THROUGHPUT AND DELAY ANALYSIS IN COGNITIVE OVERLAID NETWORKS

A Dissertation

by

LONG GAO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2009

Major Subject: Electrical Engineering

THROUGHPUT AND DELAY ANALYSIS IN COGNITIVE OVERLAID NETWORKS

A Dissertation

by

LONG GAO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Shuguang Cui
Committee Members,	Zixiang Xiong
	Srinivas Shakkottai
	Lewis Ntaimo
Head of Department,	Costas Georghiades

December 2009

Major Subject: Electrical Engineering

ABSTRACT

Throughput and Delay Analysis in Cognitive Overlaid Networks. (December 2009)

Long Gao

B.S., Beijing Jiaotong University;

M.S., Beijing University of Posts and Telecommunications

Chair of Advisory Committee: Dr. Shuguang Cui

Consider a cognitive overlaid network (CON) that has two tiers with different priorities: a primary tier vs. a secondary tier, which is an emerging network scenario with the advancement of cognitive radio (CR) technologies. The primary tier consists of randomly distributed primary radios (PRs) of density n, which have an absolute priority to access the spectrum. The secondary tier consists of randomly distributed CRs of density $m = n^\gamma$ with $\gamma \geq 1$, which can only access the spectrum opportunistically to limit the interference to PRs. In this dissertation, the fundamental limits of such a network are investigated in terms of the asymptotic throughput and packet delay performance when m and n approaches infinity. The following two types of CONs are considered: 1) selfish CONs, in which neither the primary tier nor the secondary tier is willing to route the packets for the other, and 2) supportive CONs, in which the secondary tier is willing to route the packets for the primary tier while the primary tier does not. It is shown that in selfish CONs, both tiers can achieve the same throughput and delay scaling laws as a stand-alone network. In supportive CONs, the throughput and delay scaling laws of the primary tier could be significantly improved with the aid of the secondary tier, while the secondary tier can still achieve the same throughput and delay scaling laws as a stand-alone network. Finally, the throughput and packet delay of a CON with a small number of nodes are investigated. Specifically, we investigate the power and rate control schemes for multiple CR links

in the same neighborhood, which operate over multiple channels (frequency bands) in the presence of PRs with a delay constraint imposed on data transmission. By further considering practical limitations in spectrum sensing, an efficient algorithm is proposed to maximize the average sum-rate of the CR links over a finite time horizon under the constraints on the CR-to-PR interference and the average transmit power for each CR link. In the proposed algorithm, the PR occupancy of each channel is modeled as a discrete-time Markov chain (DTMC). Based on such a model, a novel power and rate control strategy based on dynamic programming (DP) is derived, which is a function of the spectrum sensing output, the instantaneous channel gains for the CR links, and the remaining power budget for the CR transmitter. Simulation results show that the proposed algorithm leads to a significant performance improvement over heuristic algorithms.

To My Parents and Xuan

ACKNOWLEDGMENTS

The following people have helped me with this dissertation and throughout my time at Texas A&M University. First, I owe my deepest gratitude to my advisor, Dr. Shuguang Cui. I have the good fortune to be advised by him who truly cares about my professional and personal development. His dedication to his students cannot be appreciated more. Besides my advisor, I would like to thank Dr. Zixiang Xiong, Dr. Srinivas Shakkottai, and Dr. Lewis Ntaimo for agreeing to serve on my doctoral committee. Further thanks go to the administrative staff in the Department of Electrical and Computer Engineering, and in particular to Paula Evans. Their help throughout these years is much appreciated. I also want to thank Dr. Rui Zhang at A-Star, Singapore, for his advice on my research.

My time in College Station would not have been the same without my friends and colleagues at Texas A&M University. In particular, I would like to thank Dr. Changchuan Yin, Fan Zhang, Charalambos Charalambous, Liang Liu, Zhen Huang, Jing Wang, Meng Zeng, Lily Zhang, Qing Zhou, Chuan Huang, Kyle Cai, Armin Banaei, and Tarun Agarwal. Finally, I would like to thank my parents and Xuan Zhao for their love and support.

TABLE OF CONTENTS

CHAPTER Page

Page

LIST OF FIGURES

FIGURE Page

CHAPTER I

INTRODUCTION

Recently, the emergence of cognitive radio techniques results in a new type of two-tier overlaid networks, i.e., cognitive overlaid networks (CONs). The design, analysis, and deployment issues of CONs are interesting and challenging. In this chapter, we first introduce the definition of CONs. The challenges in analyzing the fundamental limits of CONs are then described with an emphasis on the throughput and delay analysis. Afterwards, we briefly describe the prior work on this topic and summarize our contributions. The overall dissertation organization is given at the end.

A. Definition of Cognitive Overlaid Networks

The fast growth in wireless services results in an over-crowded spectrum due to the current static bandwidth assignment strategy adopted by the government. In year 2002, the Federal Communications Commission (FCC) Spectrum Policy Task Force published a report [1], indicating that there is a spectrum shortage for further licensing, while more than 90 percent of the already-licensed spectrum remains idle at a given time and location. To explore the under-utilized spectrum resources, cognitive radio (CR) techniques have been proposed to implement opportunistic spectrum access over the licensed legacy bands [2-11].

The emergence of CRs introduces a new type of network, i.e., the CON as shown in Fig. 1, which has two tiers sharing the same spectrum with different priorities: a primary tier vs. a secondary tier. The primary tier consists of legacy primary radios (PRs), which have an absolute priority to access the spectrum. The secondary tier

---

The journal model is *IEEE Transactions on Automatic Control.*

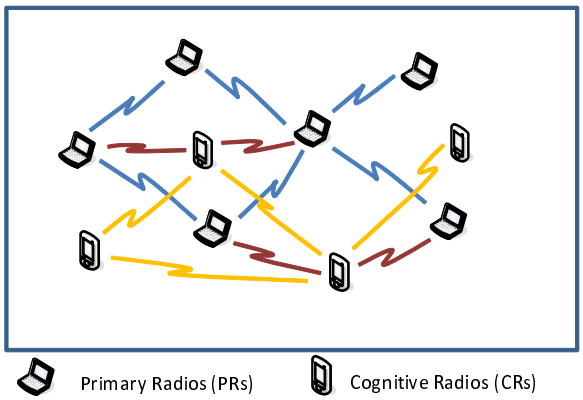

Fig. 1. Illustration of a cognitive overlaid network.

consists of CRs, which can only access the spectrum opportunistically to limit the interference to PRs. These two tiers both have their own data packets to transport and may interfere with each other during the packet transmission if not carefully coordinated. In recent years, there have been significant and increasing interests in CONs due to their promising future applications.

B. Challenges and Motivations

The design and deployment of CONs necessitate an understanding of the fundamental limits in such networks, e.g., the throughput and delay that each tier can support[1]. One possible way to address this issue is to start with small CONs, e.g., a four-node CON with one primary source-destination (S-D) pair and one secondary (cognitive) S-D pair [12], or even a smaller three-node network with one primary S-D pair and one CR as a relay [13]. Based on the results for such building blocks, one may gain some insight for larger CONs. However, this has been proved challenging due to the lack of information-theoretical understanding of the aforementioned four-node and

[1]We investigate the throughput and delay performance of the two tiers respectively, since each tier usually has its own packets to transport.

three-node CONs.

Another approach is to consider large CONs and focus on the asymptotic performance by taking the numbers of the nodes (for both PRs and CRs) to infinity. Such results could provide us high-level guidance on how to design appropriate network architectures and protocols for large CONs. To analyze the asymptotic performance of such large CONs, the following issues have to be addressed:

- **Network Model**: A model of how the PR and CR nodes are generated has to be chosen first. The channel model between an arbitrary pari of transmitter and receiver also has to be specified. Furthermore, the network performance measures, e.g., the throughput and packet delay, has to be clearly defined.

- **Interaction Model**: Given the two-tier structure of CONs, the interaction model between the two tiers has to be clarified. In particular, the questions such as whether the packet exchanging between the two tiers is allowed or not, and what kind of packet exchanging procedures that the two tiers could use, have to be answered.

- **Interference Management**: One of the most important technical requirements for the CRs is that their data transmission should not result in harmful interference to PRs. To satisfy such a requirement, interference management or avoidance has to be applied, e.g., the CRs could perform individual or cooperative spectrum sensing to detect the idle frequency bands, so-called spectrum holes, and adjust their carrier frequency, transmit power, data rate, and other transmission parameters in a timely manner to minimize the interference to the PRs.

In this dissertation, with the above issues under consideration, we mainly focus on the asymptotic analysis over throughput and packet delay in large CONs. Towards

the end, we devote one chapter to a small-network case to show how to design CR transmission strategies to maximize the CR achievable rate, while providing certain protection over PR transmissions. Some existing work related to our work will be briefly reviewed in the next section.

C. Prior Work

The explosive growth of large-scale wireless applications motivates people to study the fundamental limits over wireless networks. Initiated by the seminal work of Gupta and Kumar [14], the throughput scaling law for large-scale wireless networks has become an active research topic [15-39]. Scaling laws provide a fundamental way to measure the achievable throughput of a wireless network. Considering n nodes that are randomly distributed in a unit area and grouped independently into one-to-one S-D pairs, it was shown [14] that the typical time-slotted multi-hop architecture with a common transmission range and adjacent-neighbor communication can achieve a sum throughput that scales as $\Theta\left(\sqrt{n/\log n}\right)$[2]. Besides, it was shown that an alternative arbitrary network structure with optimally chosen traffic patterns, node locations, and transmission ranges can achieve a sum throughput of order $\Theta\left(\sqrt{n}\right)$. In [17], with percolation theory, Franceschetti *et al.* showed that the $\Theta\left(\sqrt{n}\right)$ sum throughput scaling is achievable even for randomly deployed networks under certain special conditions. In [18] [25], it was shown that by allowing the nodes to move independently and uniformly, a constant throughput scaling $\Theta(1)$ per S-D pair can be achieved. Later, Diggavi *et al.* showed that a constant throughput per S-D pair

[2]We use the following notations throughout this dissertation: i) $f(n) = O(g(n))$ means that there exists a constant c and integer N such that $f(n) < cg(n)$ for $n > N$; ii) $f(n) = \Omega(g(n))$ means that $g(n) = O(f(n))$; iii) $f(n) = \Theta(g(n))$ means that $f(n) = O(g(n))$ and $g(n) = O(f(n))$; iv) $f(n) = o(g(n))$ means that $f(n)/g(n) \to 0$ as $n \to \infty$.

is achievable even with a one-dimensional mobility model [16]. In these approaches, the network area is fixed and the throughput scales with the node density n. We call this kind of network as *dense network*. On the other hand, based on the *extended network* model where the density of nodes is fixed and the network area increases with n, the information-theoretic scaling laws of transport capacity were studied for different values of the pathloss exponent α in [19] [29-32] [38]. In particular, Ozgur *et al.* [30] proposed a hierarchical cooperation scheme to achieve a sum throughput that scales as $n^{2-\kappa/2}$ for $2 \leq \kappa < 3$, i.e., asymptotically linear for $\kappa = 2$.

In wireless networks, another key performance metric is delay, which incurs the interesting problems regarding the interactions between throughput and delay. The issues of delay-throughput tradeoff for static and mobile wireless networks were addressed in [15] [21-28]. In [21], El Gamal *et al.* established the optimal delay-throughput tradeoff for static and mobile wireless networks. For static networks, they showed that the optimal delay-throughput tradeoff is given by $D(n) = \Theta\left(n\lambda(n)\right)$, where $\lambda(n)$ and $D(n)$ are the throughput and delay per S-D pair, respectively. Using a random-walk mobility model, they showed that a much higher delay of $\Theta\left(n \log n\right)$ is associated with the higher throughput of $\Theta(1)$ for mobile networks. The delay-throughput tradeoffs in mobile wireless networks have been investigated under many other mobility models, which include the i.i.d. model [23] [25] [28], the hybrid random walk model [27], and the Brownian motion model [24]. For the hierarchical cooperation scheme in a static wireless network, Ozgur and Lévêque [26] showed that a significantly larger delay was introduced compared with the traditional multi-hop scheme, and the delay-throughput tradeoff is $D(n) = \Theta\left(n\left(\log n\right)^2 \lambda(n)\right)$ for $\lambda(n)$ between $\Theta\left(1/(\sqrt{n}\log n)\right)$ and $\Theta\left(1/\log n\right)$.

All the aforementioned results focus on the throughput scaling laws or the delay-throughput tradeoffs for a single wireless network. Consider a licensed primary net-

work and a cognitive secondary network coexisting in a unit area. The primary network has the absolute priority to use the spectrum, while the secondary network can only access the spectrum opportunistically to limit the interference to the primary network. In this overlaid regime, the throughput scaling law and the delay-throughput tradeoff for both the primary and secondary networks are interesting and challenging problems. Some preliminary work along this line appeared recently. In [34] [35], Vu *et al.* considered the throughput scaling law for a single-hop cognitive radio network, where a linear scaling law is obtained for the secondary network with an outage constraint for the primary network. In [36], Jeon *et al.* considered a multi-hop cognitive network on top of a primary network and assumed that the secondary nodes know the location of each primary node regardless of whether it is a transmitter (TX) or a receiver (RX). With an elegant transmission scheme, they showed that by defining a preservation region around each primary node, both networks can achieve the same throughput scaling law as a stand-alone wireless network, while the secondary network may suffer from a finite outage probability. However, in a practical cognitive network, it is hard for the CRs to know the locations of primary RXs since they may keep passive all the time. As such, it is more reasonable to assume that the secondary network only knows the locations of the primary TXs. Furthermore, the results in [36] are obtained without considering possible positive interactions between the primary network and the secondary network. In practice, the secondary network, which is usually deployed after the existence of the primary network for opportunistic spectrum access, can transport data packets not only for itself but also for the primary network due to their cognitive nature. As such, it is meaningful to investigate whether the throughput and/or delay performance of the primary network (whose protocol was fixed before the deployment of the secondary tier) can be improved with the opportunistic aid of the secondary network, while assuming the secondary net-

work still capable of keeping the same throughput and delay scaling laws as the case where no supportive actions are taken between the two networks. The main focus of this dissertation is to address the above issues.

