This makes sense since links with higher prices u_l in our design is those links that have heavier traffic and hence needs more spectrum (to increase the throughput on link l and consequently to increase the aggregate utility). Due to interference among links in the network, when link l is allocated with more spectrum, the available spectrum for its neighboring links $\bar{l}$ in E_l^I is decreased, which may increase the congestion level at link $\bar{l}$ (Thus decrease the throughput on link $\bar{l}$). Apparently, there is a tradeoff between local utility maximization and the aggregate utility maximization. The lagrangian multipliers v_l are used as messengers to pass this tradeoff information among links. It is interpreted as the price that needs to be paid to link l for an interfering link to take a

unit of available spectrum from link l. In Eq. (6.26), JSSRC increases the contention price v_l of link l when link l's local available spectrum b_l^{ava} (spectrum left by primary users) becomes smaller than the aggregate spectrum needs of the links in its neighborhood (s.t. $\sum_{\hat{l} \in E_l^I} (A_{l\hat{l}} b_{\hat{l}}(m+1)) > b_l^{ava}$) to prevent other links from starving link l. When there is still unallocated available spectrum (s.t. $\sum_{\hat{l} \in \mathcal{C}(l)} (A_{l\hat{l}} b_{\hat{l}}(m+1)) < b_l^{ava}$), link l decreases its contention price v_l so that other interfering links are encouraged to use this unallocated spectrum.

5.5 JSSRC phase II algorithm

In Phase I algorithm JSSRC, we determine the spectrum width that each link can use. In Phase II algorithm, we move on to determine the central frequencies of spectrum bands that these links uses. If we assume a centralized spectrum broker, the frequency scheduling problem can be trivially solved at the broker. The optimal policy given by phase I algorithm JSSRC is achieved when each node follows the command of the central broker. However, in large networks, such a central spectrum broker is not likely to exist. The links only have limited local information and thus prone to make "bad" decisions on frequency selection. Hence, an optimal frequency scheduling may not be easily obtained. In the following, we first demonstrate a case where naive scheduling algorithm fails to generate a good policy and then propose a novel distributed Timing Window based Spectrum Sharing (TWSR) algorithm that closely approximates the optimal spectrum sharing policy without requiring centralized spectrum broker.

5.5.1 A case study of distributed link layer scheduling

Consider a simple network of three nodes in Figure 6.1. Suppose there are two active one-hop sessions s_1 and s_2 from node 1 to 2 and 3 to 2, respectively. Each session requires 50MHz spectrum to transmit its traffic. Node 1 and 3 each has one radio and node 2

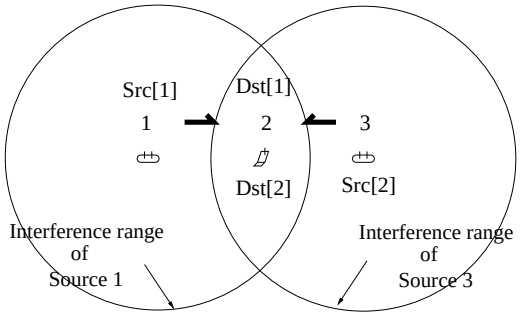


Figure 5.1: Illustration of naive distributed link scheduling algorithm

has two radios. Links (1,2) and (3,2) mutually interfer with each other. Suppose there is an available spectrum whitespace of 100MHz left by primary users and the spectrum whitespace starts from 2.4GHz. The optimal spectrum sharing policy should allocate each link a 50MHz wide spectrum. This is only possible if link (3,2) use either the first half block of the spectrum whitespace starting from 2.4GHz or use the second half block starting from 2.45GHz, while the other link uses the remaining half simultaneously. A centralized spectrum broker can easily realize such a spectrum allocation. However, when links make distributed spectrum reservation decisions, link (3,2) may first reserve a 50MHz spectrum range starting from 2.425GHz without any trouble. However, this reservation left two 25MHz spectrum fragments and link (1,2) can only use one of the spectrum fragment due to the limit of only one radio at node 1. Hence, with the limitation on radios per link, distributed spectrum reservation may not result in optimal spectrum scheduling.

5.5.2 Timing Window based Spectrum Reservation

To address the above issue, we propose the Timing Window based Spectrum Reservation (TWSR) algorithm for link layer frequency scheduling. TWSR works in scenarios where there is only one dedicated radio at a node for each of its links. The spectrum scheduling produced by TWSR approximates the optimal spectrum scheduling.

Based on b^*, which is the optimal value computed in Phase I algorithm, TWSR gives an

approximate spectrum allocation $\bar{\mathbf{b}}$ and a corresponding frequency scheduling $\bar{\mathbf{f}}$. The design of TWSR is based on the observation that the more constraints in Eq (5.9) that a link l belongs to, the more difficult to find an appropriate spectrum fragment with size b_l for link l. Hence, those links that are included in more constraints should be given higher priority in spectrum reservation.

We define the number of constraints link l belongs to in Eq. (5.9) as link l's constraint degree ϱ_l. After link l obtains its new spectrum width b_l^* in Phase I, link l computes a timing window based on its constraint degreee ϱ_l, where the lower bound of the time window is $w_l^{lb}=\delta/\varrho_l$ and the upper bound is $w_l^{ub}=\delta/(\varrho_l-1)$. Here, δ is a scalar determined by timing granularity of the network. Link l then randomly picks a time between w_l^{lb} and w_l^{ub} to declare its spectrum reservation to its interfering links. The spectrum reserved by link l is picked as the lowest unreserved spectrum fragment that has size b_l. If such a spectrum fragment does not exist, link l reserves the largest spectrum fragment available. The spectrum reservation process is a first-come-first-serve process. A link l only reserves a spectrum fragment that is not previously reserved by its interfering links. This first-come-first-serve scheduling and non-overlapping declaration mechanism ensures that links that are more difficult to schedule declare their spectrum reservation earlier and, hence, have more chance to get exactly the spectrum size that they demand. The pseudo code of TWSR is provided in Algorithm 4.

The spectrum allocation vector $\bar{\mathbf{b}}$ generated by TWSR is generally a suboptimal version of $\mathbf{b}^*$. However, as shown in section 6.4, TWSR combined with the Phase I of JSSRC shows good performance. Not only does it generate final results very close to optimal results, it also converges in many of the cases in a relatively short time limit (400 iterations). Note that the spectrum contention price $\mathbf{u}$ at the end of each master iteration (line 14 in Algorithm 1) is updated according to $\bar{\mathbf{b}}$ generated by TWSR instead of the optimal value $\mathbf{b}^*$. As a result, the update in the master iteration can partially correct the approximation error made by $\bar{\mathbf{b}}$. In Proposition 2, we give an upperbound on the approximation error of JSSRC with TWSR. It turns out that the combined approximation error of the iterative algorithm is upper bounded by how close the suboptimal schedule $\bar{\mathbf{b}}(n)$ is to the optimal schedule $\mathbf{b}^*(n)$

in each iteration.

Proposition 2: Suppose at each iteration n, the norm of the error between suboptimal $\bar{\mathbf{b}}(\mathbf{u}(n)$ and $\mathbf{b}^*(\mathbf{u}(n))$ is bounded, i.e., $||\bar{\mathbf{b}}(\mathbf{u}(n)) - \mathbf{b}^*(\mathbf{u}(n))||_2 \leq \epsilon$, then under the same assumption as in *Theorem 1*, JSSRC converges statistically to within $\gamma\frac{G^2}{2} + \epsilon log(1 + SNR_{max})$ of the optimal, where $SNR_{max} = \max_l\{SNR_l\}$.

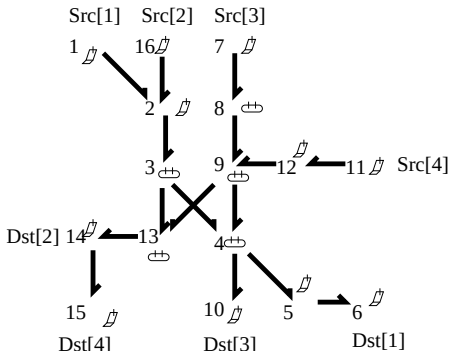


Figure 5.2: Simulated network topology and routing for JSSRC: Scenario I

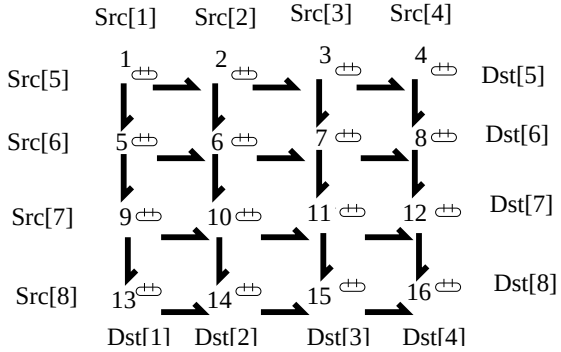


Figure 5.3: Simulated network topology and routing for JSSRC: Scenario II

5.6 Numerical Results

To evaluate the performance of JSSRC, we run extensive simulations of JSSRC with TWSR in different topologies and network sizes. The number of sessions in a simulation ranges from

Algorithm 2: Timing Window based Spectrum Reservation

Input : $\mathbf{b}^*(n)$

Output: $\bar{\mathbf{b}}(n), \bar{\mathbf{f}}(n)$

begin

  Compute timing window at each link $w_l^{lb} \leftarrow \delta/\varrho_l$, $w_l^{ub} \leftarrow \delta/(\varrho_l - 1)$

  Each link randomly picks a time point t_l between $[w_l^{lb}, w_l^{ub}]$

  Each link l sets a *timer* T_l with initial value t_l and starts counting down the timer

  while *not all timers timeout* **do**

    if $\forall T_k$ *timeout* **then**

      while *not finishing scanning the entire spectrum whitespace left by the primary users* **do**

        Link k scans for unreserved spectrum fragments starting from the lowest frequency in the spectrum whitespace left by primary users.

        if *link k finds a fragment of width* $b^w \geq b_k^*$ **then**

          Link k reserves b_k^* amount of spectrum starting from the lowest frequency f^w in this fragment.

          Link k informs the other links in $\mathcal{C}(k)$ about $\bar{b}_k \leftarrow b_k^*$ and $\bar{f}_k \leftarrow f^w + b_k^*/2$

        end

      end

      if *no spectrum fragment found* **then**

        Link k reserves the largest available spectrum fragment b^w.

        Link k informs the other links in $\mathcal{C}(k)$ about $\bar{b}_k \leftarrow b^w$ and $\bar{f}_k \leftarrow f^w + b^w/2$.

      end

    end

  end

  output $(\bar{\mathbf{b}}, \bar{\mathbf{f}})$

end

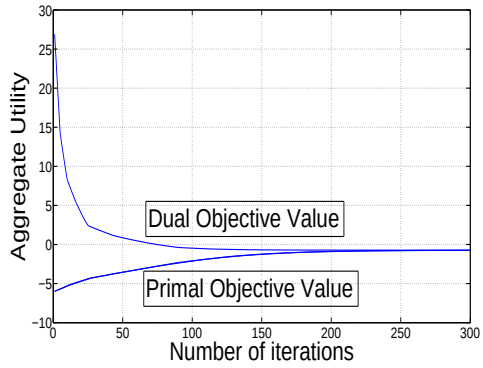


Figure 5.4: JSSRC algorithm converges to optimal in Scenario I, Primal and dual iteration converges

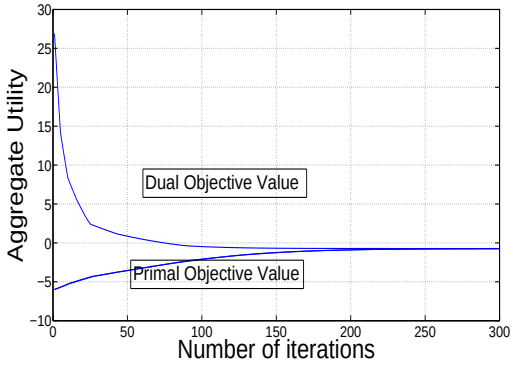


Figure 5.5: JSSRC algorithm converges to optimal in Scenario II, Primal and dual iteration converges

2 to 8. The network utility function is taken as $U(\cdot) = log(\cdot)$. In all scenarios a node can interfere with any of its one-hop neighbors. In the simulation, we set step size for both master iterations (line 15 in Algorithm 1) and secondary iterations (line 6 and 11 in Algorithm 1) to 0.01. We compare JSSRC with TWSR to the optimal solution obtained by JSSRC with a centralized frequency scheduler.

We first show two representative scenarios to characterize the behavior of JSSRC.

In Scenario 1, there are 4 active flow sessions in a 15 nodes network. We can see from Fig. 5.4 that the convergence curve of JSSRC with TWSR shows large oscillation at the early

iterations, then gradually smooths out and finally converges to the optimal. According to *Proposition 2*, this indicates that TWSR generates close to optimal schedules after enough number of iterations. JSSRC with TWSR also shows convergence speed as good as the JSSRC with a centralized frequency scheduling, which demonstrates the effectiveness of our design.

In Scenario 2, as shown in Fig. 5.3, there are 8 active sessions in a 16-node network. This network is considered a dense network with heavy load of traffic.

From the convergence curve in Fig. 5.5, we can see that JSSRC with TWSR still performs well. It closely matches with the optimal curve. The final error is also well bounded from and very close to the optimal. The simulated curve oscillates around the optimal curve with a set of scheduling solutions generated by TWSR.

To stop the oscillation we can either simply set a maximum number for iterations or add a stabilizer in the dual objective function to stabilize the network.

Other results are summarized in Table I, which compares the performance of JSSRC with TWSR against JSSRC with a centralized frequency scheduling algorithm in terms of convergence speed and achieved aggregate utility. Table 5.2 shows results from 10 scenarios with different settings of topologies and network sizes. The collective results of JSSRC with TWSR is very promising. The convergence speed of JSSRC with TWSR is almost as fast as the optimal. The gap between the stable points of JSSRC-TWSR and the optimal is averaged to be less than 1.5%.