D. Overview of Contributions

In this dissertation, the fundamental limits of CONs are investigated in terms of the asymptotic throughput and packet delay performance when the numbers of nodes (for both PRs and CRs) approach infinity. We assume that the secondary tier only knows the location of primary TXs (which is different from [36]). According to the interaction pattern of the two tiers, the following two types of CONs are considered: 1) selfish CONs, in which neither the primary tier nor the secondary tier is willing to route the packets for the other, and 2) supportive CONs, in which the secondary tier is willing to route the packets for the primary tier while the primary tier does not.

We first investigate the throughput and delay scaling laws for selfish CONs. We assume that the primary and secondary nodes are all static. Based on such an assumption, we define a preservation region just around each primary TX and propose corresponding transmission schemes for the two tiers. It is shown that in selfish CONs, both tiers can achieve the same throughput and delay scaling laws as a stand-alone network, incurring zero outage for the CRs with high probability.

We then investigate the throughput and delay scaling laws for supportive CONs. We consider the following two scenarios: i) the primary and secondary nodes are all static; ii) the primary nodes are static while the secondary nodes are mobile. With specialized protocols for the secondary tier, we show that the throughput and delay scaling laws of the primary tier could be significantly improved with the aid of the secondary tier, while the secondary tier can still achieve the same throughput and

delay scaling laws as a stand-alone network.

Note that in [14], the authors also pointed out that adding a large amount of extra pure relay nodes (which only relay traffic for other nodes), the throughput scaling can be improved at the cost of excessive network deployment. However, there are two key differences between such a statement in [14] and our results. First, in our work, the added extra relays (the secondary nodes) only access spectrum opportunistically (i.e., they need not to be allocated with primary spectrum resources, given their cognitive nature), while the extra relay nodes mentioned in [14] are regular primary nodes (just without generating their own traffic) who need to be assigned with certain primary spectrum resource in the same way as other primary nodes. As such, based on the cognitive features of the secondary nodes considered in our work, the primary throughput improvement could be achieved in an existing primary network without the need to change its current protocol; while in [14], the extra relay deployment has to be considered in the initial primary network design phase for its protocol to utilize the relays. In other words, the problem considered in this dissertation is how to improve the throughput scaling over an existing primary network by adding another supportive network tier (the secondary cognitive tier), where the primary network is already running a certain protocol as we will discuss later in the dissertation, which is different from the networking scenario considered in [14]. Second, in this dissertation, the extra relays are also source nodes on their own (i.e., they also initiate and support their own traffic within the secondary tier), and as one of the main results shows, even with their help to improve the primary-tier throughput, these extra relays (i.e., the secondary tier) could still achieve the same throughput scaling for their own traffic as a stand-alone network considered in [14].

Finally, the throughput and delay performance of a CON with a small number of nodes is investigated. Specifically, we investigate the power and rate control schemes

for multiple CR links in the same neighborhood, which operate over multiple channels (frequency bands) in the presence of PRs with a delay constraint imposed on data transmission. By further considering practical limitations in spectrum sensing, an efficient algorithm is proposed to maximize the average sum-rate of the CR links over a finite time horizon under the constraints on the CR-to-PR interference and the average transmit power for each CR link. In the proposed algorithm, the PR occupancy pattern of each channel is modeled as a discrete-time Markov chain (DTMC). Based on such a model, a novel power and rate control strategy based on dynamic programming (DP) is derived, which is a function of the spectrum sensing output, the instantaneous channel gains for the CR links, and the remaining power budget for the CR transmitter. Simulation results show that the proposed algorithm leads to a significant performance improvement over heuristic algorithms.

E. Dissertation Organization

The rest of the dissertation is organized as follows. In Chapter II, the throughput and delay scaling laws for selfish CONs are investigated. The system model and main results are described in Section II. A. The proposed protocols for the primary and secondary networks are discussed in Section II. B. The delay and throughput scaling laws for the primary network are established in Section II. C. The delay and throughput scaling laws for the secondary network are derived in Section II. D. Finally, Section II. E summarizes our conclusions.

In Chapter III, the throughput and delay scaling laws for supportive CONs are investigated. The system model is described and the main results are summarized in Section III. A. The proposed protocols for the primary and secondary tiers are described in Section III. B. The delay and throughput scaling laws for the primary

tier are derived in Section III. C. The delay and throughput scaling laws for the secondary tier are studied in Section III. D. Finally, Section III. E summarizes our conclusions.

In Chapter IV, the throughput and packet delay performance of a CON with a small number of nodes is investigated. The system model is described in Section IV. A. A special case of multiple CR links over a single time slot (MCST), which is the building block for Section IV. C, is discussed in Section IV. B. The DP-based power and rate control strategy for multiple CRs over multiple time slots (MCMT) is proposed in Section IV. C. Three heuristic algorithms are discussed in Section IV. D. Numerical results are given in Section IV. E. Finally, Section IV. F summarizes our conclusions.

In Chapter V, the dissertation conclusions are summarized in Section V. A and the future work is discussed in Section V. B.

CHAPTER II

THROUGHPUT AND DELAY SCALING LAWS IN SELFISH COGNITIVE OVERLAID NETWORKS

In this chapter[1], we investigate the throughput and delay scaling laws in selfish CONs, in which neither the primary tier nor the secondary tier is willing to route the packets for the other. We first describe the system model and the main results. We then propose the network protocols for the primary tier and the secondary tier, respectively. Afterwards, we analyze the throughput and delay scaling laws based on our proposed protocols. Finally, we summarize our conclusions in this chapter.

A. System Model and Main Results

In this section, we first describe the system model and assumptions about the CON, and then define the throughput and delay. We use $p(E)$ to represent the probability of event E and claim that an event E_n occurs with high probability $(w.h.p.)$ if $p(E_n) \to 1$ as $n \to \infty$.

1. Network Model

Consider a CON with a static primary tier and a static secondary tier coexisting over a unit square. The primary nodes are distributed according to a Poisson point process (P. P. P.) of density n and randomly grouped into one-to-one S-D pairs. The distribution of the secondary nodes is following a P. P. P. of density m. The secondary nodes are also randomly grouped into one-to-one S-D pairs. As the model in [36], we assume that the density of the secondary tier is higher than that of the primary tier,

[1]The work was submitted for publication to IEEE/ACM Transaction on Networking and IEEE must be contacted if a party wishes to reuse the paper.

i.e.,

$$m = n^{\gamma}, \tag{2.1}$$

with $\gamma > 1$.

For the wireless channel, we only consider the large-scale pathloss and ignore the effects of shadowing and small-scale multipath fading. As such, the normalized channel power gain $g(r)$ is given as

$$g(r) = \frac{A}{r^{\kappa}}, \tag{2.2}$$

where A is a system-dependent constant, r is the distance between the TX and the corresponding RX, and $\kappa > 2$ denotes the pathloss exponent. In the following discussion, we normalize A to be unity for simplicity.

The primary tier and the secondary tier share the same spectrum, time, and space, while the former one is the licensed user of the spectrum and thus has a higher priority to access the spectrum. The secondary tier opportunistically access the spectrum while keeping its interference to the primary tier at an "acceptable level". In this chapter, the "acceptable level" means that the presence of the secondary tier does not degrade the throughput scaling law of the primary tier.

We assume that the secondary tier only knows the locations of the primary TXs and has no knowledge about the locations of the primary RXs. This is the essential difference between our model and the model in [36], where the authors assumed that the secondary tier knows the locations of all the primary nodes. Some other aspects of our model are defined in a similar way to that in [36], as we will discuss later.

2. Transmission Rate and Throughput

The ambient noise is assumed as additive white Gaussian noise (AWGN) with an average power N_0. During each transmission, we assume that each TX-RX pair deploys a capacity-achieving scheme, and the channel bandwidth is normalized to be unity for simplicity. Thus the data rate of the k-th primary TX-RX pair is given by

$$R_p(k) = \log\left(1 + \frac{P_p(k)g\left(\|X_{p,\mathrm{tx}}(k) - X_{p,\mathrm{rx}}(k)\|\right)}{N_0 + I_p(k) + I_{sp}(k)}\right), \tag{2.3}$$

where $\|\cdot\|$ stands for the norm operation, $P_p(k)$ is the transmit power of the k-th primary TX-RX pair, $X_{p,\mathrm{tx}}(k)$ and $X_{p,\mathrm{rx}}(k)$ are the TX and RX locations of the k-th primary TX-RX pair, respectively, $I_p(k)$ is the sum interference from all other primary TXs to the RX of the k-th primary TX-RX pair, and $I_{sp}(k)$ is the sum interference from all the secondary TXs to the RX of the k-th primary TX-RX pair. Specifically, $I_p(k)$ can be written as

$$I_p(k) = \sum_{i=1, i\neq k}^{Q_p} P_p(k)g\left(\| X_{p,\mathrm{tx}}(i) - X_{p,\mathrm{rx}}(k) \|\right), \tag{2.4}$$

where Q_p is the number of active primary TX-RX pairs, and $I_{sp}(k)$ is given by

$$I_{sp}(k) = \sum_{i=1}^{Q_s} P_s(i)g\left(\| X_{s,\mathrm{tx}}(i) - X_{p,\mathrm{rx}}(k) \|\right), \tag{2.5}$$

where Q_s is the number of active secondary TX-RX pairs, $P_s(i)$ is the transmit power of the i-th secondary TX-RX pair, and $X_{s,\mathrm{tx}}(i)$ is the TX location of the i-th secondary TX-RX pair. Likewise, the data rate of the l-th secondary TX-RX pair is given by

$$R_s(l) = \log\left(1 + \frac{P_s(l)g\left(\|X_{s,\mathrm{tx}}(l) - X_{s,\mathrm{rx}}(l)\|\right)}{N_0 + I_s(l) + I_{ps}(l)}\right), \tag{2.6}$$

where $X_{s,\mathrm{rx}}(l)$ is the RX location of the l-th secondary TX-RX pair, $I_s(l)$ is the sum interference from all other secondary TXs to the RX of the l-th secondary TX-RX pair, and $I_{ps}(l)$ is the sum interference from all primary TXs to the RX of the l-th secondary TX-RX pair. Specifically, $I_s(l)$ is given by

$$I_s(l) = \sum_{i=1, i\neq l}^{Q_s} P_s(i) g\left(\parallel X_{s,\mathrm{tx}}(i) - X_{s,\mathrm{rx}}(l) \parallel\right), \tag{2.7}$$

and $I_{ps}(l)$ is given by

$$I_{ps}(l) = \sum_{i=1}^{Q_p} P_p(i) g\left(\parallel X_{p,\mathrm{tx}}(i) - X_{s,\mathrm{rx}}(l) \parallel\right). \tag{2.8}$$

Now we give the definitions of throughput per S-D pair and sum throughput.

Definition 1. *The* throughput per S-D pair $\lambda(n_t)$ *is defined as the average data rate that each source node can transmit to its chosen destination w.h.p. in a multi-hop fashion with a particular scheduling scheme, where n_t is the number of nodes in the network. We have*

$$p\left(\min_{1\leq i\leq n_t/2} \liminf_{t\to\infty} \frac{1}{t} M_i(t) \geq \lambda(n_t)\right) \to 1, \tag{2.9}$$

as $n_t \to \infty$, where $M_i(t)$ is the number of bits that S-D pair i transmitted in t time slots.

Definition 2. *The* sum throughput $T(n_t)$ *is defined as the product between the throughput per S-D pair $\lambda(n_t)$ and the number of S-D pairs in the network, i.e.,*

$$T(n_t) = \frac{n_t}{2} \lambda(n_t). \tag{2.10}$$

According to the network model defined in Section II. A, the number of nodes in the primary tier (or in the secondary tier) is a random variable. However, we

will show in Lemma 1 and Lemma 3 at Section II. B that the number of nodes in the primary tier (or in the secondary tier) will be bounded by functions of the node density *w.h.p.*. As such, in the following discussion, we use $\lambda_p(n)$ and $\lambda_s(m)$ to denote the throughputs per S-D pair for the primary tier and the secondary tier, respectively. We use $T_p(n)$ and $T_s(m)$ to denote the sum throughputs for the primary tier and the secondary tier, respectively.

3. Fluid Model and Delay

As in [21], we use a fluid model to study the delay-throughput tradeoffs for the primary and secondary tiers. In this model, we divide each time slot into multiple packet slots, and the size of the data packets can be scaled down to arbitrarily small with the increase of the node density n (or m).

Definition 3. *The* delay $D(n_t)$ of a packet *is defined as the average time that it takes to reach the destination node after the departure from the source node.*

Let $D_i(j)$ denote the delay of packet j for S-D pair i. The sample mean of delay over all packets transmitted for S-D pair i is defined as

$$D_i = \limsup_{k\to\infty} \frac{1}{k} \sum_{j=1}^{k} D_i(j), \tag{2.11}$$

and the average delay over all S-D pairs is given by

$$\overline{D(n_t)} = \frac{2}{n_t} \sum_{i=1}^{n_t/2} D_i.$$

The average delay over all realizations of the network is

$$D(n_t) = E\left[\overline{D(n_t)}\right] = \frac{2}{n_t} \sum_{i=1}^{n_t/2} E\left[D_i\right]. \tag{2.12}$$

As what we did over the notations of throughput, in the following discussion,

we use $D_p(n)$ and $D_s(m)$ to denote the packet delays for the primary tier and the secondary tier, respectively.

4. Main Results

The main results in this chapter are as follows.

- We propose a transmission scheme for a selfish CON with a primary tier vs. a secondary tier. We assume that the primary tier uses a typical time-slotted adjacent-neighbor transmission protocol (similar to that in [14]) and the secondary tier has a higher density and only knows the locations of the primary TXs. By a properly designed secondary protocol, we show that each secondary source node has a finite opportunity to transmit its packets to the chosen destination *w.h.p.*, i.e., no outage compared with the result in [36].

- For the primary tier, we show that the throughput per S-D pair is $\lambda_p(n) = \Theta(\sqrt{\frac{1}{n\log n}})$ *w.h.p.* and the sum throughput is $T_p(n) = \Theta(\sqrt{\frac{n}{\log n}})$ *w.h.p.*. These results are the same as those in a stand-alone ad hoc wireless network considered in [14]. Following the fluid model [21], we give the delay-throughput tradeoff for the primary tier as $D_p(n) = \Theta(n\lambda_p(n))$ for $\lambda_p(n) = O(\frac{1}{\sqrt{n\log n}})$, which is the optimal delay-throughput tradeoff for a stand-alone wireless ad hoc network established in [21].

- For the secondary tier, we prove that the throughput per S-D pair is $\lambda_s(m) = \Theta(\sqrt{\frac{1}{m\log m}})$ *w.h.p.* and the sum throughput is $T_s(m) = \Theta(\sqrt{\frac{m}{\log m}})$ *w.h.p.*. Although due to the presence of the preservation regions, the secondary packets seemingly experience larger delays compared with that of the primary tier, we show that the delay-throughput tradeoff for the secondary tier is the same as that in the primary tier, i.e., $D_s(m) = \Theta(m\lambda_s(m))$ for $\lambda_s(m) = O(\frac{1}{\sqrt{m\log m}})$.

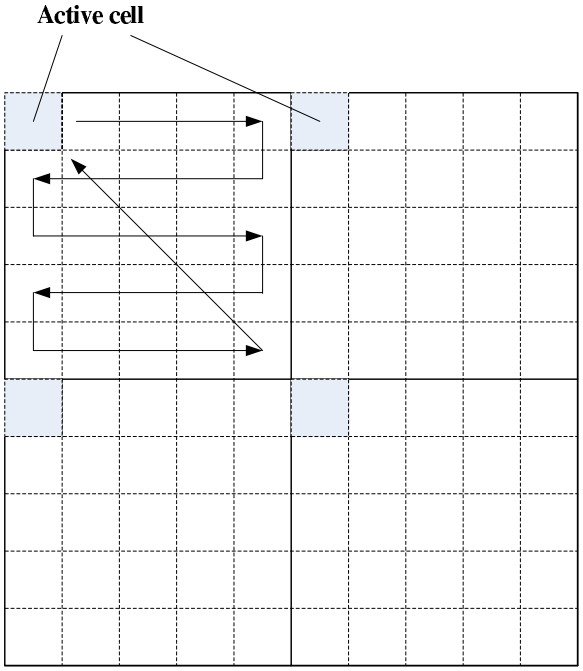

Fig. 2. A four-cluster example with 25 cells per cluster.

B. Network Protocols

In our proposed scheme, the primary tier deploys a modified time-slotted multi-hop transmission scheme over that in [36]. The secondary tier adapts its protocol according to the primary transmission scheme. We first describe the primary protocol, then introduce the secondary protocol, and finally give a lemma to show that with our proposed protocols the secondary users can communicate without outage *w.h.p.*. Similarly as in [21] [36], we claim that an outage event occurs when a node has zero opportunity to communicate. The outage probability is defined as the fraction of nodes that have zero opportunity to communicate.

1. Primary Protocol

- We divide the unit square into small-square primary cells. The area of each primary cell is $a_p = \frac{k_1 \log n}{n}$, with $k_1 \geq 1$.

- We group the primary cells into primary clusters, and each cluster has K_p primary cells. Note that the number of primary cells in a primary cluster has to satisfy $K_p \geq 25$ such that there is no outage for the secondary tier (See Lemma 5 for details). For convenience, we take $K_p = 25$ throughout the chapter. We split the transmission time into time division multiple access (TDMA) frames, where each frame has 25 time slots that correspond to the number of cells in each primary cluster with each slot of length t_p. In each time slot, one cell in each primary cluster is chosen to be active. The cells in each primary cluster take turns to be active in a round-robin fashion. All primary clusters follow the same 25-TDMA transmission pattern, as shown in Fig. 2.

- We define the data path along which the packets traverse as the horizontal line and then the vertical line connecting a source and its corresponding destination, as shown in Fig. 3. One node within a primary cell is defined as a designated relay node, which is responsible for relaying the packets of all the data paths passing through the cell. The packets will be forwarded from cell to cell by the relay nodes first along the horizontal data path (HDP), then along the vertical data path (VDP). Nodes in a particular cell take turns to serve as the designated relay node.

- When a primary cell is active, it transmits a single packet for each of the data paths passing through the cell. The transmission is also deployed in a TDMA fashion. The TDMA frame structure for the primary tier is shown in Fig. 4,

where one packet slot is assigned to one S-D data path that passes through or originates from a particular primary cell. As such, the number of packet slots is determined by the total number of data paths in the cell, which is based on the so-called fluid model [21]. The specific packet transmission procedure is as follows:

  - The designated relay node first transmits a single packet for each of the S-D paths passing through the cell; and then each of the source nodes within the cell takes turns to transmit a single packet.
  - The receiving node must be located in one of the neighboring primary cells along the predefined data path, unless it is a destination node, which may be located in the same cell. If the next-hop of the packet is the final destination, it will be directly delivered to the destination node; otherwise, the packet will be transmitted to a designated relay node.
  - The designated relay node in each primary cell maintains a buffer to temporarily store the packets received from its neighboring cells, and each packet will be transmitted to the next hop in the next active time slot of the cell.

- At each packet slot, the TX node transmits with power of $P_0 a_p^{\frac{\kappa}{2}}$, where P_0 is a constant.

The primary protocol is similar to that in [21] but with different data paths and TDMA transmission patterns. As a result, we have the following two lemmas.

Lemma 1. *Let n_{pt} denote the number of total primary nodes in the unit square; then we have $\frac{n}{2} < n_{pt} < en$* w.h.p..

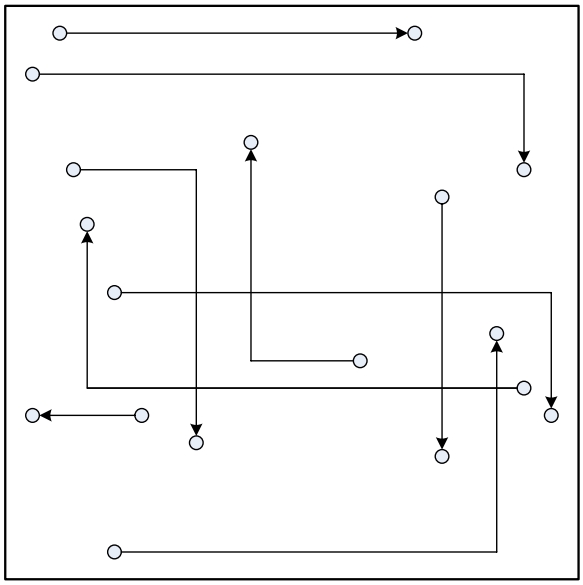

Fig. 3. Examples of HDPs and VDPs for the primary S-D pairs.

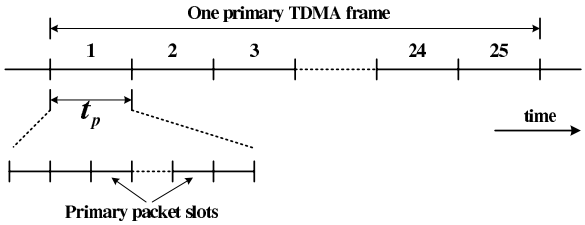

Fig. 4. Structure of the primary TDMA frame (for selfish CONs).

Proof. Since n_{pt} is a Poisson random variable with parameter $\mu = n$, using the Chernoff bound (Theorem 5.4 in [37]), we have

$$\begin{aligned} p\left(n_{pt} \leq \frac{n}{2}\right) &\leq \frac{e^{-n}(en)^{\frac{n}{2}}}{\left(\frac{n}{2}\right)^{\frac{n}{2}}} \\ &= \left(\frac{2}{e}\right)^{\frac{n}{2}} \to 0 \end{aligned} \tag{2.13}$$

as $n \to \infty$, and

$$\begin{aligned} p\left(n_{pt} \geq en\right) &\leq \frac{e^{-n}(en)^{en}}{(en)^{en}} \\ &= e^{-n} \to 0 \end{aligned} \tag{2.14}$$

as $n \to \infty$. Combining (2.13) and (2.14) via the union bound, we obtain

$$p\left(n_{pt} \leq \frac{n}{2} \text{ or } n_{pt} \geq en\right) \leq p\left(n_{pt} \leq \frac{n}{2}\right) + p\left(n_{pt} \geq en\right) \to 0$$

as $n \to \infty$. Hence

$$p\left(\frac{n}{2} < n_{pt} < en\right) = 1 - p\left(n_{pt} \leq \frac{n}{2} \text{ or } n_{pt} \geq en\right) \to 1$$

as $n \to \infty$, which completes the proof. □

We recall the following useful lemma from [33].

Lemma 2. *(Lemma 5.7 in [33]) For $k_1 \geq 1$, each primary cell contains at least one but no more than $k_1 e \log n$ primary nodes* w.h.p..

2. Secondary Protocol

- We divide the unit area into square secondary cells with size $a_s = \frac{k_2 \log m}{m}$, with $k_2 \geq 1$. Without loss of generality, we choose $k_2 = k_1$ in the following discussion.

- We group the secondary cells into secondary clusters. Each secondary cluster

has K_s cells. Note that the value of K_s is not necessarily the same as that of K_p as long as $K_s \geq 9$. Here we choose $K_s = K_p = 25$ for simplicity. Similar to the primary protocol, the secondary tier also follows a 25-TDMA pattern to communicate with t_s slot length. We let the duration of each secondary TDMA frame equal to that of one primary time slot. The relationship between the primary TDMA frame and the secondary TDMA frame is shown in Fig. 5, where each secondary time slot is further divided into packet slots.

- To limit the interference from the secondary nodes to the primary nodes, we define a preservation region as a square containing M^2 secondary cells around a particular primary cell in which an active primary TX (not the RX) is located, where M is an integer and the value will be defined later. No secondary nodes in the preservation regions are allowed to transmit.

- The designated relay nodes and data paths for the secondary tier are defined in the same way as those for the primary tier. As shown in Fig. 6, when a particular secondary cell outside the preservation region is active, its designated relay node transmits a single packet for each of the data paths passing through the cell, and each of the secondary source nodes within the cell takes turns to transmit a single packet. The packet is transmitted to the next-hop relay node or the destination node in neighboring secondary cells along the HDP or VDP path. Note that if the RX node is the destination node, it may be located in the same cell, as we discussed for the primary protocol.

- When a secondary cell falls into a preservation region[2], its designated relay

[2]Note that the secondary nodes located in the preservation regions can still receive packets from TXs outside the preservation regions, although they are not permitted to transmit packets.

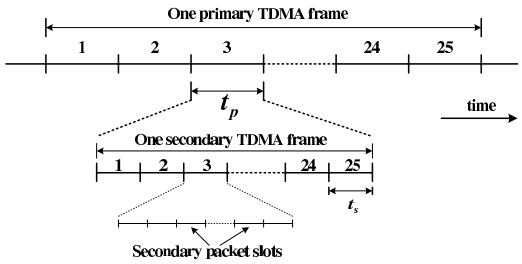

Fig. 5. Structure of the secondary TDMA frame and its relationship with the primary TDMA frame (for selfish CONs).

node buffers the packets that it receives; it waits until the preservation region is cleared and the cell is active to deliver the packets to the next hop.

- At each packet slot, the active secondary TX node transmits with power of $P_1 a_s^{\frac{\kappa}{2}}$, where P_1 is a constant.

Similarly as in the primary tier case, we have the following two lemmas for the secondary tier.

Lemma 3. *Let n_{st} denote the total number of secondary nodes in the unit square; then we have $\frac{m}{2} < n_{st} < em$* w.h.p..

Proof. The proof is similar to that of Lemma 1. □

Lemma 4. *(Lemma 5.7 in [33]) For $k_2 \geq 1$, each secondary cell contains at least one but no more than $k_2 e \log m$ secondary nodes* w.h.p..