5.7 Related Works

In literature [69][70][74], frequency agility is often treated the same as multi-channel communication capability. As pointed out in [61], SDRs/CRs are much more powerful and flexible

Table 5.2: More simulation results for JSSRC

Table I				
Group #	TWSR		Optimal	
	Aggregate Utility	No. of iterations	Aggregate Utility	No. of iterations
1	3.2209	159	3.2561	130
2	2.1285	223	2.1287	223
3	2.6354	192	2.6355	192
4	3.2027	400*	3.2561	368
5	1.2087	225	1.2087	209
6	2.3027	400*	2.3835	308
7	3.3049	342	3.3053	283
8	3.1342	400*	3.1349	294
9	3.7338	320	3.7344	302
10	0.7838	400*	0.7986	400*

than multi-channel radios. While the use of multi-channel model can significantly simplify the algorithm design for CR/SDR networks, it limits the potential of SDR/CR network. In [61], the author use an uneven bandwidth segmentation model for spectrum sharing and formulate the joint flow routing and spectrum allocation problem as a Mixed Integer and Non-Linear Programming problem. An algorithm based on Sequential Fixing (SF) is then proposed for the MINLP. A distributed algorithm for the same problem is developed in [62] under similar assumptions. Numerical results show that both algorithms developed in [61] and [62] can achieve near optimal performance. These two schemes, however, assume a known and universal spectrum segmentation among all spectrum band and for all radio access technologies. This assumption might not fit into the scenario of Dynamic Spectrum Access (DSA) networks and cognitive radio networks since white-space/hole in the spectrum might be irregularly or even continuously distributed anywhere in the whole spectrum band. More importantly, the algorithms assumes known traffic from fixed pairs of nodes (data sessions) and fixed network topology. The flow routing and spectrum allocation solution is good for stable and stationary networks. In a dynamic network, however, any change in the network may trigger a rerun of the entire algorithm to get the new near-optimal solution. This can become a problem when the network changes frequently, which is common in cognitive radio and dynamic spectrum access networks. Adaptive algorithms such as JSSRC, in contrast, are more robust in a dynamic environment and are thus more desirable in those scenarios.

On the other hand, in the Internet context, adaptive algorithms for congestion control and traffic rate control have been extensively researched over the last two decades [33]-[37]. Such research provides optimization-based framework for Internet congestion control and derives adaptive and distributed solutions through both primal and dual approaches. Recent research results [43]-[50] show that it is possible to model principles of wireless network design under a similar optimization framework. In wireless networks, however, upper layer protocol design is inherently coupled with lower layer properties and vice versa. For example, traffic rate and routing choices at upper layers actually change the existence and magnitude of

interference at the physical layer which in return affects the feasibility and efficiency of upper layer decisions on traffic rate. For this reason, cross-layer designs [43]-[50] are predominant in wireless environment.

5.8 Conclusion

We present a distributed algorithm JSSRC for joint spectrum sharing and end-to-end rate control for multihop cognitive radio networks. By taking advantage of cognitive radio's capability to dynamically reconfigure central frequency and channel width, JSSRC with perfect link layer scheduling is able to achieve the optimal spectrum sharing policy that maximizes transport layer utility. For practical implementation, we present a novel timing window based spectrum reservation scheme (TWSR) for link layer scheduling. Simulation results show that JSSRC algorithm with TWSR is highly effective.

As the preliminary results to a practical joint transport, routing and spectrum sharing adaptation algorithm, the basic framework of JSSRC can be extended in various directions. We describe some of the extensions here and briefly discuss how to handle them.

First, to accommodate time-varying channels, a stochastic channel model can be used to describe the network. For example, if we assume the channel state is described by a finite state Markov Chain, then we can define the average capacity region $\bar{\Lambda}$ as a function of these states and JSSRC can be straighforwardly extended to maximize network utility over these average capacity region.

Second, JSSRC can be extended to work asynchronously. Actually, the dual decomposition is well suited for asynchronous design. Convergence and optimality analysis for asynchronous algorithm is at the center of this extension. Classic asynchronous implementations of parallel and distributed computing systems can be used for this extension.

Third, the complexity of JSSRC is to be further reduced. This includes reducing both the

computational complexity and communication overhead. A more efficient JSSRC core using Newton-like algorithm may be designed. Efficient central frequency scheduling algorithm is also critical. As mentioned in [46], a tradeoff analysis of optimality and performance can be beneficial for improving JSSRC.

In the future, we are also planning to build up a DSA testbed that runs JSSRC or its variants to investigate the practical aspect of JSSRC.

5.9 Appendix

Proposition 1: Let $\mathbf{u}^*$ denote an optimal spectrum contention price, the iterative solution system of Eq.(6.6), (6.24), (6.26) and (6.30) converges statistically to within $\frac{\gamma G^2}{2}$ of the optimal value.

Proof: By Eq. (6.30), we have

$$\begin{aligned} & ||\mathbf{u}(t+1) - \mathbf{u}^*||_2^2 \\ = \; & ||[\mathbf{u}(t) - \gamma \mathbf{g}(t)]^+ - \mathbf{u}^*||_2^2 \\ \leq \; & ||\mathbf{u}(t) - \gamma \mathbf{g}(t) - \mathbf{u}^*||_2^2 \\ = \; & ||\mathbf{u}(t) - \mathbf{u}^*||_2^2 - 2\gamma \mathbf{g}(t)^T(\mathbf{u}(t) - \mathbf{u}^*) + \gamma^2 ||\mathbf{g}(t)||_2^2 \\ \leq \; & ||\mathbf{u}(t) - \mathbf{u}^*||_2^2 - 2\gamma (D(\mathbf{u}(t)) - D(\mathbf{u}^*)) + \gamma^2 ||\mathbf{g}(t)||_2^2 \end{aligned}$$

Since where the last inequality is derived from the definition of subgradient. Applying the inequalities recursively, we obtain

$$\begin{aligned} & ||\mathbf{u}(t+1) - \mathbf{u}^*||_2^2 \\ \leq \; & ||\mathbf{u}(1) - \mathbf{u}^*||_2^2 - 2\gamma \textstyle\sum_{\tau=1}^{t} [D(\mathbf{u}(\tau)) - D(\mathbf{t}^*)] \\ & + \gamma^2 \textstyle\sum_{\tau=1}^{n} ||\mathbf{g}(t)||_2^2 \end{aligned}$$

Since

$$||\mathbf{u}(t+1) - \mathbf{u}^*||_2^2 \geq 0 \quad ,$$

we have

$$\begin{aligned} & 2\gamma \sum_{\tau=1}^{t} [D(\mathbf{u}(\tau)) - D(\mathbf{u}^*)] \\ \leq \; & ||\mathbf{u}(1) - \mathbf{u}^*||_2^2 + \gamma^2 \sum_{\tau=1}^{t} ||\mathbf{g}(t)||_2^2 \\ \leq \; & ||\mathbf{u}(1) - \mathbf{u}^*||_2^2 + n\gamma^2 G^2 \end{aligned}$$

From this inequality we obtain

$$\frac{1}{t} \sum_{\tau=1}^{t} [D(\mathbf{u}(\tau)) - D(\mathbf{u}^*)] \leq \frac{||\mathbf{u}(1) - \mathbf{u}^*||_2^2 + t\gamma^2 G^2}{2t\gamma}$$

$$\limsup_{t \to \infty} \frac{1}{t} \sum_{\tau=1}^{t} [D(\mathbf{u}(\tau)) - D(\mathbf{u}^*)] \leq \frac{\gamma G^2}{2} \tag{5.26}$$

This shows that the dual algorithm converges statistically to within $\frac{\gamma G^2}{2}$ of the optimal value

Lemma 1: Suppose at each iteration n, the norm of the error between suboptimal $\bar{\mathbf{b}}(n+1)$ and $\mathbf{b}(n+1)$ is bounded, i.e., $||\bar{\mathbf{b}}(n) - \mathbf{b}(n)||_2 \leq \epsilon$, the norm of the error between suboptimal subgradient $\bar{\mathbf{g}}(n)$ and $\mathbf{g}(n)$ for the dual problem is also bounded, i.e., $||\bar{\mathbf{g}}(n) - \mathbf{g}(n)||_2 \leq \epsilon log(1 + SNR_{max})$ where $SNR_{max} = \max_l\{SNR_l\}$, γ is the step size in master iteration and $\mathbf{p}$ is the power vector.

Proof: Substitute $\mathbf{c(n)}$ for $\mathbf{b(n)}\max_l\{log(1 + SNR_l)\}$ in Eq.(6.30). Without loss of generality, let $\log(1 + SNR_{max}) = 1$, we have

$$\mathbf{g}^T(n+1) = \mathbf{R}\mathbf{y}(\mathbf{u}^T(n)) - \mathbf{b}(\mathbf{u}^T(n))$$

and

$$\bar{\mathbf{g}}^T(n+1) = \mathbf{R}\mathbf{y}(\mathbf{u}^T(n)) - \bar{\mathbf{b}}(\mathbf{u}^T(n)) \tag{5.27}$$

Using assumption

$$||\bar{\mathbf{b}}(\mathbf{u}(n)) - \mathbf{b}(\mathbf{u}(n))||_2 \leq \epsilon$$

and Eq. (5.27), we have

$$
\begin{aligned}
& ||\bar{\mathbf{g}}(n+1) - \mathbf{g}(n+1)||_2 \\
= \; & ||\bar{\mathbf{b}}(\mathbf{u}(n)) - \mathbf{b}(\mathbf{u}(n))||_2 \leq \; \epsilon
\end{aligned}
$$

Put back the constant $log(1 + SNR_{max})$ we have

$$||\bar{\mathbf{g}}(n) - \mathbf{g}(n)||_2 \leq \epsilon log(1 + SNR_{max})$$

This completes the proof.

Proposition 2: Suppose at each iteration n, the norm of the error between suboptimal $\bar{\mathbf{b}}(\mathbf{u}(n)$ and $\mathbf{b}(\mathbf{u}(n))$ is bounded, i.e., $||\bar{\mathbf{b}}(\mathbf{u}(n)) - \mathbf{b}(\mathbf{u}(n))||_2 \leq \epsilon$, then under the same assumption as in *Theorem 1*, JSSRC converges statistically to within $\gamma\frac{G^2}{2} + \epsilon log(1 + SNR_{max})$ of the optimal, where $SNR_{max} = \max_l\{SNR_l\}$.

Proof: In *Lemma 1* we prove that when $||\bar{\mathbf{b}}(n) - \mathbf{b}(n)||_2 \leq \epsilon$, we have

$$||\bar{\mathbf{g}}(n) - \mathbf{g}(n)||_2 \leq \epsilon log(1 + SNR_{max})$$

Define ϵ_n-subgradient at the *nth* iteration as $\bar{\mathbf{g}}(n)$, and suppose that at each iteration ϵ_n-subgradient is used instead of the optimal subgradient, then the proof for the statistical convergence using ϵ_n-subgradient employs the same approach as in *Proposition 1*.

Chapter 6

Joint Transport, Routing and Spectrum Sharing in Frequency-Agile Radio Networks

In this chapter, we present a distributed algorithm TRSS for the joint transport, routing and spectrum sharing problem for multihop frequency-agile radio networks. By taking advantage of frequency-agile radio's capability to dynamically reconfigure central frequency and channel width, TRSS with appropriate link layer scheduling is able to achieve the optimal spectrum sharing policy that maximizes transport layer utility.

The rest of this chapter is organized as follows. In Section II, we describe our system model. In Section III, we present the mathematical formulation for the joint transport, routing and Spectrum Sharing problem and derive a dual-driven optimizer based on subgradient projection method and dual decomposition. In Section IV, we present the design of TWSR. Corresponding numerical examples and simulation results are shown in Section V. We discuss related work in Section VI and discuss some extensions for TRSS in Section VII. We conclude our work in Section VIII.

6.1 System Model

We consider a general wireless network that is represented by a directed graph, $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, where each vertice $n \in \mathcal{N}$ corresponds to a node in the network, and each directed edge $l \in \mathcal{L}$ corresponds to a communication link. A link $l = (i, j)$ is in $\mathcal{L}$ if and only if node j can successfully receive and decode signal from node i, i.e., j is in the transmission range of node i. Let $E^I(l)$ denote the set of nodes that can cause interference to link l. Link $\hat{l}$ can cause interference to link l if and only if the receiver of link l is in the interference range of the transmitter of link $\hat{l}$. We assume that a radio cannot transmit and receive simultaneously on the same band. This means that $(i, j) \in E^I((j, i))$ implies $(j, i) \in E^I((i, j))$, i.e., link (i, j) and link (j, i) are mutually interfering with each other if they are using the same spectrum band.

We let $\mathbf{s} = \{(b_l, w_l)\}_{l \in \mathcal{L}}$ denote the spectrum sharing policy that specifies the width b_l and central frequency ω_l of spectrum bands each link l can use. There is an upper bound $\tilde{b}_l \in [0, \infty)$ on each b_l, which is the largest amount of spectrum that can be allocated to a single link. The upper and lower bound for the central frequency of each link l is denoted as ω_l^{max} and ω_l^{min}. In DSA networks, $\tilde{\mathbf{b}} = [\tilde{b}_1, \tilde{b}_2, ..., \tilde{b}_{|\mathcal{L}|}]$ corresponds to the spans of the white spaces that are available to secondary users. ω_l^{max} and ω_l^{min} are the boundary of the white space to link l.

Let $\boldsymbol{\mu}(\mathbf{s}) = \{\mu_l(\mathbf{s})\}$ denote a rate vector that corresponds to a feasible MAC scheduling with spectrum policy $\mathbf{s}$. Element $\mu_l(\mathbf{s})$ is defined as follows.

$$\mu_l(\mathbf{s}) = \begin{cases} c_l(\mathbf{s}) & \text{if } l \text{ is active in } \boldsymbol{\mu}(\mathbf{s}), \text{ i.e. if } b_l > 0 \text{ in s;} \\ 0 & \text{otherwise.} \end{cases}$$

We let $\hat{\Gamma}(\mathbf{s})$ denote a bounded region in the $|\mathcal{L}|$ dimensional real space, representing the set of all $\boldsymbol{\mu}$ that are attainable in a given networking setting and spectrum sharing policy $\mathbf{s}$. Note that $\hat{\Gamma}$ is not necessarily convex in general. We expand $\hat{\Gamma}(\mathbf{s})$ to its convex hull $\Gamma(\mathbf{s})$, i.e., $\Gamma(\mathbf{s}) = Conv\{\hat{\Gamma}(\mathbf{s})\}$. Through timesharing, all interior point of $\Gamma(\mathbf{s})$ is attainable. In

DSA network, since $\mathbf{s}$ can be dynamically configured, the set of all feasible rate vectors Γ is thus the union of $\Gamma(\mathbf{s})$, i.e., $\Gamma = \bigcup_{s\in\Pi}\{\Gamma(\mathbf{s})\}$.

With Γ capturing the possible space for MAC layer scheduling corresponding to feasible spectrum allocation, we can further model the relationship between spectrum allocation and routing of flows as follows. We let $A_i^{(k)}$ denote a flow session that is originated at node i and destined for node k with session rate y_i^k. The set of all flow sessions is denoted as $\mathcal{A}$. A fluid traffic model is used and multicommodity flow variables $\mathbf{x}(t) = \{x_l^{(d)}(t)\}$ is employed to describe routing of flows in the network, where each element $x_l^{(d)}(t)$ corresponds to the amount of traffic delivered over link l towards destination d in a unit time.

As in [34],[47] and [48], we define the capacity region Λ of the network as the set of all session rates matrices $(y_i^{(k)})$ for which there exist a multicommodity flow matrix $(x_{i,j}^{(k)})$ that satisfies the following.

(i) $[\sum_k x_l^k]_l \in \Gamma$, where $x_l^{(k)} \geq 0, \forall l \in \mathcal{L}, k \in \mathcal{N}$.