Now, let us discuss how to choose the value of M, i.e., the size for the preservation region. Considering the fact that the primary TX may only transmit to a node in its adjacent cells or within the same cell, the preservation region should accommodate at least 9 primary cells to protect the potential primary RX. Since the primary RX may be located close to the outer boundary of the 9-cell region, we should add another layer of protective secondary cells. As such, any active secondary TXs outside the

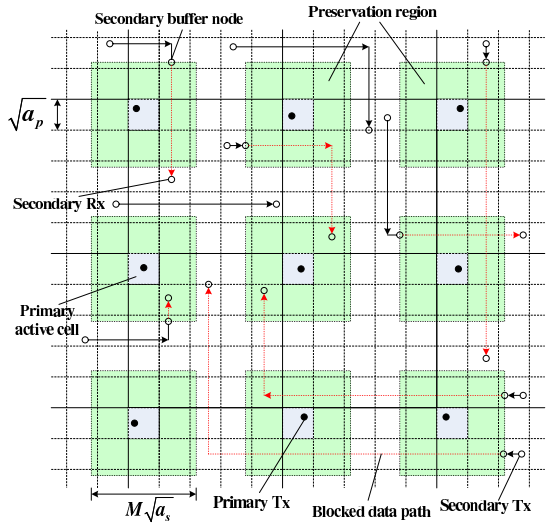

Fig. 6. Preservation region and examples of secondary data paths.

preservation region are at least certain-distance-away from the potential primary RX. Therefore, we define the side length of the preservation square region as

$$M\sqrt{a_s} \geq 3\sqrt{a_p} + 2\epsilon_p, \tag{2.15}$$

where $\epsilon_p > 0$ defines the width of the protective secondary strip around the 9 primary cells in the preservation region. There is a tradeoff in choosing the value of ϵ_p. If we choose a larger ϵ_p, the interference from the secondary tier to the primary tier will be less. However, the opportunity for the secondary tier to access the spectrum will also be less since the unpreserved area in the unit square will be reduced. In the following discussion, we set $\epsilon_p = \sqrt{a_s}$ for simplicity. Accordingly, the minimum value of M can

be set as

$$\begin{aligned} M &= \lfloor \frac{3\sqrt{a_p} + 2\sqrt{a_s}}{\sqrt{a_s}} \rfloor \\ &= \lfloor 3\sqrt{\frac{a_p}{a_s}} \rfloor + 2 \\ &\approx 3\sqrt{\frac{n^{\gamma-1}}{\gamma}}, \end{aligned} \tag{2.16}$$

where $\lfloor \cdot \rfloor$ denotes the flooring operation. In the last equation of (2.16), we applied $a_p = \frac{k_1 \log n}{n}$, $a_s = \frac{k_2 \log m}{m}$, $k_1 = k_2$, and (2.1), assuming that n is large enough. In the following discussion, "n is large" or "n is large enough" means that, for a fixed γ, n is chosen to satisfy $a_s \ll a_p$. For example, when $k_1 = k_2$, $\gamma = 2$, $n = 1000$, we have $m = 1000000$ and $\frac{a_p}{a_s} = \frac{n^{\gamma-1}}{\gamma} = 500$.

Note that the preservation region defined here is larger than that in [36] due to the fact that we only know the locations of primary TXs. If a secondary node falls inside a preservation region, it will be silenced. If not, it may become active and has an opportunity to transmit its packets. Accordingly, we call the unpreserved region as the "active region". Since the locations of preservation regions change periodically according to the active time slots in the primary TDMA frame, from the point view of a specific secondary node, it is periodically located in the active region. We define the following terminology to measure the fraction of time in which a secondary cell is located in the active region.

Definition 4. *The* opportunistic factor *of a secondary cell is defined as the fraction of time in which it is located in the active region.*

We use the following lemma to show that, with the protocols defined previously, each individual secondary source node has a finite opportunity to transmit its packets to the chosen destination *w.h.p.*.

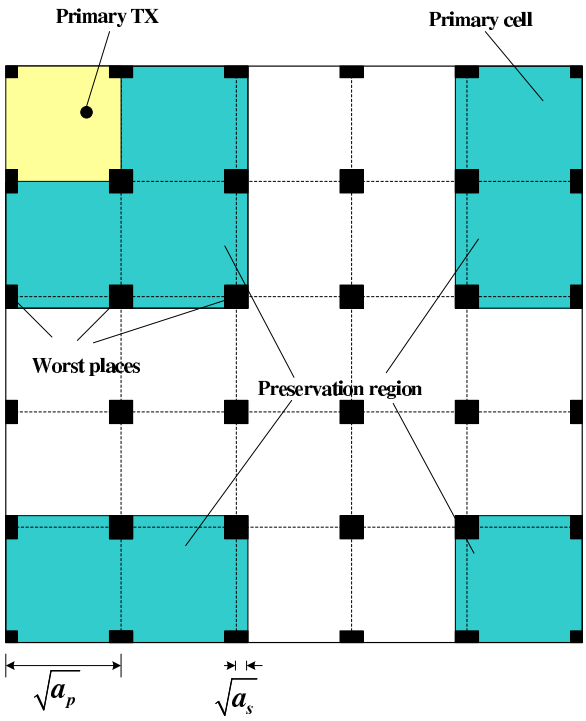

Fig. 7. Preservation regions and worst places in one primary cluster.

Lemma 5. *With the proposed transmission protocol, we have the following results:*

- *The opportunistic factor for a secondary cell is* $\frac{9}{K_p} \leq \eta \leq \frac{16}{K_p}$ *with* $K_p \geq 25$*, for* n *is large enough.*
- *Each individual secondary node has a finite opportunity to transmit its packets to the chosen destination, i.e., zero outage,* w.h.p..

Proof. Consider one primary cluster of K_p primary cells as shown in Fig. 7, where the preservation regions are illustrated as the shaded area when the upper-left primary cell is active in this and neighboring clusters. The primary cells will take turns to be active over time (see Fig. 2) and the locations of the preservation regions will change accordingly. We can easily verify that any point in the cluster has a finite opportunity to be in the active region when n is large. However, during each period of a primary TDMA frame, the fractions of time for different secondary nodes to be

in the active region are not the same. The worst places are the squares with side length of $2\sqrt{a_s}$ around the vertices of each primary cell, as shown by those deeply-shaded small squares in Fig. 7. The opportunistic factor of the secondary cells in these squares is $\frac{9}{K_p}$. The best places are the squares with side length of $\sqrt{a_p} - 2\sqrt{a_s}$ inside each primary cell, as shown by the deeply-shaded squares in Fig. 8. The opportunistic factor of the secondary cells in these squares are $\frac{16}{K_p}$. When the secondary cell lies in other places, the opportunistic factor is between $\frac{9}{K_p}$ and $\frac{16}{K_p}$.

The condition that a secondary node is located in the active region is not sufficient to ensure that it can transmit packets to the destination along the predefined data path. Recall that the secondary tier also deploys a TDMA scheme with adjacent-neighbor transmission. The sufficient condition to ensure that each individual secondary node has a finite chance to transmit packets is that the secondary cell in which the node is located will be assigned with at least one active secondary TDMA slot within each secondary frame, whenever the cell is in the active region. Since in each primary time slot, we have one complete secondary TDMA frame in our protocol, the above sufficient condition is indeed satisfied.

Based on the above discussions, during each period of a primary TDMA frame, each secondary cell has a finite opportunity to be located in the active region with an opportunistic factor of $\frac{9}{K_p} \leq \eta \leq \frac{16}{K_p}$, and each of them is assigned with a secondary TDMA slot. According to the secondary protocol, when a secondary cell is active, each packet buffered in this cell will be assigned with a packet slot *w.h.p.* to be transmitted, since the total number of data paths that pass through or originate from each secondary cell is upper-bounded *w.h.p.* (see Lemma 10 in Section II. D). Thus, the packets from any secondary source node have a finite opportunity to be transmitted along the predefined data path to the chosen destination *w.h.p.*. This completes the proof for the zero outage property. □

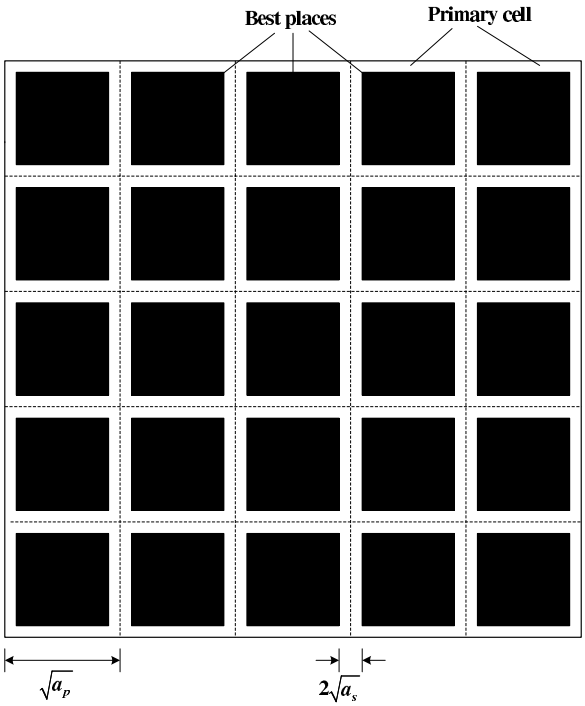

Fig. 8. The best places in one primary cluster.

There is a significant difference between our result here and that in [36]. The authors in [36] defined preservation regions of 9 secondary cells around each primary node, and the positions of the preservation regions are fixed. If the secondary nodes are located in the preservation regions, they will never be active. Therefore, the secondary tier in [36] usually suffers from a non-zero outage probability, even though the outage probability is upper-bounded *w.h.p.*. In our case, each secondary node has a finite opportunity to be active such that we have zero outage *w.h.p.*.

C. Delay and Throughput Analysis for the Primary Tier

In this section, we discuss the delay and throughput scaling laws as well as the delay-throughput tradeoff for the primary tier. The main results are given in three theorems. We first present the delay and throughput scaling laws, then establish the delay-throughput tradeoff for the primary tier.

1. Delay Analysis for the Primary Tier

The packet delay for the primary tier is given by the following theorem.

Theorem 1. *According to the primary protocol in Section II. B, the packet delay is given by*

$$D_p(n) = \Theta\left(\frac{1}{\sqrt{a_p(n)}}\right), \quad w.h.p.. \tag{2.17}$$

Proof. We first derive the average number of hops for each packet to traverse along the primary S-D data path, then use the fact that the time for each primary packet to spend at each hop is a constant, $25t_p$, as shown in Fig. 4 where , and finally calculate the average delay for each primary S-D pair .

Since each primary hop spans a distance of $\Theta\left(\sqrt{a_p(n)}\right)$ *w.h.p.*, the number of hops for a primary packet along the S-D data path i is $\Theta\left(\frac{d_p(i)}{\sqrt{a_p(i)}}\right)$ *w.h.p.*, where $d_p(i)$ is the length of the primary S-D data path i. Hence, the number of hops traversed by a primary packet, averaged over all S-D pairs, is $\Theta\left(\frac{2}{n_{pt}}\sum_{i=1}^{n_{pt}/2}\frac{d_p(i)}{\sqrt{a_p(n)}}\right)$ *w.h.p.*.

The data path length $d_p(i)$ is a random variable, with a maximum value of 2. According to the law of large numbers, as $n_{pt} \to \infty$, the average distance between primary S-D pairs is

$$\frac{2}{n_{pt}}\sum_{i=1}^{n_{pt}/2} d_p(i) = \Theta(1).$$

Therefore, the average number of hops for a primary packet to traverse is $\Theta\left(\frac{1}{\sqrt{a_p(n)}}\right)$ *w.h.p.*. Since we use a fluid model such that the packet size of the primary tier scales proportionally to the throughput $\lambda_p(n)$, each packet arrived at a primary cell will be transmitted in the next active time slot of the cell. As such, the maximum time spent at each primary hop for a particular packet is $25t_p$. Hence, the average delay for each

primary packet is given by

$$D_p(n) = \Theta\left(\frac{25t_p}{\sqrt{a_p(n)}}\right) = \Theta\left(\frac{1}{\sqrt{a_p(n)}}\right), \quad w.h.p, \tag{2.18}$$

which completes the proof. □

The above proof follows the same logic as the proof of Theorem 4 in [21]. The two differences are that we use HDPs and VDPs as the packet routing paths instead of the direct S-D links and we use a different TDMA transmission pattern.

2. Throughput Analysis for the Primary Tier

For the primary tier, the throughput per S-D pair and the sum throughput scaling laws are given in the following theorem.

Theorem 2. *With the primary protocol defined in Section II. B, the primary tier can achieve the following throughput per S-D pair and sum throughput* w.h.p.*:*

$$\lambda_p(n) = \Theta\left(\sqrt{\frac{1}{n\log n}}\right) \tag{2.19}$$

and

$$T_p(n) = \Theta\left(\sqrt{\frac{n}{\log n}}\right). \tag{2.20}$$

Before we give the proof of the above theorem, we first give two lemmas, then use these lemmas to prove the theorem. The main logical flows in the proofs of these lemmas and the theorem are motivated by that in [36] and [28].