(ii) For each $i \in \mathcal{N}$, and $k \neq i$

$y_i^{(k)} + x_{\text{into}(i)}^{(k)} \leq x_{\text{out}(i)}^{(k)}$

where

$x_{\text{into}(i)}^{(k)} = \sum_{(j,i)\in\mathcal{L}} x_{(j,i)}^k$, $x_{\text{out}(i)}^{(k)} = \sum_{(i,j)\in\mathcal{L}} x_{(i,j)}^k$

$x_{\text{into}(i)}^{(k)}$ denotes the amount of traffic that is destined to node k and flows into node i. $x_{\text{out}(i)}^{(k)}$ denotes the amount of traffic that is destined to node k and flows out of node i. Condition (i) captures the interference and link capacity constraints. Condition (ii) is standard flow-conservation constraint.

Time is slotted with slots normalized to integral units $t \in \{0, 1, 2, ...\}$. In every time slot, an optimizer determines a spectrum sharing policy $\mathbf{s} \in \mathcal{S}$ for the network.

We assume that transmitters employ spectrum reuse only when their intended receivers are not within the interference range of any other transmitter that are using the same bands.

Since all interferers are either inactive or are using different bands, the capacity c_l of a link l using b_l amount of spectrum is simply $b_l log(1 + \frac{p_l}{p_N})$, where p_l and p_N are the signal power and the noise power received at the receiving node of link l respectively. We assume that transmission power and ambient noise remain relatively constant during each execution cycle of TRSS operation and changes at a slower rate than the convergence time of the TRSS adaptation algorithm. Consequently, the capacity of a link l becomes only a function of allocated spectrum amount b_l, i.e., $c_l = b_l log(1+SNR_l)$, where SNR_l is the signal-to-noise ratio (SNR) at the receiver of link l. The link capacity vector $\mathbf{c} = [c_l]_{|\mathcal{L}|\cdot 1}$ is thus a function of the link spectrum width vector $\mathbf{b}$.

6.2 Problem Statements and Solutions

The design goal of TRSS algorithm is to jointly adapt transport, routing and MAC layer decisions to achieve fair and efficient spectrum resource allocation. This design goal can be summarized in the following utility maximization problem:

$$\max_{\mathbf{y}\in\Lambda} \sum_{i,k} U(y_i^{(k)}), \tag{6.1}$$

where the capacity region Λ is determined by the space of feasilble spectrum allocations as defined in Section II. The utility function for each flow is defined in the general form of $U(y_{i,}^{(k)})$. Assumptions on the utility functions are

C1. $U(\cdot)$ are non-decreasing, strictly concave, and twice continuously differentiable on $[0, \max_{l,s}\{c_l(s)\}]$,

C2. There exists $\alpha_i^k > 0$ such that $-\frac{1}{U''(\cdot)} \leq \alpha_i^{(k)}, \forall y_i^{(k)} \in [0, \max_{l,s}\{c_l(s)\}], \forall i, k \in \mathcal{N}$.

This utility function allows vast flexibility for achieving different types of fairness definitions, including proportional fairness and max-min fairness. In the sequel, we refer to (6.1) as *the primal problem P*. With assumption (C1) and (C2), the primal problem is convex and have

a unique optimizer $\mathbf{y}^*$, i.e. $\mathbf{y}^* = argmax_{\mathbf{y}\in\Lambda} \sum_{i,k} U(y_i^{(k)})$ which corresponds to the most fair and efficient feasible spectrum allocation.

To get a distributed solution, we resort to the dual driven decomposition approach. Consider the dual to the primal problem P:

$$\min_{u_i^{(k)}\geq 0, \forall i,k\in\mathcal{N}} D(\tilde{\mathbf{u}}) \tag{6.2}$$

with partial dual function

$$D(\tilde{\mathbf{u}}) = \max_{y_i^{(k)}\geq 0, \mathbf{x}\in\Gamma} \sum_{i,k} U(y_i^{(k)}) - \sum_{i,k} u_i^{(k)}(y_i^{(k)} + x_{\text{into}(i)}^{(k)} - x_{\text{out}(i)}^{(k)}).$$

Here, we relax the flow conservation constraints by introducing Lagrangian multiplier

$$\tilde{\mathbf{u}} = [u_i^{(k)}]_{i,k}. \tag{6.3}$$

The partial dual function can be further decomposed into the following two subproblems

$$D_1(\tilde{\mathbf{u}}) = \max_{y_i^{(k)}\geq 0} \sum_{i,k} (U(y_i^{(k)}) - y_i^{(k)} u_i^{(k)}) \tag{6.4}$$

and

$$D_2(\tilde{\mathbf{u}}) = \max_{\mathbf{x}\in\Gamma} \sum_{i,k} u_i^{(k)}(x_{\text{out}(i)}^{(k)} - x_{\text{in}(i)}^{(k)}). \tag{6.5}$$

These decompositions transform the joint spectrum sharing, routing and transport problem into three separate optimization problems: D_1, D_2 and the master dual problem as represented by Eq. (6.2). These three optimization problems are correlated through the lagrangian multiplier $\tilde{\mathbf{u}}$. $D_1(\tilde{\mathbf{u}})$ is a standard rate control problem which can be solved at the transport layer. $D_2(\tilde{\mathbf{u}})$ is a link layer resource allocation problem that maximizes the weighted sum of flow differential (outflow minus inflow) at each node. In the following, we first solve problem D_1 in Subsection A. We discuss the solution of problem D_2 in Subsection B. The updating algorithm for the master problem is presented in Subsection C.

6.2.1 Solving the subproblem D_1

D_1 can be decomposed as

$$D_1(\tilde{\mathbf{u}}) = \sum_{i,k} \left[\max_{y_i^{(k)} \geq 0} \quad \{U(y_i^{(k)}) - u_i^{(k)} y_i^{(k)}\} \right]$$

From the above equation, it can be seen that the optimal solution to D_1 can be easily obtained by distributed algorithms at the transport layer. Each source node of a session only needs to know local price $u_i^{(k)}$ to calculate the optimal rate for this session. The unique maximizer of D_1 is

$$y_i^{(k)} = U'^{-1}(u_i^{(k)}). \tag{6.6}$$

Alternatively, we can let the session rate allocation converge to the optimal value of D_1 by using iterative algorithm, such as TCP-Vegas in [38].

6.2.2 Solving the subproblem D_2

Observe that we have the following identity:

$$\begin{aligned} &\max_{\mathbf{x} \in \Gamma} \sum_{i,k} u_i^{(k)} (x_{\text{out}(i)}^{(k)} - x_{\text{in}(i)}^{(k)}), \\ &\qquad = \max_{\mathbf{x} \in \Gamma} \sum_{i,j} x_{i,j}^{(k^*)} \max_k \{u_i^{(k)} - u_j^{(k)}\} \end{aligned} \tag{6.7}$$

where $k^* = arg\max_k \{u_i^{(k)} - u_j^{(k)}\}$. If we interpret the price used in the dual problem as the queue length (or scaled queue length), it then becomes apparent that D_2 can be solved in two steps as follows.

(i) For each link $(i,j) \in \mathcal{L}$, find the maximum queue backlog among all flow sessions $k \in \mathcal{K}$ and assign a weight

$$\begin{aligned} u_{i,j} = & \qquad \max_k \{u_i^{(k)} - u_j^{(k)}\} \\ = & \qquad u_{i,j}^{(k^*)} \end{aligned} \tag{6.8}$$

to that link.

(ii) Solve problem D_2 in the reduced form as

$$\max_{\mathbf{x}\in\Gamma} \sum_{i,j} x_{i,j} u_{i,j} \quad , \tag{6.9}$$

where $x_{i,j} = x_{i,j}^{(k^*)}$.

Step (i) motivates us to design the routing module of TRSS using a queue-length-based algorithm known as the *Back-Pressure* Routing (BPR) introduced by [34] for wireless networks. The detail of the routing algorithm is provided in the following definition.

Definition (Back-Pressure Routing): At slot t, for each link $l = (i, j) \in \mathcal{L}$ find

$$u_{i,j}[t] = \max_k \{u_i^{(k)}[t] - u_j^{(k)}[t]\} \tag{6.10}$$

and

$$k_{i,j}[t] = \arg\max_k \{u_i^{(k)}[t] - u_j^{(k)}[t]\}. \tag{6.11}$$

With $u_{i,j}$ and $k_{i,j}$ determined, we can proceed to solve problem (6.9). Note that problem (6.9) may have multiple solutions. We choose $\mathbf{x}[t]$ from the set of optimal solutions and then serve the queue holding packets destined for node $k_{i,j}[t]$ over link (i, j) at rate $x_{i,j}[t]$, i.e.,

$$x_{i,j}^{k(i,j)}[t] = x_{i,j}[t] \tag{6.12}$$

and

$$x_{i,j}^{\hat{k}}[t] = 0, \quad \forall \hat{k} \neq k(i, j). \tag{6.13}$$

The essential idea of BPR as described in (6.12) and (6.13) is to always let the most backlogged flow on each link get to use the link in each iteration.

In the above, we transform the subproblem D_2 into a simpler form (6.9) using BPR. However, we still cannot solve it because the constraint set Γ is in a collective form. In the following we expand Γ. To proceed, we first give the definitions of three important formulation parameters.

Denote the set of links that can interfere with link l as $E^I(l)$. We define a *spectrum contention circle* $\mathcal{C}(l)$ for each link $l \in \mathcal{L}$ as $\mathcal{C}(l) = E^I(l) \bigcup l$. Note that, in addition to the contention circle $\mathcal{C}(l)$, a link l may also belong to other contention circles.

For a systematic approach to describe contention circles, we define spectrum contention matrix $\mathbf{A}^{|\mathbf{L}|\cdot|\mathbf{L}|}$ as follows.

$$A_{l\hat{l}} = \begin{cases} 1 & \hat{l} = l \text{ or } \hat{l} \in E^I(l), \text{i.e., } \hat{l} \text{ can produce interference} \\ & \text{to link } l\text{'s receiver;} \\ 0 & \text{otherwise.} \end{cases}$$

The nonzero elements of the lth row of $\mathbf{A}$ correspond exactly to all the members in the contention circle of link l. A nonzero element in the $\hat{l}th$ row of column l means link l belongs to the contention circle of link $\hat{l}$. If $A_{l\hat{l}} = 1$, link l and $\hat{l}$ can not be active using the same spectrum band simultaneously because of interference.

The sum of the lth column equals the number of contention circles that a link l belongs to. We define it as link l's *constraint degree* ϱ_l, which can be expressed as

$$\varrho_l = |\hat{l} : l \in \mathcal{C}(\hat{l})|. \tag{6.14}$$

In formulations, ϱ_l corresponds to the number of scheduling constraints on link l and is used as a measure of frequency scheduling difficulty for link l. This will be detailed when we discuss the design of TWSR in Section IV.

With the above model and definitions, Problem (6.9)'s constraint Γ can be expanded and D_2 can be represented in explicit constraints as

$$\begin{aligned} \max_{x_l \geq 0, l \in \mathcal{L}} \quad & \sum_l x_l u_l \\ \text{subject to} \quad & \\ & x_l \leq c_l \\ & c_l = b_l \log(1 + SNR_l) \\ & b_l + \sum_{\hat{l} \in E_l^I} b_{\hat{l}} \leq \tilde{b}_l \\ & b_l \geq 0, c_l \geq 0, \end{aligned} \tag{6.15}$$

where the Lagrangian multiplier u_l is defined as in (6.10).

To get a more compact representation of the formulation, let $\mathbf{b}=[b_1, ..., b_{|\mathcal{L}|}]^T$,$\mathbf{c}=[c_1, ..., c_{|\mathcal{L}|}]^T$ and $\tilde{\mathbf{b}}=[\tilde{b}_1, ..., \tilde{b}_{|\mathcal{L}|}]^T$ denote the spectrum width allocation vector, the link capacity vector and the spectrum availability vector respectively. Let $\mathbf{N}$ be a $|\mathcal{L}| \cdot |\mathcal{L}|$ diagonal matrix where $n_{ll} = \log(1 + SNR_l)$ and $n_{l\hat{l}} = 0$ for $l \neq \hat{l}$. Let $\mathbf{x} = [x_1, ..., x_{|\mathcal{L}|}]^T$ denote the vector of session rates. We can obtain an equivalent but more compact form of the formulation as:

$$\begin{aligned} \max_{\mathbf{x} \geq \mathbf{0}} \quad & \mathbf{u}^T \mathbf{x} \\ \text{subject to} \quad & \\ & \mathbf{x} \leq \mathbf{c} \\ & \mathbf{c} = \mathbf{N}\mathbf{b} \\ & \mathbf{A}\mathbf{b} \leq \tilde{\mathbf{b}} \\ & \mathbf{b} \geq \mathbf{0}, \mathbf{c} \geq \mathbf{0}. \end{aligned} \tag{6.16}$$

Note that the link indexed Lagrangian multipliers $\mathbf{u}$ are the vector form of $u_{i,j}$ as defined in (6.10) and is different from the node-flow-pair indexed $\tilde{\mathbf{u}}$ as defined in (6.3) which is only used in the definition of primal problem P in Section III.

We refer to the above formulation as $\tilde{D}_2$. A centralized algorithm to solve $\tilde{D}_2$ is straightforward to devise and easy to solve. However, a distributed algorithm is more desirable and non-trivial. In the following, we use dual decomposition and subgradient method to obtain a distributed algorithm to solve $\tilde{D}_2$.

Consider the dual of $\tilde{D}_2$

$$\min_{\mathbf{v},\mathbf{w}\geq\mathbf{0}} L(\mathbf{v},\mathbf{w}) \tag{6.17}$$

with lagrangian dual function

$$L(\mathbf{v},\mathbf{w}) = \max_{\mathbf{x},\mathbf{b}\leq\mathbf{0}} \mathbf{u}^T\mathbf{x} - \mathbf{w}^T(\mathbf{x}-\mathbf{N}\mathbf{b}) - \mathbf{v}^T(\mathbf{A}\mathbf{b}-\tilde{\mathbf{b}}), \tag{6.18}$$

where $\mathbf{w}$ and $\mathbf{v}$ are the Lagrangian multipliers.