Lemma 6. *With the primary protocol defined in Section II. B, each TX node in a primary cell can support a constant data rate of K_1, where $K_1 > 0$ is independent of n.*

Proof. In a given primary packet slot, suppose we have Q_p active primary cells and Q_s active secondary cells. The data rate supported for a TX node in the i-th active primary cell can be calculated as follows:

$$R_p(i) = \frac{1}{25} \log \left(1 + \frac{P_p(i) g(\|X_{p,\text{tx}}(i) - X_{p,\text{rx}}(i)\|)}{N_0 + I_p(i) + I_{sp}(i)} \right), \tag{2.21}$$

where $\frac{1}{25}$ denotes the rate loss due to the 25-TDMA transmission in the primary tier. Note that since there is only one active primary link initiated in each primary cell at a given time, we index the active link initiated in the i-th active primary cell as the i-th active primary link in the whole network. In Fig. 9, we show the primary interference sources to the primary RX of the i-th active primary link, where the shaded cells represent the active primary cells based on the 25-TDMA protocol. From the figure, we see that we have 8 primary interferers with a distance of at least $3\sqrt{a_p}$, 16 primary interferers with a distance of at least $7\sqrt{a_p}$, and so on. Thus, $I_p(i)$ is upper-bounded as

$$\begin{aligned} I_p(i) &= \sum_{k=1, k\neq i}^{Q_p} P_p(k) g(\|X_{p,\text{tx}}(k) - X_{p,\text{rx}}(i)\|) \\ &< P_0 \sum_{t=1}^{\infty} 8t(4t-1)^{-\kappa} \\ &= I_p < \infty, \end{aligned} \tag{2.22}$$

where we used the relationship that $P_p(k) = P_0 a_p^{\frac{\kappa}{2}}$ for all k's and the fact that the series $\sum_{t=1}^{\infty} 8t(4t-1)^{-\kappa}$ converges to a constant for $\kappa > 2$ (see Remark 6.4 in [33]). Due to the preservation regions, a minimum distance $\sqrt{a_s}$ can be guaranteed from all

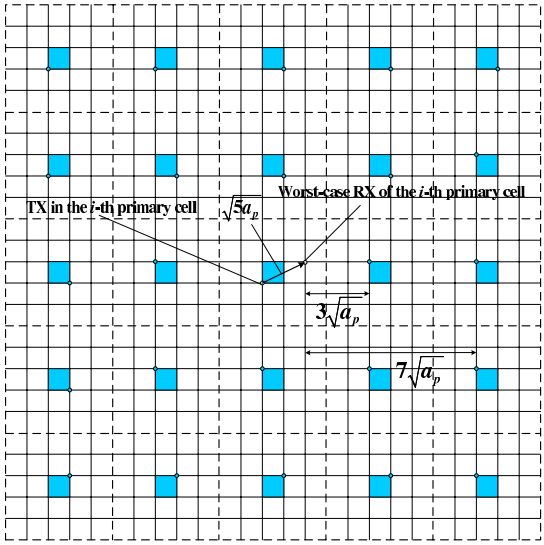

Fig. 9. Interference from the concurrent primary transmissions to the worst-case primary RX of the transmission from the i-th primary cell.

secondary active TXs to any active primary RXs. Thus, $I_{sp}(i)$ is upper-bounded as

$$\begin{aligned} I_{sp}(i) &= \sum_{k=1}^{Q_s} P_s(k) g(\|X_{s,\mathrm{tx}}(k) - X_{p,\mathrm{rx}}(k)\|) \\ &\quad + P_1 a_s^{\frac{\kappa}{2}} \left(\sqrt{a_s}\right)^{-\kappa} \\ &< P_1 \sum_{t=1}^{\infty} 8t(4t-1)^{-\kappa} + P_1 \\ &= I_{sp} < \infty, \end{aligned} \tag{2.23}$$

where we used the fact that $P_s(k) = P_1 a_s^{\frac{\kappa}{2}}$ for all k's. Therefore, we have

$$R_p(i) > \frac{1}{25} \log \left(1 + \frac{P_0 (\sqrt{5})^{-\kappa}}{N_0 + I_p + I_{sp}} \right) = K_1 > 0, \tag{2.24}$$

where the relationship that $\|X_{p,\mathrm{tx}}(i) - X_{p,\mathrm{rx}}(i)\| \leq \sqrt{5a_p}$ is used (see Fig. 9). This completes the proof. □

Lemma 7. *For $a_p(n) = k_1 \log n/n$, the number of primary S-D paths (including both HDPs and VDPs) that pass through or originate from each primary cell is*

$O\left(n\sqrt{a_p(n)}\right)$ w.h.p..

Proof. See the proof of Lemma 3 in [36] or the proof of Lemma 2 in [28]. □

Now we give the proof for Theorem 2.

Proof. Consider the proof of the per-node throughput in (2.19). We need to show that there are deterministic constants $c_2 > 0$ and $c_1 < +\infty$ to satisfy

$$\lim_{n\to\infty} p\left(\frac{c_2}{\sqrt{n\log n}} \leq \lambda_p(n) \leq \frac{c_1}{\sqrt{n\log n}}\right) = 1. \tag{2.25}$$

A loose upper bound of the per-node throughput for the primary tier is achieved when the secondary tier is absent. Gupta and Kumar [14] have already showed that such an upper bound given in (2.25) exists. We then only need to consider the proof for the lower bound.

Since a given TX node in each primary cell can support a constant data rate of K_1 (see Lemma 6), each primary S-D pair can achieve a data rate of at least K_1 divided by the maximum number of data paths that pass through and originate from the primary cell. From Lemma 7, we know that the number of data paths that pass through or originate from each primary cell is $O\left(n\sqrt{a_p(n)}\right)$ *w.h.p.*. Therefore, the throughput per S-D pair $\lambda_p(n)$ is lower-bounded by $\Omega\left(\frac{1}{n\sqrt{a_p(n)}}\right)$ *w.h.p.*, i.e., the lower bound is $\Omega\left(\frac{1}{\sqrt{n\log n}}\right)$ *w.h.p.*.

From Lemma 1, the number of primary S-D pairs is lower-bounded by $\frac{n}{4}$ *w.h.p.*. Thus, the sum throughput $S_p(n)$ is lower-bounded by $\frac{n}{4}\lambda_p(n)$ *w.h.p.*, i.e., the lower bound is $\Omega\left(\sqrt{\frac{n}{\log n}}\right)$ *w.h.p.*. The upper bound of $S_p(n)$ is already established in [14]. This completes the proof. □

From the proof of Theorem 2, the throughput per S-D pair for the primary tier can be written as

$$\lambda_p(n) = \Theta\left(\frac{1}{n\sqrt{a_p(n)}}\right), \quad w.h.p.. \tag{2.26}$$

3. Delay-throughput Tradeoff for the Primary Tier

Combining the results in (2.17) and (2.26), the delay-throughput tradeoff for the primary tier is given by the following theorem.

Theorem 3. *With the primary protocol defined in Section II. B, the delay-throughput tradeoff is*

$$D_p(n) = \Theta\left(n\lambda_p(n)\right), \ \ for \ \ \lambda_p(n) = O\left(\frac{1}{\sqrt{n\log n}}\right). \tag{2.27}$$

D. Delay and Throughput Analysis for the Secondary Tier

The difference between the primary and the secondary transmission schemes arises from the presence of the preservation regions. When their paths are blocked by the preservation regions, the secondary relay nodes buffer the packets and wait until the next hop is available. Due to the presence of the preservation region, the secondary packets will experience a larger delay compared with the primary packets. However, the average packet delay per hop for each secondary S-D data path is still a constant as we discussed later. Thus, we can show that the throughput scaling law and the delay-throughput tradeoff for the secondary tier are the same as those in the primary tier. In the following discussion, we first analyze the average packet delay, then discuss the throughput scaling law, and finally describe the delay-throughput tradeoff.

1. Delay Analysis for the Secondary Tier

The average packet delay for the secondary tier is given by the following theorem.

Theorem 4. *According to the proposed secondary tier protocol in Section II. B, the packet delay is given by*

$$D_s(m) = \Theta\left(\frac{1}{\sqrt{a_s(m)}}\right), \quad w.h.p.. \tag{2.28}$$

Before giving the proof of Theorem 4, we present the following lemma.

Lemma 8. *The average packet delay for each secondary hop is* $\Theta(1)$.

Proof. Let $D^j_{s,h}(i)$ denote the packet delay for the secondary tier over hop j and S-D pair i. As shown in Fig. 5, if there are no preservation regions, each secondary cell has one active time slot in each primary time slot. In another word, each secondary packet will experience a worst-case delay of t_p at each hop, i.e., $D^j_{s,h}(i) = t_p$. When we have the preservation regions, according to Lemma 5, $D^j_{s,h}(i)$ is a bounded random variable. It depends on the location of the active TX from which the secondary packet departs. As shown in Fig. 5 and Fig. 7, when the active TX is located in the worst places as shown in Fig. 7, $D^j_{s,h}(i)$ is $\frac{1}{\eta_{\min}} t_p$, where $\eta_{\min} = \frac{9}{25}$ is the minimum value of the opportunistic factor η. Similarly, when the active TX is located in the best places as shown in Fig. 8, $D^j_{s,h}(i)$ is $\frac{1}{\eta_{\max}} t_p$, where $\eta_{max} = \frac{16}{25}$ is the maximum value of the opportunistic factor η. Hence, the ensemble average of $D^j_{s,h}(i)$ will be a constant c_0, where $\frac{1}{\eta_{\max}} t_p < c_0 < \frac{1}{\eta_{\min}} t_p$, i.e., $E\left[D^j_{s,h}(i)\right] = \Theta(1)$. This completes the proof. □

Now, let us prove Theorem 4.

Proof. Since each secondary hop covers a distance of $\Theta\left(\sqrt{a_s(m)}\right)$ *w.h.p.*, and similarly as in the proof of Theorem 2, the average length of each secondary S-D data path is $\Theta(1)$, the average number of hops for each secondary packet is $\Theta\left(\frac{1}{\sqrt{a_s(m)}}\right)$ *w.h.p.*. From Lemma 8, the average packet delay for each secondary hop is $\Theta(1)$. Therefore, the average packet delay for the secondary tier is

$$D_s(m) = \Theta\left(\frac{1}{\sqrt{a_s(m)}}\right)$$

w.h.p., which completes the proof. □

2. Throughput Analysis for the Secondary Tier

For the secondary tier, the throughput scaling law is given by the following theorem.

Theorem 5. *With the secondary protocol defined in Section II. B, the secondary tier can achieve the following throughput per-node and sum throughput* w.h.p.*:*

$$\lambda_s(m) = \Theta\left(\sqrt{\frac{1}{m\log m}}\right) \tag{2.29}$$

and

$$T_s(m) = \Theta\left(\sqrt{\frac{m}{\log m}}\right). \tag{2.30}$$

Similarly as in the primary tier case, we first present two lemmas, then use these lemmas to prove Theorem 5.

Lemma 9. *With the proposed secondary protocol, each TX node in a secondary cell can support a data rate of K_2, where $K_2 > 0$ is independent of m.*

Proof. Due to the presence of the preservation regions, a minimum distance of $1.5\sqrt{a_p}$ from all primary TXs to a specific active secondary RX can be guaranteed. At a given secondary packet slot and at the i-th secondary link (i.e., the active transmission initiated in the i-th secondary cell), the interference from all active primary TXs is

upper-bounded as

$$\begin{aligned} I_{ps}(i) &< P_0 a_p^{\frac{\kappa}{2}} \sum_{t=1}^{\infty} 8t((3t-1)\sqrt{a_p})^{-\kappa} \\ &\quad + P_0 a_p^{\frac{\kappa}{2}} \left(1.5\sqrt{a_p}\right)^{-\kappa} \\ &< P_0 \sum_{t=1}^{\infty} 8t(3t-1)^{-\kappa} + P_0(1.5)^{-\kappa} \\ &\quad = I_{ps} < \infty, \end{aligned} \tag{2.31}$$

where we applied the same technique as in the proof of Lemma 6 to obtain the upper bound. Likewise, $I_s(i)$ is upper-bounded by $I_s = P_1 \sum_{t=1}^{\infty} 8t(4t-1)^{-\kappa}$, which converges to a constant (see Remark 6.4 in [33]). Considering the effects of the preservation region, the lower bound of the data rate that is supported in each secondary cell can be written as

$$R_s(i) > \frac{1}{25}\eta_{\min} \log\left(1 + \frac{P_0(\sqrt{5})^{-\kappa}}{N_0 + I_{ps} + I_s}\right) = K_2 > 0, \tag{2.32}$$

where $\eta_{\min} = \frac{9}{25}$ represents the penalty due to the presence of the preservation region. Thus, we can guarantee a constant data rate $K_2 > 0$ for a given TX node in each secondary cell, which completes the proof. □

Lemma 10. *For $a_s(m) = k_2 \log m/m$, the number of secondary S-D paths (including both HDPs and VDPs) that pass through or originate from each secondary cell is $O\left(m\sqrt{a_s(m)}\right)$* w.h.p..

Proof. The proof of Lemma 10 follows the same logic as that in the proof of Lemma 7. □

Now, let us prove Theorem 5.

Proof. The proof of Theorem 5 is similar to the proof of Theorem 2. □

Similarly as in Theorem 2, the throughput per S-D pair of the secondary tier can be written as

$$\lambda_s(m) = \Theta\left(\frac{1}{m\sqrt{a_s(m)}}\right), \quad w.h.p.. \tag{2.33}$$

3. Delay-throughput Tradeoff for the Secondary Tier

Combining the results in (2.28) and (2.33), the delay-throughput tradeoff for the secondary tier is given by the following theorem.

Theorem 6. *With the secondary protocol defined in Section II. B, the delay-throughput tradeoff for the secondary tier is*

$$D_s(m) = \Theta\left(m\lambda_s(m)\right), \ \textit{for } \lambda_s(m) = O\left(\frac{1}{\sqrt{m\log m}}\right). \tag{2.34}$$

E. Summary

In this chapter, we studied the selfish CON where neither the primary tier nor the secondary tier is willing to route the packets for the other. When the secondary tier has a higher density, with our proposed protocols, both of the two tiers can achieve the throughput scaling law promised by Gupta and Kumar in [14]. Comparing with the recent result in [36], we only assumed the knowledge about the primary TX locations and there is no outage penalty for the secondary nodes. By using a fluid model, we also showed that both tiers can achieve the same delay-throughput tradeoff as the optimal one established for a stand-alone wireless network in [21].

CHAPTER III

THROUGHPUT AND DELAY SCALING LAWS IN SUPPORTIVE COGNITIVE OVERLAID NETWORKS

In this chapter[1], we investigate the throughput and delay scaling laws in supportive CONs, in which the secondary tier is willing to route the packets for the primary tier while the primary tier does not. We first describe the system model and the main results. We then propose the network protocols for the primary tier and the secondary tier, respectively. Afterwards, we analyze the throughput and delay scaling laws based on our proposed protocols. Finally, we summarize our conclusions in this chapter.