Due to the linearity of $\tilde{D}_2$, the dual problem may have multiple optimal points with the same dual objective value. By solving (6.17), the primal optimal variables are not immediately available. To remove this complication, we add two small regularization term in $\tilde{D}_2$, namely, quadratic term $\sigma\mathbf{b}^T\mathbf{b}$ and $\sigma\mathbf{x}^T\mathbf{x}$. Here, σ is very small positive constant. Subsequently, we have

$$L(\mathbf{v},\mathbf{w}) = \max_{\mathbf{x},\mathbf{b}\leq\mathbf{0}} \mathbf{u}^T\mathbf{x} - \mathbf{w}^T(\mathbf{x}-\mathbf{N}\mathbf{b}) - \mathbf{v}^T(\mathbf{A}\mathbf{b}-\tilde{\mathbf{b}}) + \sigma\mathbf{b}^T\mathbf{b} + \sigma\mathbf{x}^T\mathbf{x}. \tag{6.19}$$

Using gradient projection method, we can derive an iterative algorithm to solve the regularized dual problem as follows:

$$\mathbf{b}^* = \{\frac{1}{2\sigma}(\mathbf{N}\mathbf{w} - \mathbf{A}^T\mathbf{v}[t])\}^+, \tag{6.20}$$

$$\mathbf{x}^* = \{\frac{1}{2\sigma}(\mathbf{u} - \mathbf{w}[t])\}^+, \tag{6.21}$$

$$\mathbf{v}[t+1] = \{\mathbf{v}[t] + \beta(\mathbf{A}\mathbf{b}^* - \tilde{b})\}^+, \tag{6.22}$$

$$\mathbf{w}[t+1] = \{\mathbf{w}[t] + \beta(\mathbf{x}^* - \mathbf{N}\mathbf{b}^*\}^+, \tag{6.23}$$

where β is a positive stepsize. For each link, the iterative algorithm in (6.20)-(6.23) can be expressed as:

$$b_l^* = \{\frac{1}{2\sigma}\{\log(1+SNR_l)w_l - \sum_{\hat{l}:l\in\mathcal{C}(\hat{l})} v_{\hat{l}}[t]A_{\hat{l}l}\}\}^+, \tag{6.24}$$

$$x_l^* = \{\frac{1}{2\sigma}(u_l - w_l[t])\}^+, \tag{6.25}$$

$$v_l[t+1] = \{v_l[t] + \beta(\sum_{\hat{l}\in\mathcal{C}(l)} (A_{l\hat{l}}b_{\hat{l}}[t+1]) - \tilde{b}_l\}^+, \tag{6.26}$$

$$w_l[t+1] = \{w_l[t] + \beta(x_l^* - b_l^*\log(1+SNR_l))\}^+. \tag{6.27}$$

Observe that, given $\mathbf{u}$, all computation involved in (6.24)-(6.27) requires only local information to complete. For example, according to (6.24) and (6.26), to compute b_l^* and v_l , link l only needs to collect the values of $v_{\hat{l}}$ and $b_{\hat{l}}$ from the links that can interfere with link l. In (6.25) and (6.27), to compute x_l^* and w_l, only the contention price at the link itself is needed.

The convergence of the above iterative algorithms to the optimal solution follows from the more general case of the convergence of gradient projection method which is well established in literature [88]. More specifically, define $\bar{a}_1 = \arg\max_{l\in\mathcal{L}}\{|\mathcal{C}_l|\}$, and $\bar{a}_2 = \arg\max_{l\in\mathcal{L}}\{\varrho_l\}$. The range of the step size with which the algorithm converges can be defined as in $Theorem$ 1 of [39]:

Proposition 1: Gradient projection iteration (6.24)-(6.27) converges when $\beta, \sigma, \bar{a}_1, \bar{a}_2$ satisfy the following equation,

$$0 \le \beta \le \frac{4\sigma}{\bar{a}_1\bar{a}_2}. \tag{6.28}$$

6.2.3 Solving the master dual problem

Let $\mathbf{x}^*$ denote the optimal link flow solution obtained by solving (6.17). We can then assign the multicommodity flow as:

$$x_{i,j}^k[t] = \begin{cases} x_{(i,j)} & \text{if } k = k_{i,j}[t] \\ \\ 0 & \text{otherwise.} \end{cases}$$

Now we come to solve the master dual problem (6.2). Since the dual function is not necessarily differentiable, we use projection subgradient method to solve (6.2). It is easy to verify that

$$x_{\text{out}(i)}^{(k)} - (y_i^{(k)} + x_{\text{into}(i)}^{(k)}) \tag{6.29}$$

is a subgradient of $D(\tilde{\mathbf{u}})$ at node i for flow k in time slot t. The projection subgradient method based adaption can then be expressed as

$$u_i^{(k)}[t+1] = \{u_i^{(k)}[t] + \gamma(y_i^{(k)} + x_{\text{into}(i)}^{(k)} - x_{\text{out}(i)}^{(k)})\}^+, \tag{6.30}$$

where γ is a positive stepsize. This updating algorithm achieves optimal when the stepsize is sufficiently small.

6.2.4 Convergence of Phase I algorithm of TRSS

Let $\mathbf{g}(t)$ denote the vector of all subgradients. If we assume the norm of the subgradients of the dual function is bounded, i.e., there exists a constant G such that

$$||\mathbf{g}(t)||_2 \leq G, \tag{6.31}$$

then the convergence of the Phase I algorithm is guaranteed by the following proposition.

Proposition 2: If (6.31) holds, the aggregate utility $\sum_{i,k} U(y_i^{(k)})$ generated by the iterative computation of TRSS converges statistically to within $\frac{\gamma G^2}{2}$ of the optimal value.

Proof: the proof is provided in Appendix.

6.2.5 Understanding the computational goal of TRSS

Equations (6.6),(6.24)-(6.27) and (6.30) form the mathematical foundation of TRSS phase I algorithm for the joint transport, routing and spectrum width allocation problem **P**. The pseudo code for the Phase I algorithm of TRSS is summarized in *Algorithm* 1 Before exploring any further, let us take a closer look at Phase I algorithm to gain some insight on the physical meaning of this Phase I algorithm. The key is to understand the functions of the Lagrangian multipliers $\tilde{\mathbf{u}}$, **u,** **v** and **w**.

In our formulation, $u_i^{(k)}$ can be interpreted as the scaled queue length at node i for packets destined to node k. The transport layer at each source node determines the flow rate destined for node k based on the queue length of the flow at the source node. This is distinct from link-centric formulations where the source node of a flow needs queue length information of all links along the flow's path. This is because back-pressure routing is able to indirectly carry this information hop-by-hop back to the source node by the back-pressure load-balance mechanism. Indeed, this backpressure information is carried by the link level queue backlog $u_i^{(k)} - u_j^{(k)}$ for each link as in (6.7). The routing layer then balances the load according to the backlog information. The reason that the routing module chooses the most backlogged flow on each link is to maximize the flow utility. To this end, $u_i^{(k)} - u_j^{(k)}$ can be interpreted as the gain that the network can get by delivering a unit traffic of flow k over link (i, j). Note that the aggregate flow on each link per time slot is upper-bounded by link's capacity. It then becomes apparent that the optimal choice is to dedicate all the capacity to the most profitable flow on that link, i.e. the flow with the largest backlog.

By interpreting $u_i^{(k)}$ as scaled queue length, the master dual problem's iterative algorithm in (6.30) can be treated as an evaluation of queue length for the next interval. Indeed, if we let stepsize $\gamma = 1$, Eq. (6.30) is the exact queue length for packets at node i destined for k. We need to mention that despite the clear physical meaning, the scalar stepsize is normally

Algorithm 3: The TRSS Algorithm

Input : $\mathcal{G}(\mathcal{N},\mathcal{L}), \mathbf{y}(t_0)$

Output: **b, f,x, y**

begin

Initialize (**b,x, y, u, v, w**)

while *variation of* **u** *is larger than the termination threshold* **do**

while *variation of* **y** *is larger than termination threshold* **do**

Every session source $s(i,k) \in S$ update $y_i^{(k)}$ as follows,

$y_i^{(k)} \leftarrow \left[y_i^{(k)} = U'^{-1}(u_i^{(k)})\right]^+$

end

For each link $(i,j) \in \mathcal{L}$, find $u_{i,j}[t]$ and $k_{i,j}[t]$ as follows

$u_{i,j}[t] = \max_k\{u_i^{(k)}[t] - u_j^{(k)}[t]\}$

$k_{i,j}[t] = \arg\max_k\{u_i^{(k)}[t] - u_j^{(k)}[t]\}$

while *variation of* $\mathbf{w}, \mathbf{v}$ *is larger than the termination threshold* **do**

Every link $l \in L$ updates x_l and b_l according to the following equations

$x_l^* \leftarrow \{\frac{1}{2\sigma}(u_l - w_l[t])\}^+$

$b_l^* \leftarrow \{\frac{1}{2\sigma}\{\log(1+SNR_l)u_l - \sum_{\hat{l}:l\in E_{\hat{l}}^I} v_{\hat{l}}[t]\}^+$

Every link $l \in \mathcal{L}$ updates w_l and v_l as

$w_l[t+1] \leftarrow \{w_l[t] + \beta(x_l^* - b_l^*\log(1+SNR_l))\}^+$

$v_l[t+1] \leftarrow \{v_l[t] + \beta(\sum_{\hat{l}\in E_l^I} b_{\hat{l}}[t+1] - \tilde{b}_l)\}^+$

end

Assign flow routing according to $x_{i,j}^*$ as follows,

$$x_{i,j}^k[t] = \begin{cases} x_{(i,j)} & \text{if } k = k_{i,j}[t] \\ 0 & \text{otherwise.} \end{cases}$$

Every node $i \in \mathcal{N}$ update $u_i^{(k)}$ for each flow as follows,

$u_i^{(k)}[t+1] = \{u_i^{(k)}[t] + \gamma(y_i^{(k)} + x_{\text{into}(i)}^{(k)} - x_{\text{out}(i)}^{(k)})\}^+$

$(\bar{\mathbf{b}}, \bar{\mathbf{f}}) \leftarrow TWSR(\mathbf{b}^*, \mathbf{f}^*)$

end

end

chosen to be smaller than 1. This is because when γ is large, the TRSS algorithm tends to be aggressive in queue length control, which may cause fluctuation of in the value of $u_i^{(k)}$ and delay the convergence of TRSS.

w_l is interpreted as the link price for congestion control. In (6.27), link l adjusts its link price based on its congestion level, which is captured by $x_l - c_l$. When link l is overloaded so that the traffic goes through link l (i.e., x_l) is larger than link l's capacity c_l, link l increases its price u_l to encourage sessions that route through it to reduce their rates, which prevents congestion at link l. When link l is lightly loaded so that the traffic goes through link l is smaller than link l's capacity, link l decreases its price u_l to encourage sessions that route through it to increase their rates, so that link l's spectrum utilization can be improved.

Treating $u_i^{(k)} - u_j^{(k)}$ as link price, the subproblem D_2 becomes the problem of maximizing the aggregated link layer revenue through allocation of spectrum. Intuitively, links with higher prices are likely to be allocated with more spectrum as they generate more revenues. This makes sense since a link with a higher price $u_i^{(k)} - u_j^{(k)}$ is the link that has heavier traffic and hence needs additional spectrum to alleviate its congestion. The optimal spectrum allocation, however, is not that simple. Due to interference among links, when a link l is allocated with more spectrum, the available spectrum for its neighboring links ($\bar{l}$ in $\mathcal{C}(l)$) is decreased. In other words, there is an inevitable contention among nodes for spectrum. This contention is captured by the lagrangian multiplier $v_{\bar{l}}$. The update algorithm of Eq. (6.26) increases the contention price v_l of link l when link l's local available spectrum $\tilde{b}_l$ (spectrum left by primary users) becomes smaller than the spectrum demands from the links in its contention circle (s.t. $\sum_{\hat{l}\in\mathcal{C}(l)} (A_{l\hat{l}} b_{\hat{l}}(m+1)) > \tilde{b}_l$), discouraging these links in $\mathcal{C}(l)$ from demanding more spectrum from link l. When there is still unallocated available spectrum at link l (s.t. $\sum_{\hat{l}\in\mathcal{C}(l)} (A_{l\hat{l}} b_{\hat{l}}(m+1)) < \tilde{b}_l$), link l decreases its contention price v_l so that other links in its contention circle $\mathcal{C}(l)$ can use this unallocated spectrum at lower prices. Based on this rationale, Eq. (6.24) essentially says that the spectrum width allocated to a link l is proportional to its income per unit of allocated spectrum, which is the price $(\log(1+SNR_l)u_l)$ that link l can charge for relaying traffic minus the total cost

$(\sum_{\hat{l}:l\in\mathcal{C}(\hat{l})}(v_{\hat{l}}(m)A_{\hat{l}l}))$ for spectrum usage that link l has to pay to its neighboring links that it can interfere.

6.3 TRSS phase II algorithm, Timing Window based Spectrum Reservation

Phase I of TRSS solves a joint spectrum width allocation and rate control problem, where only the total amount of allocated spectrum for each link is determined. In Phase II, TRSS distributively schedules a feasible spectrum sharing policy for each radio by further determining the central frequency of each link according to the spectrum width allocation vector $\mathbf{b}$ generated in Phase I. The objective of Phase II is to assign each link l a spectrum band with width $\bar{b}_l$ and a center frequency $\bar{f}_l$ without spectrum overlap in any spectrum contention circle, which is called frequency scheduling in this chapter.

Note that the spectrum contention constraint (4) in $\mathbf{P}$ is a sufficient condition for link layer schedulability. Any feasible spectrum width allocation $\mathbf{b}$ generated by the Phase I algorithm corresponds to at least one feasible spectrum sharing policy $(\mathbf{b}, \mathbf{f})$, where $\mathbf{f}$ is the central frequency vector. If we assume a centralized spectrum broker, the scheduling problem is trivial. The optimal policy is achieved immediately when each node follows the command of the central broker. However, in large networks, such a central spectrum broker is not likely to exist. The links only have limited local information and thus is prone to make "bad" decisions on spectrum selection. Hence, an optimal frequency scheduling may not be easily obtained distributively. In the following, we first demonstrate a case where naive scheduling algorithm fails to generate a good policy and then propose a novel distributed Timing Window based Spectrum Sharing (TWSR) algorithm that closely approximates the optimal spectrum sharing policy without requiring centralized spectrum broker.

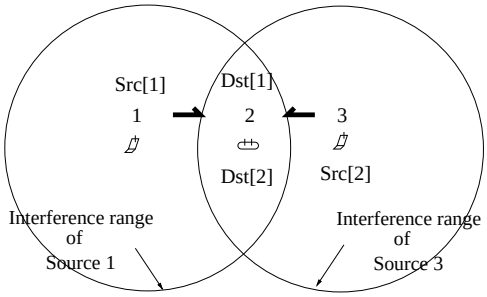


Figure 6.1: "Bad" spectrum allocation may reduce low spectrum efficiency

6.3.1 A case study of distributed link layer scheduling

Consider a simple network of 3 nodes in Fig. 6.1. Suppose there are two active one-hop sessions s_1 and s_2 from node 1 to 2 and 3 to 2, respectively. Each session requires 50MHz spectrum to transmit its traffic. Node 1 and 3 each has one radio and node 2 has two radios. Links (1,2) and (3,2) mutually interfere with each other. Suppose there is an available spectrum whitespace of 100MHz left by primary users and the spectrum whitespace starts from 2.4GHz. The optimal spectrum sharing policy should allocate each link a 50MHz wide spectrum. This is only possible if link (3,2) uses either the first half block of the spectrum whitespace starting from 2.4GHz or uses the second half block starting from 2.45GHz, while the other link uses the remaining half simultaneously. A centralized spectrum broker can easily realize such a spectrum allocation. However, when links make distributed spectrum reservation decision, link (3,2) may first reserve a 50MHz spectrum range starting from 2.425GHz without any trouble. However, this reservation leaves two 25MHz spectrum fragments and link (1,2) can only use one of the spectrum fragment because it has only one radio. Hence, with the limitation on radios per link, distributed spectrum reservation may not result in optimal spectrum scheduling.

6.3.2 Timing Window based Spectrum Reservation

To address the above issue, we propose the Timing Window based Spectrum Reservation (TWSR) algorithm. TWSR works in scenarios where there is only one dedicated radio at a node for each of its link. The spectrum scheduling produced by TWSR approximates the optimal spectrum scheduling.

Based on b*, TWSR produces an approximate spectrum scheduling $\bar{\mathbf{b}}$ and $\bar{\mathbf{f}}$. The design of TWSR is based on the observation that the more contention circles that a link l belongs to, the more difficult to find an appropriate spectrum fragment with size b_l for link l as it takes coordination among more links. Hence, those links that are included in more contention circles of other links should be given priority in spectrum reservation. Recall the definition of constraint degree ϱ in Section III which is the number of contention circles that a link l belongs to.

Based on the aforementioned rationale, after link l obtains its new spectrum width b_l^* in Phase I, link l computes a timing window based on its constraint degree ϱ_l, where the lower bound of the time window is w_l^{lb}=δ/ϱ_l and the upper bound is w_l^{ub}=$\delta/(\varrho_l - 1)$. Here, δ is a scalar determined by timing granularity of the network. Link l then randomly picks a time between w_l^{lb} and w_l^{ub} to declare to its interfering links its spectrum reservation. The spectrum reserved by link l is picked as the lowest unreserved spectrum fragment that has size b_l. If such a spectrum fragment does not exist, link l reserves the largest spectrum fragment available. The spectrum reservation process is a first-come-first-serve process. A link l only reserves a spectrum fragment that is not previously reserved by its interfering links in $\mathcal{C}(l)$. This first-come-first-serve scheduling and non-overlapping declaration mechanism ensures that links that are more difficult to schedule declare their spectrum reservation earlier and, hence, have more chance to get exactly the spectrum size that they demand.

The spectrum width allocation vector $\bar{\mathbf{b}}$ given by TWSR is generally a suboptimal version of $\mathbf{b}^*$. The loss of bandwidth for link l leads to $(b_l - \bar{b}_l)\log(1 + SNR_l)$ loss of link capacity. Therefore, For $\bar{\mathbf{b}}$, the corresponding flow $\mathbf{x}^*$ in (6.25) may no longer be attainable. We then

scale $\mathbf{x}^*$ to $\bar{\mathbf{x}}$ as follows.

$$\bar{x}_l = \{x_l - (b_l - \bar{b}_l)\log(1 + SNR_l)\}^+. \tag{6.32}$$

TWSR itself is a heuristic algorithm falling into the category of greedy packing problems. However, as shown in Section 6.4, TRSS with TWSR produces results that are very close to the optimal results. Note that the shadow price $\tilde{\mathbf{u}}$ at the end of each master iteration (line 18 in Algorithm I) is updated according to $\bar{\mathbf{x}}$ generated by TWSR instead of the optimal value $\mathbf{x}^*$. As a result, the update in the master iteration can partially correct the approximation error made by $\bar{\mathbf{x}}$. In Theorem 2 in the next section, we give an upper-bound on the approximation error of TRSS with TWSR. It turns out that the combined approximation error of the iterative algorithm is upper bounded by how close the suboptimal schedule $\bar{\mathbf{b}}$ are to the optimal schedule $\mathbf{b}^*$.

The pseudo code for TWSR is summarized in Algorithm 2. The performance of TRSS combin[ed with TWSR is characterized by the following proposition.

Proposition 3: Suppose at each iteration t, the norm of the error between suboptimal $\bar{b}_l[t]$ and $b_l^*[t]$ is bounded, i.e., $|\bar{b}_l[t] - b_l^*[t]| \leq \epsilon$, then under the same assumption as in *Proposition 2*, TRSS with TWSR converges statistically to within $\gamma\frac{G^2}{2} + K\gamma\epsilon log(1 + SNR_{max})$ of the optimal, where $SNR_{max} = \max_l\{SNR_l\}$ and γ is the step size of choice.

proof: The proof is provided in the appendix.

6.4 Numerical Results

We investigate TRSS with a representative network and discuss the implications of the results. More results are summarized in Table 1 due to space limit.

The network utility function is taken as $U(\cdot) = log(\cdot)$. The step size and regularization term

Algorithm 4: Timing Window based Spectrum Reservation

Input : $\mathbf{b}^*(n)$

Output: $\bar{\mathbf{b}}(n), \bar{\mathbf{f}}(n)$

begin
- Compute timing window at each link $w_l^{lb} \leftarrow \delta/\varrho_l$, $w_l^{ub} \leftarrow \delta/(\varrho_l - 1)$
- Each link randomly picks a time point t_l between $[w_l^{lb}, w_l^{ub}]$
- Each link l sets a *timer* T_l with initial value t_l and starts counting down the timer
- **while** *not all timers timeout* **do**
  - **if** $\forall T_k$ *timeout* **then**
    - **while** *not finishing scanning the entire spectrum whitespace left by the primary users* **do**
      - Link k scans for unreserved spectrum fragments starting from the lowest frequency in the spectrum whitespace left by primary users.
      - **if** *link k finds a fragment of width* $b^w \geq b_k^*$ **then**
        - Link k reserves b_k^* amount of spectrum starting from the lowest frequency f^w in this fragment.
        - Link k informs the other links in $\mathcal{C}(k)$ about $\bar{b}_k \leftarrow b_k^*$ and $\bar{f}_k \leftarrow f^w + b_k^*/2$
      - **end**
    - **end**
    - **if** *no spectrum fragment found* **then**
      - Link k reserves the largest available spectrum fragment b^w.
      - Link k informs the other links in $\mathcal{C}(k)$ about $\bar{b}_k \leftarrow b^w$ and $\bar{f}_k \leftarrow f^w + b^w/2$.
    - **end**
  - **end**
- **end**
- output $(\bar{\mathbf{b}}, \bar{\mathbf{f}})$

end

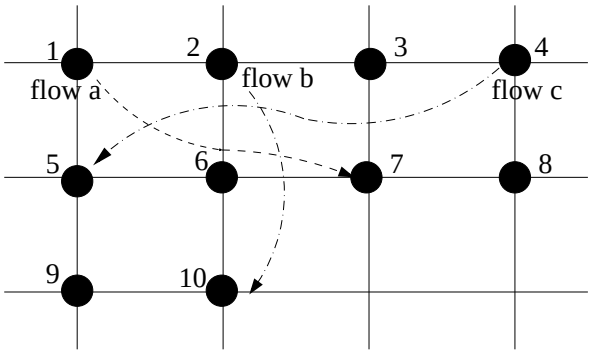


Figure 6.2: Simulated network topology and traffic flows for the TRSS algorithm: Example 1

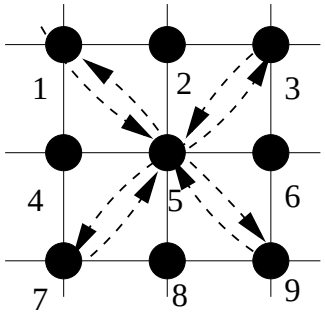


Figure 6.3: Simulated network topology and traffic flows for the TRSS algorithm: in Example 2

are chosen as $\gamma = 0.1$ and $\sigma = 0.1$ respectively in most simulations.

In a representative example, we consider a simple ad hoc network shown in Fig. 6.2 with 10 nodes and 3 flows a (from node 1 to node 7), b (from node 4 to node 5) and c (from node 2 to node 10). Each node contends with all its 1-hop neighbors for the available spectrum. We see in Fig. 6.4 and 6.5 that the aggregate utility and source rates quickly converge to a neighborhood of the fixed point. Due to the non-differentiability of the dual function, source rates oscillate around the optimal. This oscillation is practically smoothed out in Fig. 6.6 where the averaged aggregate utility is shown. We note that TRSS with TWSR produces negligible performance degradation (less than 1%) compared to the TRSS with optimal frequency scheduling. This is not a special case. In Table I, we will see that TRSS with

TWSR is profoundly effective and generally produces optimal or close-to-optimal solutions.

In the second example, we set up a topology to simulate the traffic of a WLAN, where subscriber nodes $1, 2, 3, 4, 6, 7, 8, 9$ communicate with access point nodes 5. In Fig. 6.3 , there are 4 flows from 1, 3, 7, 9 to node 5 and another 4 flows from node 5 to nodes 1, 3, 7, 9. Other settings remain the same as in example 1. As the number of flows increases, the optimization parameters are significantly increased. However, TRSS converges only slightly slower than in Example 1. The difference between TRSS with optimal frequency scheduling and with TWSR remains very small.

In Table 6.4 we summarize the data from ten groups of simulations. In the simulations, various topologies and traffic patterns are used to examine the robustness of TRSS. The maximum network size is 50 nodes. The largest number of flows is 16. As the final results indicate, TRSS converges statistically in all experiments. The comparison between TRSS with optimal frequency scheduling and TRSS with TWSR shows that only an average of approximately 2% of gap is introduced by TWSR, which signals the effectiveness of our design.

6.5 Related Works

In literature [69][70][74], frequency agility is often treated the same as multi-channel communication capability.

As pointed out in [61], SDRs/CRs are much more powerful and flexible than multi-channel radios. While the use of multi-channel model can significantly simplify the algorithm design for CR/SDR networks, it limits the potential of SDR/CR network. In [61], the author use an uneven bandwidth segmentation model for spectrum sharing and formulate

the joint flow routing and spectrum allocation problem as a Mixed Integer and Non-Linear

Table 6.1: More simulation results for TRSS

Table I				
Group #	Optimal		TWSR	
	Aggregate Utility	No. of iterations	Aggregate Utility	No. of iterations
1	0.5274	12	0.5274	15
2	1.4573	46	1.4075	36
3	1.0134	71	1.0132	61
4	1.8012	101	1.7936	106
5	1.0954	431	1.0948	406
6	1.1328	421	1.1048	586
7	1.3683	626	1.3498	836
8	1.1197	811	1.1123	786
9	2.0169	926	1.9234	891
10	0.3481	706	0.3036	751

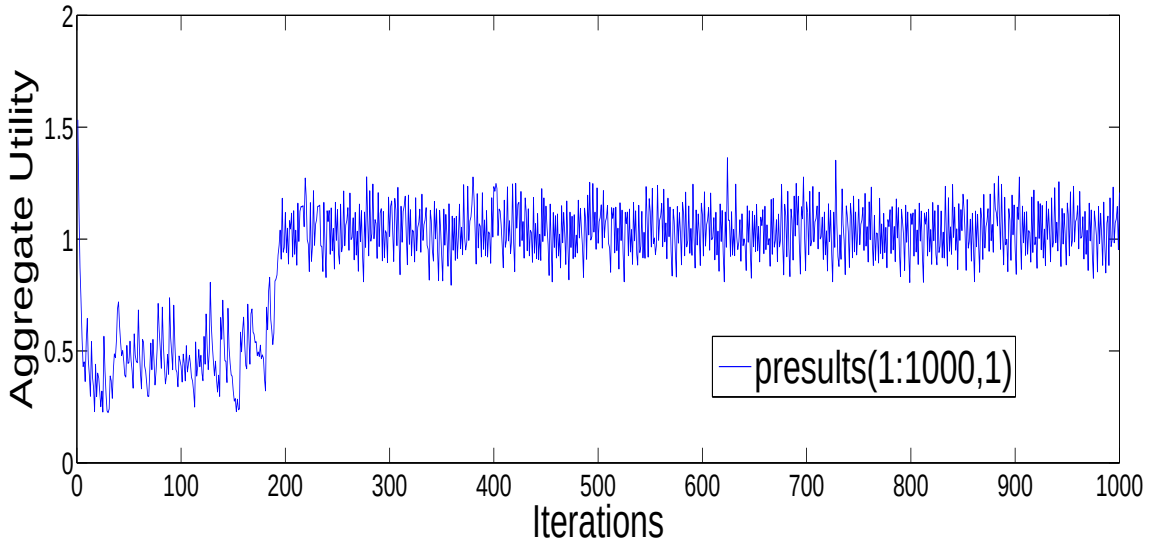


Figure 6.4: TRSS algorithm converges and push the aggregate utility of the network to optimal in Example 1

Programming problem. An algorithm based on Sequential Fixing (SF) is then proposed for the MINLP. A distributed algorithm for the same problem is developed in [62] under similar assumptions. Numerical results show that both algorithms developed in [61] and [62] can achieve near optimal performance. These two schemes, however,

assume known traffic from fixed pairs of nodes (data sessions) and fixed network topology. The flow routing and spectrum allocation algorithms are good for stable and stationary networks. In a dynamic network, however, any change in the network may trigger a rerun of the entire algorithm to get the new near-optimal solution. This can become a problem when the network changes frequently, which is common in cognitive radio and dynamic spectrum access networks. Adaptive algorithms such as JSSRC, in contrast, are more robust in a dynamic environment and are thus more desirable in those scenarios.

On the other hand, in the Internet context, adaptive algorithms for congestion control and traffic rate control have been extensively researched over the last two decades [33]-[37]. Such research provides optimization-based framework for Internet congestion control and

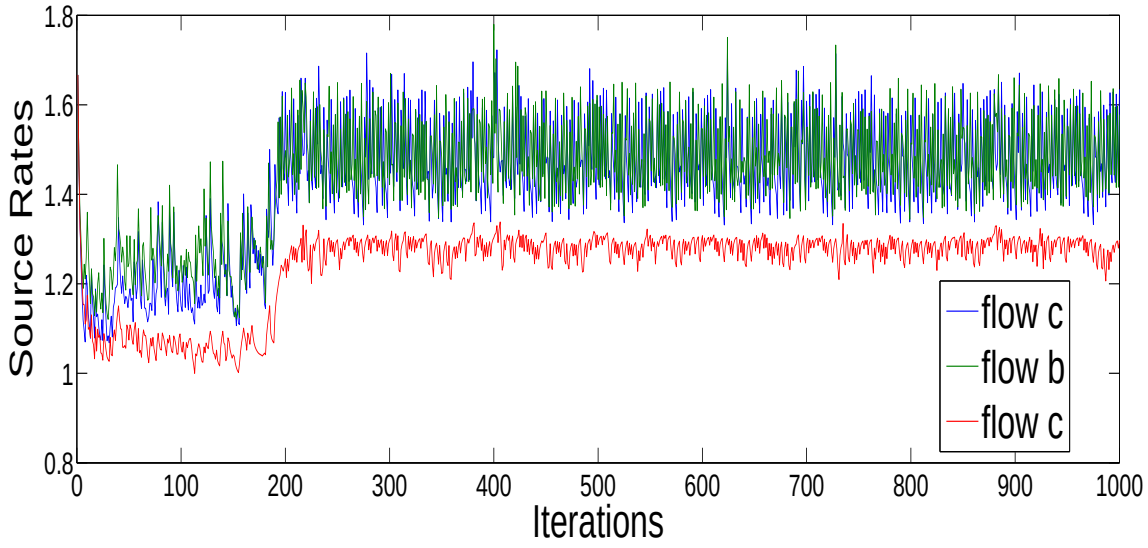


Figure 6.5: TRSS algorithm converges and push the source rates to optimal in Example 1

derives adaptive and distributed solutions through both primal and dual approaches. Recent research results [43]-[50] show that it is possible to model principles of wireless network design under a similar optimization framework. In wireless networks, however, upper layer protocol design is inherently coupled with lower layer properties and vice versa. For example, traffic rate and routing choices at upper layers actually change the existence and magnitude of interference at the physical layer, which in return affects the feasibility and efficiency of upper layer decisions on traffic rate. For this reason, cross-layer designs [43]-[50] are predominant in wireless environment.