A. System Model and Main Results

Consider a supportive CON with a primary tier and a denser secondary tier over a unit square. We assume that the primary nodes are static, and consider the following two scenarios: i) the secondary nodes are also static; ii) the secondary nodes are mobile. We first describe the network model, the interaction model between the two tiers, and the mobility models for the mobile secondary nodes in the second scenario. Then we summarize the main results in terms of the delay and throughput scaling laws for the proposed two-tier network.

1. Network Model

The network model for supportive CONs is the same as that for selfish CONs. Refer to Section II. A for details.

[1]The work was submitted for publication to IEEE/ACM Transaction on Networking and IEEE must be contacted if a party wishes to reuse the paper.

2. Interaction Model

As shown in the previous work [36], although the opportunistic data transmission in the secondary tier does not degrade the scaling law of the primary tier, it may reduce the throughput in the primary tier by a constant factor due to the fact that the interference from the secondary tier to the primary tier cannot be reduced to zero. To completely compensate the throughput degradation or even improve the throughput scaling law of the primary tier in the two-tier setup, we could allow certain positive interactions between the two tiers. Specifically, we assume that the secondary nodes are willing to act as relay nodes for the primary tier, while the primary nodes are not assumed to do so. When a primary source node transmits packets, the surrounding secondary nodes could pretend to be primary nodes to relay the packets (which is feasible since they are software-programmable cognitive radios). In the scenario where the primary and secondary nodes are all static, the secondary nodes chop the received primary packets into smaller pieces suitable for secondary-tier transmissions. The small data pieces will be reassembled before they are delivered to the primary destination nodes. In the scenario where the secondary nodes are mobile, the received packets are stored in the secondary nodes and delivered to the corresponding primary destination node only when the secondary nodes move into the neighboring area of the primary destination node. As such, the primary tier is expected to achieve better throughput and/or delay scaling laws. More details can be found in the secondary protocols proposed in Section III. B. Note that, these "fake" primary nodes do not have the same priority as the real primary nodes in terms of spectrum access, i.e., they can only use the spectrum opportunistically in the same way as a regular secondary node. The assumption that the secondary tier is allowed to relay the primary packets is the essential difference between our model and the

models in [36].

3. Mobility Model

In the scenario where the secondary nodes are mobile, we assume that the positions of the primary nodes are fixed whereas the secondary nodes stay static in one primary time slot[2] and change their positions at the next slot. In particular, we consider the following two mobility models for the secondary nodes.

Two-dimensional i.i.d. mobility model [25]: The secondary nodes are uniformly and randomly distributed in the unit area at each primary time slot. The node locations are independent of each other, and independent from time slot to time slot, i.e., the nodes are totally reshuffled over each primary time slot.

Two-dimensional random walk (RW) model [21] [39]: We divide the unit square into $1/S$ small-square RW-cells, each of them with size S. The RW-cells are indexed by (x, y), where $x, y \in \{1, 2, \cdots, 1/\sqrt{S}\}$. A secondary node that stays in a RW-cell at a particular primary time slot will move to one of its eight neighboring RW-cells at the next slot with equal probability (i.e., 1/8). For the convenience of analysis, when a secondary node hits the boundary of the unit square, we assume that it jumps over the opposite edge to eliminate the edge effect [21] [39]. The nodes within a RW-cell are uniformly and randomly distributed. Note that the unit square are also divided into primary cells and secondary cells in the proposed protocols as discussed in Section III. B, which are different from the RW-cells defined above. In this chapter, we only consider the case where the size of the RW-cell is greater than or equal to that of the primary cell.

---

[2]As we will see in Section III. B, the data transmission is time-slotted in the primary and secondary tiers.

4. Main Results

We summarize the main results in terms of the throughput and delay scaling laws for supportive CONs here. The definitions of the throughput and delay are the same as those in Chapter II (refer to Section II. B and Section II. C for details). We first present the results for the scenario where the primary and secondary nodes are all static and then describe the results for the scenario with mobile secondary nodes.

i)The primary and secondary nodes are all static.

- It is shown that the primary tier can achieve a per-node throughput scaling of $\lambda_p(n) = \Theta\left(1/\log n\right)$ and a delay scaling of $D_p(n) = \Theta\left(\sqrt{n^{\gamma}\log n}\lambda_p(n)\right)$ for $\lambda_p(n) = O\left(1/\log n\right)$.
- It is shown that the secondary tier can achieve a per-node throughput scaling of $\lambda_s(m) = \Theta\left(\frac{1}{\sqrt{m\log m}}\right)$ and a delay scaling of $D_s(m) = \Theta(m\lambda_s(m))$ for $\lambda_s(m) = O\left(\frac{1}{\sqrt{m\log m}}\right)$.

ii) The primary nodes are static and the secondary nodes are mobile.

- It is shown that the primary tier can achieve a per-node throughput scaling of $\lambda_p(n) = \Theta\left(1/\log n\right)$, and delay scaling laws of $\Theta(1)$ and $\Theta(1/S)$ with the i.i.d. mobility model and the RW mobility model, respectively.
- It is shown that the secondary tier can achieve a per-node throughput scaling of $\lambda_p(n) = \Theta(1)$, and delay scaling laws of $\Theta(m)$ and $\Theta\left(m^2 S\log\frac{1}{S}\right)$ with the i.i.d. mobility model and the RW mobility model, respectively.

B. Network Protocols

In this section, we describe the proposed protocols for the primary tier and the secondary tier, respectively. The primary tier deploys a modified time-slotted multi-hop

transmission scheme from those for the primary network in [36], while the secondary tier chooses its protocol according to the given primary transmission scheme.

1. Primary Protocol

The main sketch of the primary protocol is given as follows:

- Divide the unit square into small-square primary cells with size $a_p(n)$. In order to maintain the full connectivity within the primary tier even without the aid of the secondary tier and enable the possible support from the secondary tier (see Theorem 10 for details), we have $a_p(n) \geq \sqrt{2}\gamma \log n/n$ such that each cell has at least one primary node w.h.p..

- Group every N_c primary cells into a primary cluster. The cells in each primary cluster take turns to be active in a round-robin fashion. We divide the transmission time into TDMA frames, where each frame has N_c primary time slots that correspond to the number of cells in each primary cluster. Note that the number of primary cells in a primary cluster has to satisfy $N_c \geq 64$ such that we can appropriately arrange the preservation regions and the collection regions, which will be formally defined later in the secondary protocol. For convenience, we take $N_c = 64$ throughout the chapter.

- Define the S-D data path along which the packets are routed from the source node to the destination node: The data path follows a horizontal line and a vertical line connecting the source node and the destination node, which is the same as that defined in [36]. Pick an arbitrary node within a primary cell as the designated relay node, which is responsible for relaying the packets of all the data paths passing through the cell.

- When a primary cell is active, each primary source node in it takes turns to transmit one of its own packets with probability p. Afterwards, the designated relay node transmits one packet for each of the S-D paths passing through the cell. The above packet transmissions follow a time-slotted pattern within the active primary time slot, which is divided into packet slots. Each source node reserves a packet slot no matter it transmits or not. If the designated relay node has no packets to transmit, it does not reserve any packet slots. For each packet, if the destination node is found in the adjacent cell, the packet will be directly delivered to the destination. Otherwise, the packet is forwarded to the designated relay node in the adjacent cell along the data path. At each packet transmission, the TX node transmits with power of $Pa_p^{\frac{\kappa}{2}}(n)$, where P is a constant.

- We assume that all the packets for each S-D pair are labelled with serial numbers (SNs). The following handshake mechanism is used when a TX node is scheduled to transmit a packet to a destination node: The TX sends a request message to initiate the process; the destination node replies with the desired SN; if the TX has the packet with the desired SN, it will send the packet to the destination node; otherwise, it stays idle. As we will see in the proposed secondary protocol for the scenario with mobile secondary nodes, the helping secondary relay nodes will take advantage of the above handshake mechanism to remove the outdated (already-delivered) primary packets from their queues. We assume that the length of the handshake message is negligible compared to that of the primary data packet in the throughput analysis for the primary tier as discussed in Section III. C.

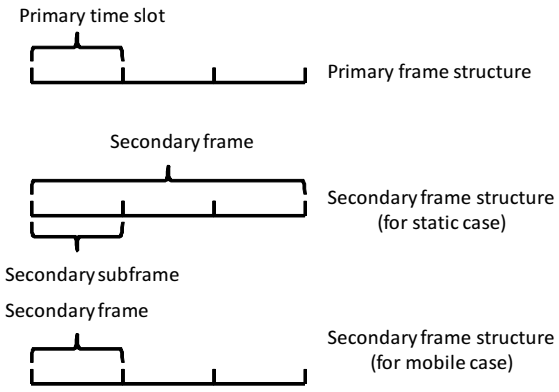

Fig. 10. Frame relationship between the two tiers (for supportive CONs).

Note that running of the above protocol for the primary tier is independent of whether the secondary tier is present or not. When the secondary tier is absent, the primary tier can achieve the throughput scaling law as a stand-alone network discussed in [14]. When the secondary tier is present as shown in Section III. C, the primary tier can achieve a better throughput scaling law with the aid of the secondary tier.

2. Secondary Protocol

In the following, we first present the proposed secondary protocol for the scenario with static secondary nodes, and then describe the one for the scenario with mobile secondary nodes.

Protocol for Static Secondary Tier

We assume that the secondary nodes have the necessary cognitive features such as software-programmability to "pretend" as primary nodes such that they could be chosen as the designated primary relay nodes within a particular primary cell. As later shown by Lemma 12 in Section III. C, a randomly selected designated relay node for the primary packet in each primary cell is a secondary node w.h.p.. Once a

secondary node is chosen to be a designated primary relay node for primary packets, it keeps silent and receives broadcasted primary packets during active primary time slots when only primary source nodes transmit their packets. Furthermore, we use the time-sharing technique to guarantee successful packet deliveries from the secondary nodes to the primary destination nodes as follows. We divide each secondary frame into three equal-length subframes, such that each of them has the same length as one primary time slot as shown in Fig. 10. The first subframe is used to transmit the secondary packets within the secondary tier. The second subframe is used to relay the primary packets to the next relay nodes. Accordingly, the third subframe of each secondary frame is used to deliver the primary packets from the intermediate destination nodes[3] in the secondary tier to their final destination nodes in the primary tier. Specifically, for the first subframe, we use the following protocol:

- Divide the unit area into square secondary cells with size $a_s(m)$. In order to maintain the full connectivity within the secondary tier, we have to guarantee $a_s(m) \geq 2 \log m/m$ with a similar argument to that in the primary tier.

- Group the secondary cells into secondary clusters, with each secondary cluster of 64 cells. Each secondary cluster also follows a 64-TDMA pattern to communicate, which means that the first subframe is divided into 64 secondary time slots.

- Define a preservation region as nine primary cells centered at an active primary TX and a layer of secondary cells around them, shown as the square with dashed edges in Fig. 11. Only the secondary TXs in an active secondary cell outside

[3]An "intermediate" destination node of a primary packet within the secondary tier is a chosen secondary node in the primary cell within which the final primary destination node is located.

all the preservation regions can transmit data packets; otherwise, they buffer the packets until the particular preservation region is cleared. When an active secondary cell is outside the preservation regions in the first subframe, it allows the transmission of one packet for each secondary source node and for each S-D path passing through the cell in a time-slotted pattern within the active secondary time slot w.h.p.. The routing of secondary packets follows similarly defined data paths as those in the primary tier.

- At each transmission, the active secondary TX node can only transmit to a node in its adjacent cells with power of $Pa_s^{\frac{\kappa}{2}}(m)$.

In the second subframe, only secondary nodes who carry primary packets take the time resource to transmit. Note that each primary packet is broadcasted from the primary source node to its neighboring primary cells where we assume that there are N secondary nodes in the neighboring cell along the primary data path successfully decode the packet and ready to relay. In particular, each secondary node relays $1/N$ portion of the primary packet to the intermediate destination node in a multi-hop fashion, and the value of N is set as

$$N = \Theta\left(\sqrt{\frac{m}{\log m}}\right). \tag{3.1}$$

From Lemma 11 in Section IV, we can guarantee that there are more than N secondary nodes in each primary cell w.h.p. when $\gamma \geq 2$. When $1 < \gamma < 2$, the number of the secondary nodes in each primary cell is less than N w.h.p.. In this regime, the proposed protocols could be modified by using the maximum number of the secondary nodes in the neighboring primary cell of a primary TX along the S-D data path. We leave this issue in our future work. The specific transmission scheme in the second subframe is the same as that in the first subframe, where the subframe is divided into

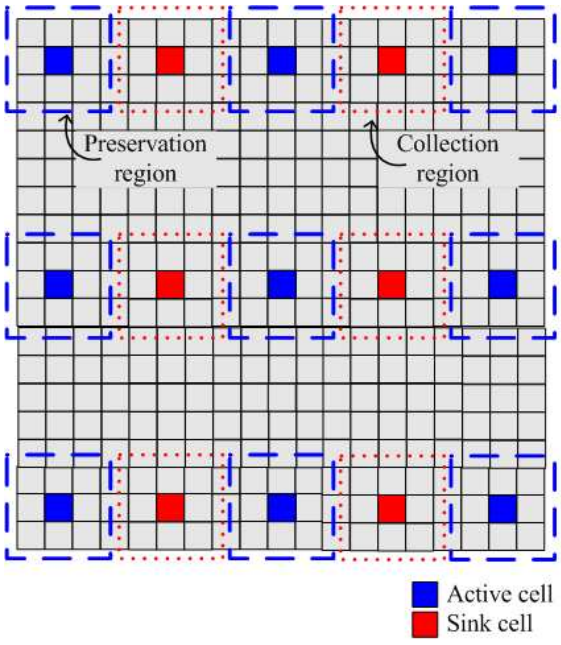

Fig. 11. Preservation regions and collection regions.