6.6 Conclusion

We present TRSS, a joint optimization algorithm for cross-layer resource management in frequency-agile radio based networks. By taking advantage of frequency-agile radio's capability to dynamically reconfigure central frequency and channel width, TRSS with perfect link layer scheduling is able to achieve the optimal spectrum efficiency with regard to max-

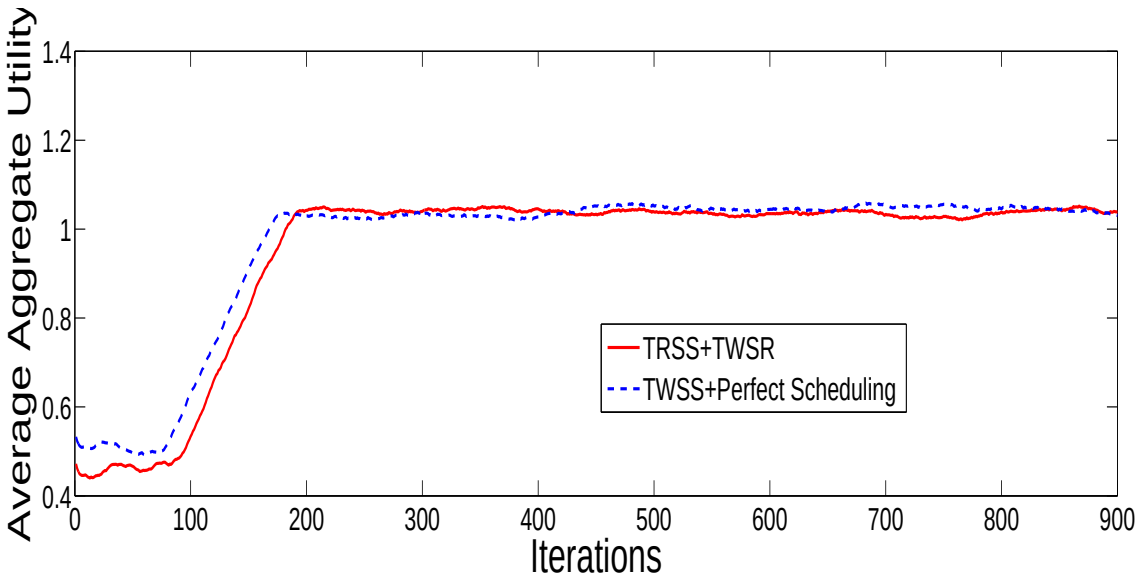


Figure 6.6: Comparing TRSS with optimal frequency scheduling and TRSS with TWSR in Example 1

imizing network utility. For practical implementation, we present a novel timing window based spectrum reservation scheme (TWSR) for link layer scheduling. Simulation results show that TRSS algorithm with TWSR achieves good performance despite TWSR's heuristic nature.

As the preliminary results to a practical joint transport, routing and spectrum sharing adaptation algorithm, the basic framework of TRSS can be extended in various directions. We describe some of the extensions here and briefly discuss how to handle them.

First, to accommodate time-varying channels, a stochastic channel model can be used to describe the network. For example, if we assume the channel state is described by a finite state Markov Chain, then we can define the average capacity region $\bar{\Lambda}$ as a function of these states and TRSS can be straighforwardly extended to maximize network utility over these average capacity region.

Second, TRSS can be extended to work asynchronously. Actually, the dual decomposition is well suited for asynchronous design. Convergence and optimality analysis for asynchronous

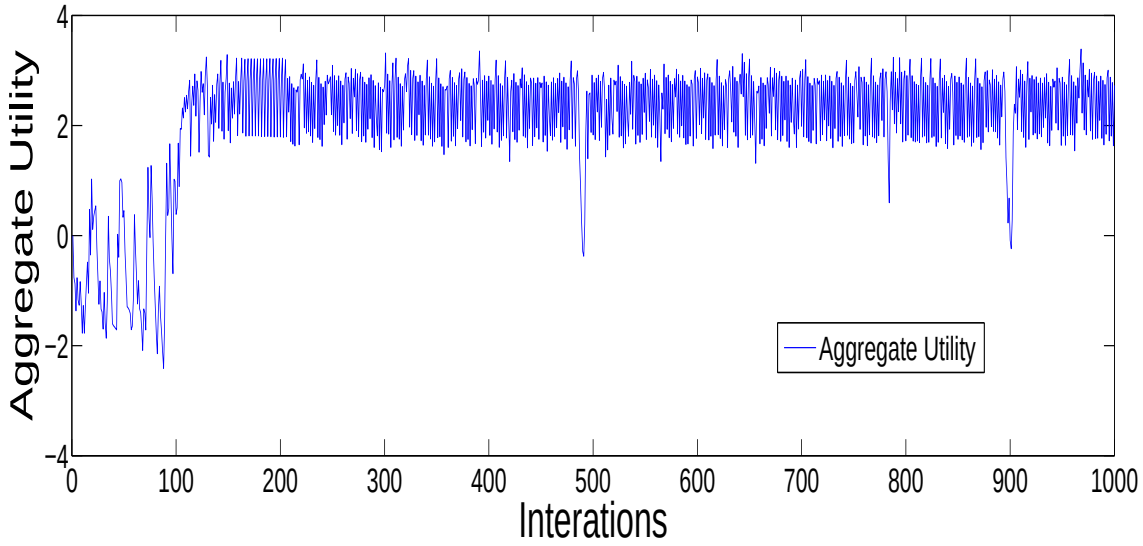


Figure 6.7: TRSS algorithm converges and push the aggregate utility of the network to optimal in Example 2

algorithm is at the center of this extension. Classic asynchronous implementations of parallel and distributed computing systems can be used for this extension.

Third, the complexity of TRSS is to be further reduced. This includes reducing both the computational complexity and communication overhead. A more efficient TRSS core using Newton-like algorithm may be designed. Efficient frequency scheduling algorithm is also critical. As mentioned in [46], a tradeoff analysis of optimality and performance can be beneficial for improving TRSS.

In the future, we are also planning to build up a DSA testbed that runs TRSS or its variants to investigate the practical aspect of TRSS.

6.7 Appendix

Proof of Proposition 2

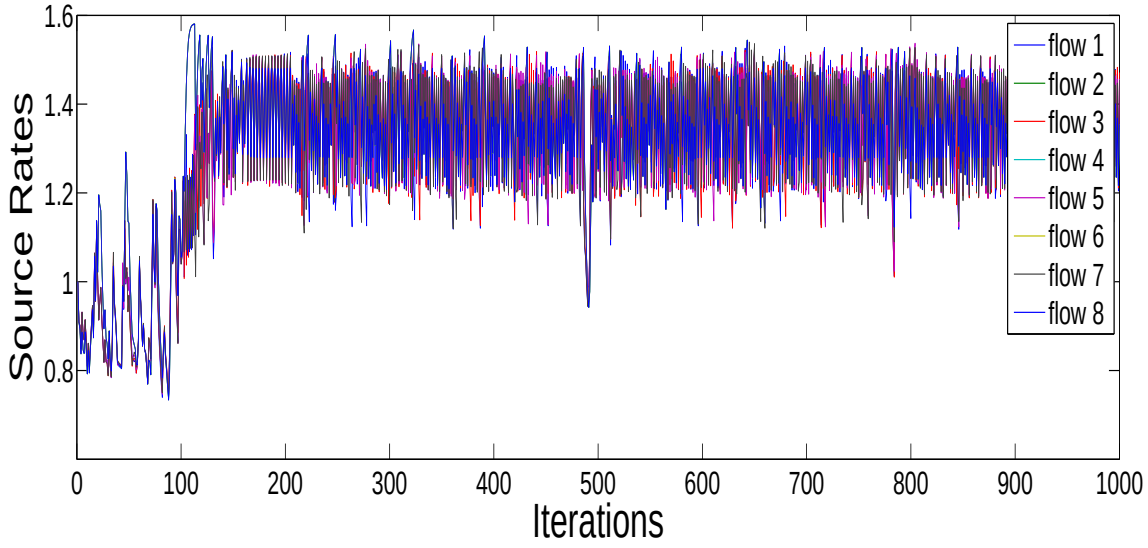


Figure 6.8: TRSS algorithm converges and push the aggregate utility of the network to optimal in Example 2

Proof: Substitute $c_l[t]$ using $b_l[t]\log(1+SNR_l)$ for each link $l \in L$ in $\tilde{D}_2$.

Note that

$$\begin{aligned} & ||\tilde{\mathbf{u}}(t+1)-\tilde{\mathbf{u}}^*||_2^2 \\ = \; & ||[\tilde{\mathbf{u}}(t)-\gamma\mathbf{g}(\tilde{\mathbf{u}}(t))]^+-\tilde{\mathbf{u}}^*||_2^2 \\ \leq \; & ||\tilde{\mathbf{u}}(t)-\gamma\mathbf{g}(\tilde{\mathbf{u}}(t))-\tilde{\mathbf{u}}^*||_2^2 \\ = \; & ||\tilde{\mathbf{u}}(t)-\tilde{\mathbf{u}}^*||_2^2-2\gamma\mathbf{g}(\tilde{\mathbf{u}}(t))^T(\tilde{\mathbf{u}}(t)-\tilde{\mathbf{u}}^*)+\gamma^2||\mathbf{g}(\tilde{\mathbf{u}}(t))||_2^2 \\ \leq \; & ||\tilde{\mathbf{u}}(t)-\tilde{\mathbf{u}}^*||_2^2-2\gamma(\mathbf{D}(\tilde{\mathbf{u}}(t))-\mathbf{D}(\tilde{\mathbf{u}}^*))+\gamma^2||\mathbf{g}(\tilde{\mathbf{u}}(t))||_2^2 \end{aligned}$$

where the last equation follows from the definition of subgradient. Apply the above inequality recursively, we have

$$\begin{aligned} & ||\tilde{\mathbf{u}}(t+1)-\tilde{\mathbf{u}}^*||_2^2 \\ \leq \; & ||\tilde{\mathbf{u}}(1)-\tilde{\mathbf{u}}^*||_2^2-2\gamma\textstyle\sum_{\tau=1}^{t}(\mathbf{D}(\tilde{\mathbf{u}}(\tau))-\mathbf{D}(\tilde{\mathbf{u}}^*)) \\ & +\gamma^2\textstyle\sum_{\tau=1}^{t}||\mathbf{g}(\tilde{\mathbf{u}}(\tau))||_2^2 \end{aligned}$$

Since

$$||\tilde{\mathbf{u}}(t+1)-\tilde{\mathbf{u}}^*||_2^2 \geq 0 \quad ,$$

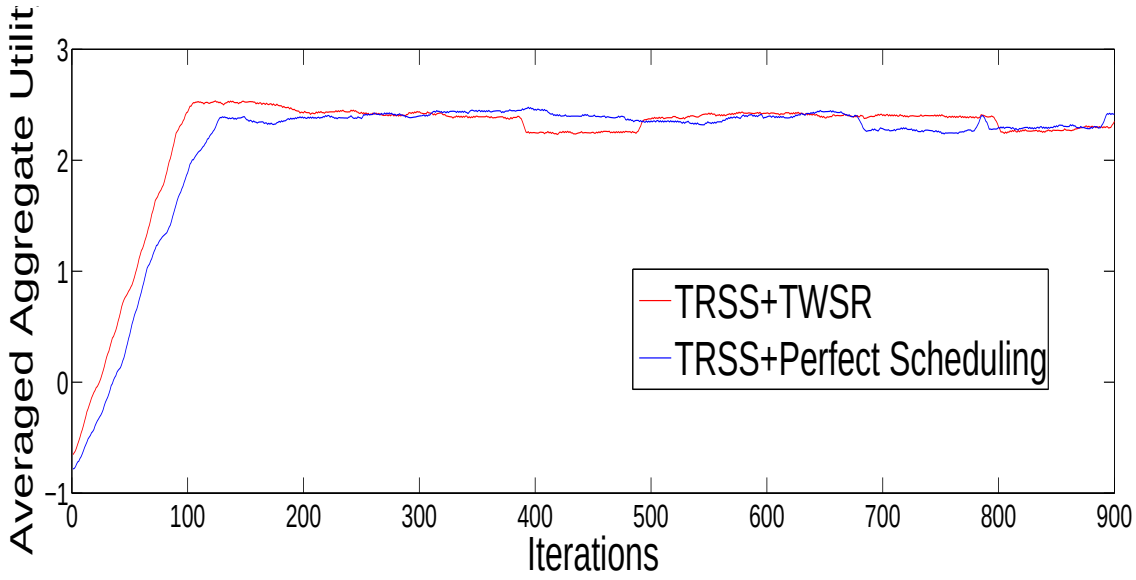


Figure 6.9: Comparing the average utility of TRSS with perfect scheduling and TRSS with TWSR in Example 2

we have

$$\begin{aligned} & 2\gamma \sum_{\tau=1}^{t} (\mathbf{D}(\tilde{\mathbf{u}}(\tau)) - \mathbf{D}(\tilde{\mathbf{u}}^*)) \\ \leq \; & ||\tilde{\mathbf{u}}(1) - \tilde{\mathbf{u}}^*||_2^2 + \gamma^2 \sum_{\tau=1}^{t} ||\mathbf{g}(\tilde{\mathbf{u}}(\tau))||_2^2 \\ \leq \; & ||\tilde{\mathbf{u}}(1) - \tilde{\mathbf{u}}^*||_2^2 + t\gamma^2 G^2 \end{aligned} \tag{6.33}$$

From Eq. (6.33), we have

$$\frac{1}{t} \sum_{\tau=1}^{t} (\mathbf{D}(\tilde{\mathbf{u}}(\tau)) - \mathbf{D}(\tilde{\mathbf{u}}^*)) \leq \frac{||\tilde{\mathbf{u}}(1) - \tilde{\mathbf{u}}^*||_2^2 + t\gamma^2 G^2}{2t\gamma}$$

and

$$\limsup_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t} (\mathbf{D}(\tilde{\mathbf{u}}(\tau)) - \mathbf{D}(\tilde{\mathbf{u}}^*)) \leq \frac{\gamma G^2}{2}$$

This shows that the adaptive algorithm converges statistically to within $\frac{\gamma G^2}{2}$ of optimal.