64 time slots and all the traffic is for primary packets.

At the intermediate destination nodes, the received primary packet segments are reassembled into the original primary packets. Then in the third subframe, we use the following protocol to deliver the packets to the primary destination nodes:

- Define a collection region as nine primary cells and a layer of secondary cells around them, shown as the square with dotted edges in Fig. 11, where the collection region is located between two preservation regions along the horizontal line and they are not overlapped with each other.

- Deliver the primary packets from the intermediate destination nodes in the secondary tier to the corresponding primary destination nodes in the sink cell, which is defined as the center primary cell of the collection region. The primary destination nodes in the sink cell take turns to receive data by following a time-slotted pattern, where the corresponding intermediate destination node

in the collection region transmits by pretending as a primary TX node. Given that the third subframe is of an equal length to one primary slot, each primary destination node in the sink cell can receive one primary packet from the corresponding intermediate destination node.

- At each transmission, the intermediate destination node transmits with the same power as that for a primary node, i.e., $Pa_p^{\frac{\kappa}{2}}(n)$.

Protocol for Mobile Secondary Tier

Like in the scenario with static secondary nodes, we assume that the secondary nodes have the necessary cognitive features to "pretend" as primary nodes such that they could be chosen as the designated primary relay nodes within a particular primary cell. Divide the transmission time into TDMA frames, where the secondary frame has the same length as that of one primary time slot as shown in Fig. 10. To limit the interference to primary transmissions, we define preservation regions in a similar way to that in the scenario with static secondary nodes.

To faciliate the description of the secondary protocol, we define the *separation threshold time* of random walk as [40]

$$\tau = \min\{t : s(t) \leq e^{-1}\} \tag{3.2}$$

where $s(t)$ measures the separation from the stationary distribution at time t, which is given by

$$\begin{aligned} s(t) &= \min\Big\{ s : p_{(x,y),(u,v)}(t) \geq (1-s)\pi_{(u,v)}, \\ &\qquad \text{for all } x, y, u, v \in \{1, 2, \cdots, 1/\sqrt{S}\} \Big\} \end{aligned} \tag{3.3}$$

where $p_{(x,y),(u,v)}(t)$ denotes the probability that a secondary node hits RW-cell (u, v) at time t starting from RW-cell (x, y) at time 0, and $\pi_{(u,v)} = S$ is the probability of

staying at RW-cell (u, v) at the stationary state. We have $\tau = \Theta(1/S)$ [40].

The secondary nodes perform the following two operations according to whether they are in the preservation regions or not:

i) If a secondary node is in a preservation region, it is not allowed to transmit packets. Instead, it receives the packets from the active primary transmitters and store them in the buffer for future deliveries. Each secondary node maintains Q separate queues for each primary S-D pair. For the i.i.d. mobility model, we take $Q = 1$, i.e., only one queue is needed for each primary S-D pair. For the RW model, Q takes the value of τ given by (3.2). The packet received at time slot t is considered to be 'type k' and stored in the kth queue, if $\left\{\lfloor \frac{t}{64} \rfloor \bmod Q\right\} = k$, where $\lfloor x \rfloor$ denotes the flooring operation.

ii) If a secondary node is not in a preservation region, it transmits the primary and secondary packets in the buffer. In order to guarantee successful deliveries for both primary and secondary packets, we evenly and randomly divide the secondary S-D pairs into two classes: Class I and Class II. Define a collection region in a similar way to that in the scenario with static secondary nodes. In the following, we describe the operations of the secondary nodes of Class I based on whether they are in the collection regions or not. The secondary nodes of Class II perform a similar task over switched timing relationships with the odd and even primary time slots.

- If the secondary nodes are in the collection regions, they keep silent at the odd primary time slots and deliver the primary packets at the even primary time slots to the primary destination nodes in the sink cell, which is defined as the center primary cell of the collection region. In a particular primary time slot, the primary destination nodes in the sink cell take turns to receive packets following a time-slotted pattern. For a particular primary destination node at

time t, we choose an arbitrary secondary node in the sink cell to send a request message to the destination node. The destination node replies with the desired SN, which will be heard by all secondary nodes within the nine primary cells of the collection region. These secondary nodes remove all outdated packets for the destination node, whose SNs are lower than the desired one. For the i.i.d. mobility model, if one of these secondary nodes has the packet with the desired SN and it is in the sink cell, it sends the packet to the destination node. For the RW model, if one of these secondary nodes has the desired packet in the kth queue with $k = \left\{ \lfloor \frac{t}{64} \rfloor \bmod Q \right\}$ and it is in the sink cell, it sends the packet to the destination node. At each transmission, the secondary node transmits with the same power as that for a primary node, i.e., $Pa_p^{\frac{\kappa}{2}}(n)$.

- If the secondary nodes are not in the collection regions, they keep silent at the even primary time slots and transmit secondary packets at the odd primary time slots as follows. Divide the unit square into small-square secondary cells with size $a_s(m) = 1/m$ and group every 64 secondary cells into a secondary cluster. The cells in each secondary cluster take turns to be active in a round-robin fashion. In a particular active secondary cell, we could use Scheme 2 in [21] to transmit secondary packets with power of $Pa_s^{\frac{\kappa}{2}}(m)$ within the secondary tier.

C. Throughput and Delay Analysis for the Primary Tier

In the following, we first present the throughput and delay scaling laws for the primary tier in the scenario where the primary and secondary nodes are all static, and then discuss the scenario where the secondary nodes are mobile.

1. The Scenario with Static Secondary Nodes

We first give the throughput and delay scaling laws for the primary tier, followed by the delay-throughput tradeoff.

Throughput Analysis

In order to obtain the throughput scaling law, we first give the following lemmas.

Lemma 11. *The numbers of the primary nodes and secondary nodes in each primary cell are* $\Theta(na_p(n))$ *and* $\Theta(ma_p(n))$ *w.h.p., respectively.*

The proof can be found in Appendix A.

Lemma 12. *If the secondary nodes compete to be the designated relay nodes for the primary tier by pretending as primary nodes, a randomly selected designated relay node for the primary packet in each primary cell is a secondary node w.h.p..*

Proof. Let η denote the probability that a randomly selected designated relay node for the primary packet in a particular primary cell is a secondary node. We have $\eta = \frac{\Theta(ma_p(n))}{\Theta(ma_p(n)+na_p(n))}$ from Lemma 11, which approaches one as $n \to \infty$. This completes the proof. □

Lemma 13. *With the protocols given in Section III. B, an active primary cell can support a constant data rate of* K_1*, where* $K_1 > 0$ *independent of* n *and* m*.*

The proof can be found in Appendix B.

Lemma 14. *With the protocols given in Section III. B, the secondary tier can deliver the primary packets to the intended primary destination node at a constant data rate of* K_2*, where* $K_2 > 0$ *independent of* n *and* m*.*

The proof can be found in Appendix B.

Based on Lemmas 11-14, we have the following theorem.

Theorem 7. *With the protocols given in Section III. B, the primary tier can achieve the following throughput per S-D pair and sum throughput w.h.p. when* $\gamma \geq 2$*:*

$$\lambda_p(n) = \Theta\left(\frac{1}{na_p(n)}\right) \tag{3.4}$$

and

$$T_p(n) = \Theta\left(\frac{1}{a_p(n)}\right), \tag{3.5}$$

where $a_p(n) \geq \sqrt{2}\gamma \log n/n$ *and* $a_p(n) = o(1)$*.*

Proof. From Lemma 13 and Lemma 14, we know that the primary TX can pour its packets into the secondary tier at a constant rate $K = \min(K_1, K_2)$. Since the primary nodes take turns to be active in each active primary cell, and the number of the primary nodes in each primary cell is of $\Theta(na_p(n))$ as shown in Lemma 11, the theoretically maximum throughput per S-D pair is of $\Theta(K/na_p(n)) = \Theta(1/na_p(n))$. Next, we show that with the proposed protocols, the maximum throughput scaling is achievable. In the proposed protocols, each primary source node pours all its packets into the secondary tier w.h.p. (from Lemma 12) by splitting data into $\Theta\left(\sqrt{m/\log m}\right)$ secondary data paths, each of them at a rate of $\Theta(\frac{1}{m\sqrt{a_s(m)}})$. Set $\sqrt{a_s(m)} = \frac{na_p(n)}{\sqrt{m \log m}}$, which satisfies $a_s(m) \geq 2\log m/m$. As such, each primary source node achieves a throughput scaling law of $\Theta\left(1/na_p(n)\right)$. Since the total number of primary nodes in the unit square is of $\Theta(n)$ w.h.p., we have $T_p(n) = \Theta(n\lambda_p(n)) = \Theta\left(1/a_p(n)\right)$ w.h.p.. This completes the proof. □

By setting $a_p(n) = \sqrt{2}\gamma \log n/n$, the primary tier can achieve the following throughput per S-D pair and sum throughput w.h.p.:

$$\lambda_p(n) = \Theta\left(\frac{1}{\log n}\right) \tag{3.6}$$

and

$$T_p(n) = \Theta\left(\frac{n}{\log n}\right). \tag{3.7}$$

Delay Analysis

We now analyze the delay performance of the primary tier with the aid of a static secondary tier. In the proposed protocols, we know that the primary tier pours all the primary packets into the secondary tier w.h.p. based on Lemma 12. In order to analyze the delay of the primary tier, we have to calculate the traveling time for the N segments of a primary packet to reach the corresponding intermediate destination node within the secondary tier. Since the data paths for the N segments are along the route and an active secondary cell (outside all the preservation regions) transmits one packet for each data path passing through it within a secondary time slot, we can guarantee that the N segments depart from the N nodes, move hop by hop along the data paths, and finally reach the corresponding intermediate destination node in a synchronized fashion. According to the definition of packet delay, the N segments experience the same delay later given in (3.27) within the secondary tier, and all the segments arrive the intermediate destination node within one secondary slot.

Let L_p and L_s denote the durations of the primary and secondary time slots, respectively. According to the proposed protocols, we have

$$L_p = 64L_s. \tag{3.8}$$

Since we split the secondary time frame into three fractions and use one of them for the primary packet relaying, each primary packet suffers from the following delay:

$$D_p(n) = \frac{3}{64}D_s(m) + C = \Theta\left(\frac{1}{\sqrt{a_s(m)}}\right) \tag{3.9}$$

where the secondary-tier delay $D_s(m)$ is later derived in (3.27), C denotes the average

time for a primary packet to travel from the primary source node to the N secondary relay nodes plus that from the intermediate destination node to the final destination node, which is a constant. We see from (3.9) that the delay of the primary tier is only determined by the size of the secondary cell $a_s(m)$. In order to obtain a better delay performance, we should make $a_s(m)$ as large as possible. However, a larger $a_s(m)$ results in a decreased throughput per S-D pair in the secondary tier and hence a decreased throughput for the primary tier, for the primary traffic traverses over the secondary tier w.h.p.. In Appendix D, we derive the relationship between $a_p(n)$ and $a_s(m)$ in our supportive two-tier setup as

$$a_s(m) = \frac{n^2 a_p^2(n)}{m \log m} \tag{3.10}$$

where we have $a_s(m) \geq 2 \log m/m$ when $a_p(n) \geq \sqrt{2}\gamma \log n/n$.

Substituting (3.10) into (3.9), we have the following theorem.

Theorem 8. *According to the proposed protocols in Section III. B, the primary tier can achieve the following delay w.h.p. when* $\gamma \geq 2$.

$$D_p(n) = \Theta\left(\frac{\sqrt{m \log m}}{n a_p(n)}\right) = \Theta\left(\frac{\sqrt{n^\gamma \log n}}{n a_p(n)}\right). \tag{3.11}$$

Delay-Throughput Tradeoff

Combining the results in (3.4) and (3.11), the delay-throughput tradeoff for the primary tier is given by the following theorem.

Theorem 9. *With the protocols given in Section III. B, the delay-throughput tradeoff in the primary tier is given by*

$$D_p(n) = \Theta\left(\sqrt{n^{\gamma}\log n}\lambda_p(n)\right) \text{ for } \lambda_p(n) = O\left(\frac{1}{\log n}\right). \tag{3.12}$$

2. The Scenario with Mobile Secondary Nodes

Throughput Analysis

In order to obtain the throughput scaling law, we first give the following lemmas.

Lemma 15. *With the protocols given in Section III. B, an active primary cell can support a constant data rate of K_3, where $K_3 > 0$ independent of n and m.*

The proof can be found in Appendix C.

Lemma 16. *With the protocols given in Section III. B, the secondary tier can deliver the primary packets to the intended primary destination node in a sink cell at a constant data rate of K_4, where $K_4 > 0$ independent of n and m.*

The proof can be found in Appendix C.

Based on Lemmas 11-12 and Lemmas 15-16, we have the following theorem.

Theorem 10. *With the protocols given in Section III. B, the primary tier can achieve the following throughput per S-D pair and sum throughput w.h.p.:*

$$\lambda_p(n) = \Theta\left(\frac{1}{na_p(n)}\right) \tag{3.13}$$

and

$$T_p(n) = \Theta\left(\frac{1}{a_p(n)}\right), \tag{3.14}$$

when $a_p(n) \geq \sqrt{2}\gamma \log n/n$ *and* $a_p(n) = o(1)$.