Proof of Lemma 1 Define $\mathcal{I}_{\text{in}}(i,k,l)$ and $\mathcal{I}_{\text{out}}(i,k,l)$ as binary indicators which are assigned value 1 if and only if there are traffic flows in or out of nodes i through link l and are destined for destination k.

Lemma 1: Suppose at each iteration t, for any link l, the error between suboptimal $\bar{b}_l(t)$ and $b_l(t)$ is bounded, i.e., $|\bar{b}_l(t) - b_l(t)| \leq \epsilon$, then the error between suboptimal subgradient $\bar{\mathbf{g}}(t)$

and $\mathbf{g}(t)$ for the dual problem is also bounded, i.e., $|\bar{g}_i^{(k)}(t) - g_i^{(k)}(t)| \leq K\epsilon log(1 + SNR_{max})$ where $SNR_{max} = \max_l\{SNR_l\}$, γ is the step size in master iteration and $\mathbf{p}$ is the power vector and $K = \sum_{l\in\mathcal{L}} |(\mathcal{I}_{\text{in}}(i,l) - \mathcal{I}_{\text{out}}(i,l))|$

Proof: For a link l, $b_l = 0$ means that link l is not used by any flow and thus can be removed from consideration. In the following we assume that $b_l > 0, \forall l \in \mathcal{L}$. We have

$$g_i^{(k)}(t+1) = y_i^{(k)} + x_{\text{into}(i)}^{(k)} - x_{\text{out}(i)}^{(k)}, \forall i,k \in \mathcal{N}$$

and

$$\bar{g}(t+1) = y_i^{(k)} + \bar{x}_{\text{into}(i)}^{(k)} - \bar{x}_{\text{out}(i)}^{(k)}, \forall i,k \in \mathcal{N}.$$

We have the approximation error in subgradient as

$$\begin{aligned}
& |\bar{g}_i^{(k)}(t+1) - g_i^{(k)}(t+1)| \\
= \; & |x_{\text{into}(i)}^{(k)} - x_{\text{out}(i)}^{(k)} - (\bar{x}_{\text{into}(i)}^{(k)} - \bar{x}_{\text{out}(i)}^{(k)})| \\
= \; & |\textstyle\sum_{(j,i)\in\mathcal{L}} x_{(j,i)}^k - \sum_{(i,j)\in\mathcal{L}} x_{(i,j)}^k \\
& -(\textstyle\sum_{(j,i)\in\mathcal{L}} \bar{x}_{(j,i)}^k - \sum_{(i,j)\in\mathcal{L}} \bar{x}_{(i,j)}^k)| \\
= \; & |\textstyle\sum_{l\in\mathcal{L}} x_l^k \mathcal{I}_{\text{in}}(i,k,l) - \sum_{l\in\mathcal{L}} x_l^k \mathcal{I}_{\text{out}}(i,k,l) \\
& -(\textstyle\sum_{l\in\mathcal{L}} \bar{x}_l^k \mathcal{I}_{\text{in}}(i,k,l) - \sum_{l\in\mathcal{L}} \bar{x}_l^k \mathcal{I}_{\text{out}}(i,k,l))| \\
= \; & |\textstyle\sum_{l\in\mathcal{L}} [x_l^k \mathcal{I}_{\text{in}}(i,k,l) - x_l^k \mathcal{I}_{\text{out}}(i,k,l) \\
& -(\bar{x}_l^k \mathcal{I}_{\text{in}}(i,k,l) - \bar{x}_l^k \mathcal{I}_{\text{out}}(i,k,l))]|
\end{aligned}$$

According to the back-pressure routing algorithm, if link l is assigned to traffic that is destined for node $\hat{k} \neq k$ at node i, then $x_l^{(k)} = x_l^{(k)} = 0$ and there is no approximation error. The above equation equals 0 and we are done.

Otherwise, if link l is assigned to traffic that is indeed destined for node k, we have the following

$$\begin{aligned}
& |\bar{g}_i^{(k)}(t+1) - g_i^{(k)}(t+1)| \\
= \; & |\textstyle\sum_{l\in\mathcal{L}} [x_l \mathcal{I}_{\text{in}}(i,l) - x_l \mathcal{I}_{\text{out}}(i,l) \\
& -(\bar{x}_l \mathcal{I}_{\text{in}}(i,l) - \bar{x}_l \mathcal{I}_{\text{out}}(i,l))]| \\
= \; & |\textstyle\sum_{l\in\mathcal{L}} [(x_l - \bar{x}_l)\mathcal{I}_{\text{in}}(i,l) - (x_l - \bar{x}_l)\mathcal{I}_{\text{out}}(i,l)]| \\
= \; & |\textstyle\sum_{l\in\mathcal{L}} [(x_l - \bar{x}_l)(\mathcal{I}_{\text{in}}(i,l) - \mathcal{I}_{\text{out}}(i,l))]|
\end{aligned}$$

With the scaling (6.32) on x_l, we have

$$x_l - \bar{x}_l \leq (b_l - \bar{b}_l)\log(1 + SNR_l) \tag{6.34}$$

and

$$\begin{aligned}
& |\textstyle\sum_{l\in\mathcal{L}}[(x_l - \bar{x}_l)(\mathcal{I}_{\text{in}}(i,l) - \mathcal{I}_{\text{out}}(i,l))]| \\
\leq\ & \textstyle\sum_{l\in\mathcal{L}}|(x_l - \bar{x}_l)(\mathcal{I}_{\text{in}}(i,l) - \mathcal{I}_{\text{out}}(i,l))| \\
\leq\ & \textstyle\sum_{l\in\mathcal{L}}|(b_l - \bar{b}_l)log(1 + SNR_l)(\mathcal{I}_{\text{in}}(i,l) - \mathcal{I}_{\text{out}}(i,l))| \\
\leq\ & \textstyle\sum_{l\in\mathcal{L}}|\epsilon log(1 + SNR_{max})(\mathcal{I}_{\text{in}}(i,l) - \mathcal{I}_{\text{out}}(i,l))| \\
\leq\ & \epsilon log(1 + SNR_{max})\textstyle\sum_{l\in\mathcal{L}}|(\mathcal{I}_{\text{in}}(i,l) - \mathcal{I}_{\text{out}}(i,l))| \\
\leq\ & \epsilon K log(1 + SNR_{max})
\end{aligned}$$

where $K = \sum_{l\in\mathcal{L}}|(\mathcal{I}_{\text{in}}(i,l) - \mathcal{I}_{\text{out}}(i,l))|$

Thus, we have

$$|\bar{g}_i^{(k)}(t) - g_i^{(k)}(t)| \leq K\epsilon log(1 + SNR_{max}) \tag{6.35}$$

This completes the proof.

Proof for Proposition 3

In *Lemma 1* we prove that when $||\bar{\mathbf{b}}(t) - \mathbf{b}(t)||_2 \leq \epsilon$, we have

$$|\bar{\mathbf{g}}(t) - \mathbf{g}(t)| \leq K\epsilon log(1 + SNR_{max})$$

Define ϵ_t-subgradient in time slot t as $\bar{\mathbf{g}}(t)$. It can be shown that $\epsilon_t \leq K\gamma\epsilon log(1 + SNR_{max})$, where γ is the step size. Suppose ϵ_t-subgradient is used at each iteration instead of the optimal subgradient, then the proof for the statistical convergence using ϵ_t-subgradient is basically the same as using normal subgradient. An in-depth treatment of ϵ_t-subgradient algorithm can be found in [88].

Chapter 7

Conclusions and Future Work

7.1 Conclusions

In this chapter, we summarize the work presented in this dissertation and discuss some directions for future study.

In Chapter 3, we present the first comprehensive model for multihop DSA networks based on two dimensional Poisson Point Process (PPP). This model can be used as an analytical tool to study the interaction between primary and secondary users in a DSA network. We incorporate essential DSA mechanisms such as spectrum sensing and primary interference avoidance in the model. The subsequent numerical study links the model to representative DSA network scenarios and the results agree well with intuition and empirical studies. Moreover, the numerical results provide fundamental insights on the design of future DSA networks and show potential as an evaluation tool for practical implementations. Theoretically, the model generalizes Kleinrock and Takagi's classical work [78] on the one-hop throughput of slotted aloha network and converges to classical models in the absence of primary network which lends credibility to our own theoretical model.

In Chapter 4, we introduce a simple and robust model to compute aggregate network ca-

pacity for wireless networks using partially overlapping channels. We present a quantitative study on the relationship between network performance and critical network settings such as node density, network topology and traffic load. Both our mathematical induction and numerical results show that the use of partially overlapped channels significantly improves network performance. Additionally, our analytical results show that denser networks with higher traffic load favor POCs based designs. Our research also reveals that the capacity improvement ratio is not a monotonic function of node density. In extreme cases, such as when distribution area is at the smallest interference range level, the marginal gain of capacity improvement by using POCs can be diminished or even reversed. We also propose a joint channel scheduling and flow routing optimization model that achieves maximum network throughput in POC based WiFi and ad-hoc networks. Numerical examples shows the effectiveness of our design.

In Chapter 5, we present a distributed algorithm JSSRC for joint spectrum sharing and end-to-end rate control for multihop cognitive radio networks. By taking advantage of frequency-agile radio's capability to dynamically reconfigure central frequency and channel width, JSSRC with appropriate link layer scheduling is able to achieve the optimal spectrum sharing policy that maximizes transport layer utility. For practical implementation, we present a novel timing window based spectrum reservation scheme (TWSR) for link layer scheduling. Simulation results show that JSSRC algorithm with TWSR achieves good performance despite TWSR's heuristic nature.

In Chapter 6, we present TRSS, a joint optimization algorithm for cross-layer resource management in frequency-agile radio based networks. TRSS is proposed to include routing considerations in the adaptive design. TRSS is different from JSSRC in both transport layer design and network layer design. For the transport layer, TRSS requires only queuing information at the local nodes instead of queuing information along the end-to-end path as is the case for JSSRC. This reduces control overhead. For the routing layer, TRSS uses state of the art back-pressure routing to move a packet along the most efficient path in the network. This addition of routing design also makes TRSS a complete cross-layer solution

for optimal spectrum sharing in wireless networks that can be plugged on top of any physical layer design.

In both JSSRC and TRSS, we exploit the frequency agility of modern radios and provide a systematic treatment for the inherently cross-layer problem of spectrum sharing in wireless networks. Moreover, the derivation of the algorithm reveals fundamental mechanisms that are critical to performance and spectrum efficiency in wireless networks, e.g., local buffer saturation level, wireless link capacity saturation level, number of contenders for the same spectrum band and number of links a specific link can interfere. The follow-up discussion provides insights on the inner mechanism that are unique to spectrum sharing in dynamic spectrum access networks.

The work presented in this dissertation effectively exploit frequency-agile radio and DSA network technology to push efficient spectrum utilization in wireless networks. Both the mathematical models and algorithms designed represent state of the art results. The numerical and simulation results obtained cross-validate the soundness of our approach to address the spectrum shortage problem, an increasingly prominent and urgent problem.

7.2 Future Work

Based on the research results obtained from this dissertation, we propose some future directions for exploration.

Firstly, the basic frameworks of JSSRC and TRSS can be easily extended. The ideas include:

· In the framework, JSSRC and TRSS assumes that the radio environment of the network changes at a slower pace than the convergence speed of the algorithms. In many practical systems especially in fading environment, wireless channels can vary relatively fast. To accommodate time-varying channels, a stochastic channel model can be used

to describe the network. For example, if we assume the channel state is described by a finite state Markov Chain, then we can define the average capacity region $\bar{\Lambda}$ as a function of these states and JSSRC(TRSS) can be extended to maximize network utility over these average capacity region.

· JSSRC and TRSS can be extended to work asynchronously. Fortunately, the dual decomposition method have many good properties for asynchronous design. Convergence and optimality analysis for asynchronous algorithm is at the center of this extension. Classic asynchronous implementations of parallel and distributed computing systems can be used for this extension.

· The complexity of JSSRC(TRSS) can be further reduced. This is crucial for practical adoption of the proposed algorithms. Complexity reduction includes reducing both the computational complexity and communication overhead. For example, a more efficient JSSRC(TRSS) core using Newton-like algorithm may be designed and is expected to significantly reduce the computational complexity and speed the convergence of the distributed algorithm. To reduce communication overhead, JSSRC and TRSS can take advantages of cached channel information if no significant changes has occurred since last adaptation. The tradeoff between optimality, convergence and computational overhead is the most challenging and critical issue in this extension. As mentioned in [46], a tradeoff analysis of optimality and performance can be beneficial for improving JSSRC(TRSS).

· Efficient spectrum scheduling algorithm is also critical. Discussion and proposals on bin-packing, realtime scheduling will be helpful in both design and theoretical analysis of such algorithms.

Secondly, the model presented in Chapter 3 can be extended to investigate CSMA based DSA network. CSMA is a precursor of many modern random access based protocol/standards, e.g., the IEEE 802.11 family of protocol/standard. The study of CSMA behavior in DSA

networks will provide important insight for design and deployment of important network technologies such WLAN and WiMax. Unlike in the slotted aloha network discussed in the preliminary work which has many interesting properties that are good for theoretical analysis, in CSMA based networks, the interaction between users are much more complex. The interaction between primary users and secondary users in a DSA network will also add to the complexity. Another challenging issue is the well know hidden node problem. Due to the asymmetry of radio environment between the transmitter and receiver as well as their corresponding neighbors, the modeling of spectrum sensing and successful packet reception is much more difficult. Some approximations has to be made to make the theoretical model remains tractable as in many classical works [60] [79]. The tradeoff between modeling accuracy and tractability will be an prominent issue throughout the analysis. A finite state Markov Chain model can also be used to capture the time-domain behavior of the network and extend the analysis from the instant behavior of the the network in a single time slot to the long-term evolution of the network in the time domain.

Bibliography

[1] CNN iReporter, ireport.cnn.com/

[2] Citi Bank mobile banking, www.citibank.com/citimobile

[3] Wells Fargo mobile banking, https://www.wellsfargo.com/mobile/

[4] American Telemedicine Association, http://www.americantelemed.org/i4a/pages /index.cfm?pageid=1

[5] Cisco predicts mobile data traffic to grow, http://www.cisco.com/en/US/solutions /collateral/ns341/ns525/ns537/ns705/ns827/white-paper-c11-520862.html

[6] Erricsson predicts mobile data traffic to grow, http://www.ericsson.com/news/1437680

[7] FCC warns of spectrum shortage, http://techdailydose.nationaljournal.com/2010/10/fcc-forecasts-major-spectrum-s.php

[8] U.S. Radio Frequency Allocation Chart, http://www.ntia.doc.gov/osmhome/allochrt.pdf

[9] AT&T data plan, http://www.att.com/shop/wireless/plans/data-plans.jsp

[10] Att phases out unlimited data plan, http://www.computerworld.com/s/article/9177598/AT-T-dumps-iPhone-unlimited-data-plan-adds-tethering

[11] I.F. Akyildiz, W.Y. Lee, M.C. Vuran, S. Mohanty, "NeXt Generation / Dynamic Spectrum Access / Cognitive Radio Wireless Networks: A Survey," Computer Networks Journal (Elsevier), Vol. 50, No. 13, pp. 2127-2159, September 2006.

[12] Q. Zhao, B. Sadler, " Survey of Dynamic Spectrum Access," IEEE Signal Processing Magazine, pp. 79-89, May 2007.

[13] S. Gandhi, C. Buragohain, L. Cao, H. Zheng, S. Suri, "A General Framework for Wireless Spectrum Auctions," in Proceedings of IEEE DySpan 2007.

[14] T. Yucek, H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," IEEE Communications Surveys and Tutorials, Vol. 11, pp. 116-130, First Quarter 2009.

[15] Y. Thomas Hou, Y. Shi, and H. D. Sherali, "Spectrum sharing for multi-hop networking with cognitive radios," IEEE Journal on Selected Areas of Communications, Vol. 26, No. 1, pp. 146-155, January 2008.

[16] D. Raychaudhuri, N. B. Mandayam, etc., "CogNet - An Architecture for Experimental Cognitive Radio Networks within the Future Internet," in Proceedings of ACM MobiArch 2006.

[17] P. Kyasanur, N. H. Vaidya, "Protocol Design Challenges for Multi-hop Dynamic Spectrum Access Networks," in Proceedings of DySpan 2005.

[18] Daniel Ugarte, A. Bruce McDonald, "On the capacity of Dynamic Spectrum Access Enabled Networks," in Proceedings of Dyspan 2005.

[19] Syed Ali Jafar, Sudhir Srinivasa, "Capacity Limits of Cognitive radio with distributed and dynamic spectral activity," IEEE Journal on Selected Areas in Communications, Vol. 25, No. 3, April 2007.

[20] A. Goldsmith, S. A. Jafar, I. Maric and S. Srinivasa, "Breaking Spectrum Gridlock with Cognitive Radios: An Information Theoretic Perspective," Proceedings of the IEEE, Vol. 97, No. 5, May 2009, pp. 894-914

[21] Jeffrey H. Reed, "Software Radio: A Modern Approach to Radio Engineering," Prentice Hall Communications Engineering and Emerging Technologies Series.

[22] J. Mitola, III and G.Q. Maguire, Jr., "Cognitive Radio: Making Software Radios More Personal," IEEE Personal Communications, Vol. 6, No. 4, 1999.

[23] S. Haykin, "Cognitive Radio: Brain-empowered WIreless Communications," IEEE Journal on Selected Areas in Communications 23 (2005) (2), pp. 201220.

[24] M. Kodialam, T. Nandagopal,"Characterizing the capacity region in multi-radio multi-channel wireless mesh networks," in Proceedings of ACM MobiCom 2005.

[25] M. Alicherry, R. Bhatia, and L.E. Li,"Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks," in Proceedings of ACM MobiCom 2005.

[26] Z. Feng and Y. Yang,"Throughput Analysis of Secondary Networks in Dynamic Spectrum Access Networks," in Proceedings of IEEE INFOCOM 2010 Workshop on Cognitive Wireless Communications and Networking.

[27] Z. and Y. Yang, "Characterizing the Impact of Partially Overlapped Channel on the Performance of Wireless Networks," in Proceedings of IEEE GLOBECOM 2008.

[28] Z. Feng, Y. Yang, " How Much Improvement Can We Get From Partially Overlapped Channels?" in Proceeding of IEEE WCNC 2008.

[29] Z. Feng, Y. Yang, "Two Phase Spectrum Sharing for Frequency-Agile Radio Networks ," in Proceeding of IEEE ICC 2009.

[30] Z. Feng, Y. Yang, "Joint Transport, Routing and Spectrum Sharing Optimization for Wireless Networks with Frequency-Agile Radios," in Proceedings of IEEE INFOCOM 2009.

[31] Z. Feng, Y. Yang, "Distributed spectrum sharing optimization for frequency-agility enabled wireless networks," submitted to IEEE Transaction on Mobile Computing.

[32] Radar Frequency Agility, http://en.wikipedia.org/wiki/Frequency_agility

[33] F.P.Kelly, A. Maulloo, and D. Tan, "rate control in communication networks: Shadow prices, proportional fairness and stability," Journal of the Operation Research Society, Vol. 49, pp.237-252, 1998.

[34] L. Tassiulas and A. Ephremides, "Stability Properties of Constrained Queueing Systems and Scheduling Policies for Maximum Throughput in Multihop Radio Networks," IEEE Transaction on Automatic Control, Vol. 37, No. 12, December 1992.

[35] S. Kunniyur and R. Srikant, "End-to-end congestion control: Utility functions, random losses and ECN marks," in Porceedings of IEEE INFOCOM, 2000.

[36] J. Mo, J. Walrand, "Fair end-to-end window-based congestion control," IEEE/ACM Transaction on Networking, Vol. 8, Issue 5, pp. 225-267, October 2000.

[37] F. Paganini, J. Doyle, and S. Low, "Scalable laws for stable network congestion control," in Proceedings of IEEE CDC, 2001.

[38] L. S. Brakmo and L. L. Peterson,"TCP Vegas: end to end congestion avoidance on a global internet," IEEE Journal on Selected Areas in Communications, 13(8), October 1995.

[39] S. H. Low and David E. Lapsley,"Optimization flow control, I: basic algorithm and convergence ," IEEE/ACM Transaction on Networking, Vol. 7, No. 6,pp.861-874, December 1999.

[40] D. X. Wei, C. Jin, S. H. Low and S. Hegde, "FAST TCP: motivation, architecture, algorithms, performance," in Proceedings of IEEE INFOCOM 2004

[41] Fastsoft, http://www.fastsoft.com/home/

[42] S. H. Low, L. L. Peterson, L. Wang, "Understanding TCP Vegas: a duality model," Jounal of the ACM , Vol.49, No.2, pp.207-235, March 2002.

[43] L. Chen, S. H. Slow, J.C. Doyle, "Joint congestion control and media access control design for ad hoc wireless networks," in Proceedings of IEEE INFOCOM 2005.

[44] Mung Chiang, S. Zhang, P. Hande, "Distributed rate allocation for inelastic flows: optimization frameworks, optimality conditions, and optimal algorithms," in Proceedings of IEEE INFOCOM 2005.

[45] Mung Chiang, "Balancing transport and physicallayers in wireless multihop networks: jointly optimal congestion control and power control," IEEE Journal on Selected Areas in Communications, Vol. 23, No. 1, January 2005.

[46] X. Lin, N. Shroff, "The Impact of Imperfect Scheduling on Cross-Layer Congestion Control in Wireless Networks," IEEE/ACM Transaction on Networking, Vol. 14, No. 2, April 2006

[47] M.J. Neely, E. Modiano, C.R. Rohrs, "Dynamic Power Allocation and Routing for Time-Varying Wireless Networks," IEEE Journal on Selected Areas of Communications, Vol. 23, No. 1, January 2005.

[48] A. Eryilmaz, R. Srikant, "Joint Congestion Control, Routing, and MAC for Stability and Fairness in Wireless Networks," IEEE Journal on Selected Areas of Communications, Vol. 24, No. 8, August 2006.

[49] Xin Wang, Koushik Kar, "Cross-layer rate control for end-to-end proportional fairness in wireless networks with random access," in Proceedings of ACM MobiHoc 2005.

[50] Loc Bui, A. Eryilmaz, R. Srikant, and Xinzhou Wu, "Joint asynchronous congestion control and distributed scheduling for multi-hop wireless networks," in Proceedings of IEEE INFOCOM 2006.

[51] A. Mishra, E. Rozner, S. Banerjee, and W. Arbaugh, "Using partially overlapped channels in wireless meshes," in Proceedings of IEEE WiMesh 2005.

[52] A. Mishra, V. Shrivastava, S. Banerjee, and W. Arbaugh, "Partially overlapped channels not considered harmful," SIGMETRICS Performance Evaluation Review, 34(1), 2006.

[53] A. Mishra, E. Rozner, S. Banerjee, and W. Arbaugh, "Exploiting partially overlapping channels in wireless networks: Turning a peril into an advantage," in Proceedings of ACM/USENIX Internet Measurement Conference 2005.

[54] H. Liu, H. Yu, X. Liu, C. Chuah, P. Mohapatra, "Scheduling multiple partially overlapped channels in wireless mesh networks," in Proceedings of IEEE ICC 2007.

[55] A. Rad, Vincent W.S. Wong, "Partially overlapped channel assignment for Multi-Channel wireless mesh network," in Proceedings of IEEE ICC 2007.

[56] Y. Shi, Y.T. Hou, J. Liu, and S. Kompella,"How to correctly use the protocol interference model for multi-hop wireless networks," in Proceedings of ACM MobiHoc, 2009.

[57] R. Jain, D. Chiu, and W. Hawe,"A Quantitative Measure Of Fairness And Discrimination For Resource Allocation In Shared Computer Systems," DEC Research Report TR-301, September 1984.

[58] S.M.S. Hasan, S.W. Ellingson,"Multiband Antenna-Receiver Integration using an RF multiplexer with sensitivity constrained design," in Proceedings of IEEE International Symposium on Antenna Propogation 2008.

[59] A. Raniwala and T.-C. Chiueh, "Architecture and algorithms for an IEEE 802.11-based multi-channel wireless mesh network," in Proceedings of IEEE INFOCOM 2005.

[60] Giuseppe Bianchi, "Performance Analysis of the IEEE 802.22 Distributed Coordination Function," IEEE Journal on Selected Areas in Communications, Vol. 18, No. 3, March 2000.

[61] Y. Thomas Hou, Yi Shi, and Hanif D. Sherali, "Optimal spectrum sharing for multi-hop software defined radio Networks," in Proceedings of IEEE INFOCOM 2007.

[62] Yi Shi and Y. Thomas Hou, "A distributed optimization algorithm for multi-hop cognitive radio networks," in Proceedings of IEEE INFOCOM 2008.

[63] P. Kyasanur and N. H. Vaidya, "Capacity of multi-channel wireless networks: impact of number of channels and interfaces," in Proceedings of ACM MobiCom 2005.

[64] P. Kyasanur and N.H. Vaidya, "Capacity of multi-channel wireless networks: impact of number of channels and interfaces," in Proceedings of IEEE MobiCom 2005.

[65] J. So and N. Vaidya, "Multi-Channel MAC for Ad Hoc Networks: Handling Multi-Channel Hidden Terminals Using A Single Transceiver," in Proceedings of MobiHoc 2004.

[66] R. Chandra, R. Mahajan, T. Moscibroda, R. Raghavendra, P. Bahl,"A case for adapting channel width in wireless networks," in Proceedings of ACM SIGCOMM 2008.

[67] P. Bahl, R. Chandra, and J. Dungan, "SSCH: slotted seeded channel hopping for capacity improvement in IEEE 802.11 ad-hoc wireless networks," in Proceedings of MoboHoc 2004.

[68] W. Ren, Q. Zhao, and A. Swami, "Power Control in Cognitive Radio Networks: How to Cross a Multi-Lane Highway," IEEE Journal on Selected Areas of Communications, Vol. 27, No. 7, pp. 12831296, September 2009.

[69] Lili Cao and Haitao Zheng, "Distributed spectrum allocation via local bargaining," in Proceedings of SECON 2005.

[70] X. Liu and W. Wang, "On the characteristics of spectrum-agile communication networks," in Proceedings of DySpan 2005.

[71] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," in Proceedings of ACM MOBICOM 2003.

[72] B. Ko, V. Misra, J. Padhye, D. Rubenstein, "Distributed channel assignment in multi-radio 802.11 mesh networks," in Proceedings of IEEE WCNC 2007.

[73] A. Raniwala, K. Gopalan, and T. cker Chiueh, "Centralized channel assignment and routing algorithms for multi-channel wireless mesh networks," ACM SIGMOBILE Mobile Computing and Communication Review, 8(2), 2004.

[74] Nie Nie, Cristina Comaniciu, "Adaptive channel allocation spectrum etiquette for cognitive radio networks," Mobile Networks and Applications, Vol. 11 pp. 779–797, December 2006.

[75] A. Subramanian, H. Gupta, S. R. Das, "Mimimum-Interference channel assignment in multi-radio wireless mesh networks," in Proceedings of SECON 2007.

[76] E. Ziouva, T. Antonakopoulos, "CSMA/CA performance under high traffic conditions: throughput and delay analysis," Elsevier Computer Communications, Vol. 25, 3, Febuary 2002, 313-321.

[77] A. E. Leu, M. McHenry, B. L. Mark, "Modeling and analysis of interference in Listen-Before-Talk spectrum access schemes," International Journal of Network Management 2006; 16:131-147.

[78] Hideaki Takagi, Leonard Kleinrock, "Optimal Tranmission Ranges for Ransomly Distributed Packet Radio Terminals," IEEE Transaction on Communications, Vol. Com-32, No. 3, March 1984.

[79] F. A. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part I-Carrier Sense Multiple-Access Modes and Their Throughput-Delay Characteristics," IEEE Transaction on Communications, Vol. Com-23, No. 12, December 1975.

[80] Wimax forum whitepapers: http://www.wimaxforum.org/.

[81] Cisco whitepaper, "802.11n: The Next Generation of Wireless Performanc," 2007.

[82] FCC, ET Docket No 03-222 Notice of proposed rule making and order, December 2003.

[83] "ILOG CPLEX", available at http://www.ilog.com/products/cplex/.

[84] Gauss's Circle, http://mathworld.wolfram.com/GausssCircleProblem.html

[85] MATLAB, www.mathworks.com

[86] Poisson point process, ww.stat.berkeley.edu/users/pitman/s205f02/lecture27.ps.

[87] Thomas M. Cover, Joy A. Thomas, "Elements of Information Theory 2nd Edition," Wiley Series in Telecommunications and Signal Processing

[88] D. Bertsekas, Nonlinear Programming, 2nd edtion, Athena Scientific, 1999.

[89] Nemhauser, G.L. and L.A. Wolsey, "Integer and Combinatorial Optimization," John Wiley & Sons, 1988.

[90] A. Goldsmith, "Wireless Communications", Cambridge University Press, 2005.