Proof. From Lemma 15 and Lemma 16, we know that a primary TX can pour its packets into the secondary tier at rate $K = \min(K_3, K_4)$ w.h.p.. Since the primary nodes take turns to be active in each active primary cell, and the number of primary source nodes in each primary cell is of $\Theta(na_p(n))$ w.h.p. as shown in Lemma 11, the maximum throughput per S-D pair is of $\Theta\left(K/(na_p(n))\right) = \Theta\left(1/(na_p(n))\right)$ w.h.p.. Next, we show that with the proposed protocols, the above maximum throughput scaling is achievable. In the proposed protocols, we know that a randomly selected designated relay node for the primary packet in each primary cell is a secondary node w.h.p. from Lemma 12. As such, when a primary cell is active, the current primary time slot is just used for the primary source nodes in the primary cell to transmit their own packets w.h.p.. Therefore, the achievable throughput per S-D pair is of $\Theta\left(pK/(na_p(n))\right) = \Theta\left(1/(na_p(n))\right)$ and thus a achievable sum throughput of $\Theta(1/a_p)$ for the primary tier w.h.p.. This completes the proof. □

By setting $a_p(n) = \sqrt{2}\gamma \log n/n$, the primary tier can achieve the following throughput per S-D pair and sum throughput w.h.p.:

$$\lambda_p(n) = \Theta\left(\frac{1}{\log n}\right) \tag{3.15}$$

and

$$T_p(n) = \Theta\left(\frac{n}{\log n}\right). \tag{3.16}$$

Delay Analysis

Based on the proposed supportive protocols, we know that the delay for each primary packet has two components: i) the hop delay, which is the transmission time for two hops (from the primary source node to a secondary relay node and from the secondary relay node to the primary destination node); ii) the queueing delay,

which is the time a packet spends in the relay-queue at the secondary node until it is delivered to its destination. The hop delay is two primary time slots, which can be considered as a constant independent of m and n. Next, we quantify the primary-tier delay performance by focusing on the expected queueing delay at the relay based on the two mobility models described in Section III. A.

a. The i.i.d. Mobility Model

We have the following theorem regarding the delay of the primary tier.

Theorem 11. *With the protocols given in Section III. B, the primary tier can achieve the following delay w.h.p. when $\gamma \geq 2$:*

$$D_p(n) = \Theta(1). \tag{3.17}$$

Proof. According to the secondary protocol, within the secondary tier we have $\Theta(m)$ secondary nodes act as relays for the primary tier, each of them with a separate queue for each of the primary S-D pairs. Therefore, the queueing delay is the expected delay at a given relay-queue. By symmetry, all such relay-queues incur the same delay w.h.p.. For convenience, we fix one primary S-D pair and consider the $\Theta(m)$ secondary nodes together as a virtual relay node as shown in Fig. 12 without identifying which secondary node is used as the relay. As such, we can calculate the expected delay at a relay-queue by analyzing the expected delay at the virtual relay node. Denote the selected primary source node, the selected primary destination node, and the virtual relay node as S, D, and R, respectively. To calculate the expected delay at node R, we first have to characterize the arrival and departure processes. A packet arrives at R when a) the primary cell containing S is active, and b) S transmits a packet. According to the primary protocol in Section III. B, the primary cell containing S

becomes active every 64 primary time slots. Therefore, we consider 64 primary time slots as an observation period, and treat the arrival process as a Bernoulli process with rate p ($0 < p < 1$). Similarly, packet departure occurs when a) D is in a sink cell, and b) at least one of the relay nodes that have the desired packets for D is in the sink cell containing D. Let q detnote the probability that event b) occurs, which can be expressed as

$$\begin{aligned} q &= 1-(1-a_p(n))^M, \\ &\sim 1-e^{-Ma_p(n)}, \\ &\to 1, \text{ as } n \to \infty, \text{ for } \gamma \geq 2, \end{aligned} \tag{3.18}$$

where $f \sim g$ means that f and g have the same limit when $n \to \infty$, $M = \Theta(ma_p(n))$ denotes the number of the secondary nodes that have desired packets for D in the sink cell containing D and belong to Class I (Class II) if D is in a sink cell at even (odd) time slots. As such, the departure process is an asymptotically deterministic process with departure rate $q = 1$. Let W_1 denote the delay of the queue at the virtual relay node based on the i.i.d. model. Thus, the queue at the virtual relay node is an asymptotically Bernoulli/deterministic queue, with the expected queueing delay given by [41]

$$E\{W_1\} = 64\frac{1-p}{q-p} \to 64, \text{ as } n \to \infty, \tag{3.19}$$

where $E\{\cdot\}$ denotes the expectation and the factor 64 is the length of one observation period. Note that the queueing length of this asymptotically Bernoulli/deterministic queue is at most one primary packet length w.h.p..

Next we need to verify that the relay-queue at each of the $\Theta(m)$ secondary nodes is stable over time. Note that based on the proposed protocol every secondary node removes the outdated packets that have the SNs lower than the desired one for D

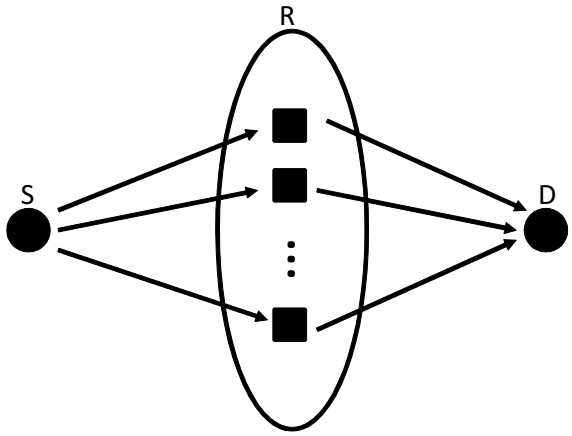

Fig. 12. Illustration of the virtual relay node R.

when it jumps into the sink cell containing D. Since the queueing length at R can be upper-bounded by one, by considering the effect of storing outdated packets, the length of the relay-queue at each secondary node can be upper-bounded by

$$L = n + 1 \tag{3.20}$$

where n can be considered as an upper-bound for the inter-visit time of the primary cell containing D, since $(1 - a_p(n))^n \to 0$ as $n \to \infty$. Thus, the relay-queues at all secondary nodes are stable over time for each given n, which completes the proof. □

b. The RW Mobility Model

For the RW model, we have the following theorem regarding the delay of the primary tier.

Theorem 12. *With the protocols given in Section III. B, the primary tier can achieve the following delay w.h.p. when* $\gamma \geq 2$*:*

$$D_p(n) = \Theta\left(\frac{1}{S}\right) = O\left(\frac{1}{a_p(n)}\right) \tag{3.21}$$

where $S \geq a_p(n)$.

Proof. Like the proof in the i.i.d. mobility case, we fix a primary S-D pair and consider the $\Theta(m)$ secondary nodes together as a virtual relay node. Denote the selected primary source node, the selected primary destination node, and the virtual relay node as S, D, and R, respectively. Based on the proposed secondary protocol in Section III. B, each secondary node maintains $Q = \tau$ queues for each primary S-D pair. Equivalently, R also maintains Q queues for each primary S-D pair where each queue is a concatenated one from $\Theta(m)$ small ones, and the packet that arrives at time t is stored in the kth queue, where $k = \left\{ \lfloor \frac{t}{64} \rfloor \bmod \tau \right\}$. By symmetry, all such queues incur the same expected delay. Without loss of generality, we analyze the expected delay of the kth queue by characterizing its arrival and departure processes. A packet that arrives at time t enters the kth queue when a) the primary cell containing S is active, b) S transmits a packet, and c) $\left\{ \lfloor \frac{t}{64} \rfloor \bmod \tau \right\} = k$. Consider 64τ primary time slots as an observation period. The arrival process is a Bernoulli process with arrival rate p. Similarly, a packet departure occurs at time t when a) D is in a sink cell, b) at least one of the relay nodes that have the desired packets for D is in the sink cell containing D, and c) $\left\{ \lfloor \frac{t}{64} \rfloor \bmod \tau \right\} = k$. Let q denote the probability that event b) occurs during one observation period, which can be expressed as

$$\begin{aligned} q &= 1 - \left(1 - \prod_{i \in \mathcal{I}} q_0 p_{(x_i, y_i)(x_d, y_d)}(t_d) \right), \\ &\geq 1 - (1 - q_0(1 - e^{-1})S)^M, \\ &\sim 1 - e^{-q_0(1-e^{-1})SM}, \\ &\to 1, \text{ as } n \to \infty, \text{ for } \gamma \geq 2, \end{aligned} \tag{3.22}$$

where $\mathcal{I}$ denotes the set of the secondary nodes that have the desired packets for D and belong to Class I (Class II) if D is in a sink cell at even (odd) time slots; (x_i, y_i)

represents the index of the RW-cell, in which the ith secondary node in $\mathcal{I}$ is located when S sends the desired packet; (x_d, y_d) is the index of the RW-cell, in which D is located; t_d stands for the difference between the arrival time and the departure time for the desired packet, which can be lower-bounded by $64(\tau - 1)$; and q_0 denotes the probability that a secondary node is within the sink cell containing D when it moves into RW-cell (x_d, y_d), which is given by $q_0 = a_p(n)/S$. As such, the departure process is an asymptotically deterministic process with departure rate $q = 1$. Let W_2 denote the delay of the queue at node R based on the RW model. Thus, the queue at node R is an asymptotically Bernoulli/deterministic queue, with the queueing delay given by

$$E\{W_2\} = 64\tau \frac{1-p}{q-p} \sim 64\tau = \Theta(\frac{1}{S}), \tag{3.23}$$

where the factor 64τ is the length of one observation period. Since $S \geq a_p(n)$, we have $E\{W_2\} = O\left(1/a_p(n)\right)$.

Using the similar argument as in the i.i.d. case, we can upper-bound the length of the kth relay-queue at any secondary node by (3.20) for any k. Thus, the relay-queues at all secondary nodes are stable, which completes the proof. □

Delay-Throughput Tradeoff

For the RW model, we have the following delay-throughput tradeoff for the primary tier by combining (3.4) and (3.21).

$$D_p(n) = O\left(\frac{n}{\lambda_p(n)}\right), \quad \text{for } \lambda_p(n) = O\left(\frac{1}{\log n}\right). \tag{3.24}$$

We see that the delay-throughput tradeoff for the primary tier with the aid of the secondary tier is even better than the optimal delay-throughput tradeoff given in [21] for a static stand-alone network. Note that the above throughput and delay analysis is based on the assumption $\gamma \geq 2$, and we leave the case with $1 < \gamma < 2$ in

our future work.

D. Throughput and Delay Analysis for the Secondary Tier

1. The Scenario with Static Secondary Nodes

Throughput Analysis

In this section, we discuss the delay and throughput scaling laws for the secondary tier. According to the protocol for the secondary tier, we split the time frame into three equal-length fractions and use one of them for the secondary packet transmissions. Since the above time-sharing strategy only incurs a constant penalty (i.e., 1/3) on the achievable throughput and delay within the secondary tier, the throughput and delay scaling laws are the same as those given in Chapter II, which are summarized by the following theorems.

Theorem 13. *With the secondary protocol defined in Section III. B, the secondary tier can achieve the following throughput per S-D pair and sum throughput w.h.p.:*

$$\lambda_s(m) = \Theta\left(\frac{1}{m\sqrt{a_s(m)}}\right) \tag{3.25}$$

and

$$T_s(m) = \Theta\left(\frac{1}{\sqrt{a_s(m)}}\right), \tag{3.26}$$

where $a_s(m) \geq 2\log m/m$ and the specific value of $a_s(m)$ is determined by $a_p(n)$ as shown in Appendix IV.

Delay Analysis

Theorem 14. *With the secondary protocol defined in Section III. B, the packet delay is given by*

$$D_s(m) = \Theta\left(\frac{1}{\sqrt{a_s(m)}}\right). \tag{3.27}$$

Delay-Throughput Tradeoff

Combining the results in (3.25) and (3.27), the delay-throughput tradeoff for the secondary tier is given by the following theorem.

Theorem 15. *With the secondary protocol defined in Section III. B, the delay-throughput tradeoff is*

$$D_s(m) = \Theta(m\lambda_s(m)), \text{ for } \lambda_s(m) = O\left(\frac{1}{\sqrt{m \log m}}\right). \tag{3.28}$$

For detailed proofs of the above theorems, please refer to Chapter II.

2. The Scenario with Mobile Secondary Nodes

When a secondary RX receives its own packets, it suffers from two interference terms from all active primary TXs and all active secondary TXs. We can use a similar method as in the proof of Lemma 15 to prove that each of the two interference terms can be upper-bounded by a constant independent of m and n. Thus, the asymptotic results for a stand-alone network in [21] [25] hold in this scenario. In the following, we summarize these results for completeness.

Throughput Analysis

We have the following theorem regarding the throughput scaling law for the secondary tier.

Theorem 16. *With the protocols given in Section III. B, the secondary tier can achieve the following throughput per S-D pair and sum throughput w.h.p.:*

$$\lambda_s(m) = \Theta(1) \tag{3.29}$$

and

$$T_s(m) = \Theta(m). \tag{3.30}$$

Delay Analysis

Next, we provide the delay scaling laws of the secondary tier for the two mobility models as discussed in Section II.C.

Theorem 17. *With the protocols given in Section III. B, the secondary tier can achieve the following delay w.h.p. based on the i.i.d. mobility model:*

$$D_s(m) = \Theta(m). \tag{3.31}$$

Theorem 18. *With the protocols given in Section III. B, the secondary tier can achieve the following delay w.h.p. based on the RW model:*

$$D_s(m) = \Theta\left(m^2 S \log \frac{1}{S}\right). \tag{3.32}$$

Note that (3.32) is a generalized result for $S \geq 1/m$. When $S = 1/m$, the delay $D_s(m) = \Theta(m \log m)$ is the same as that in [21].

E. Summary

In this chapter, we studied the throughput and delay scaling laws for a supportive CON, where the secondary tier is willing to relay packets for the primary tier. When the secondary tier has a much higher density, the primary tier can achieve a better throughput scaling law compared to non-interactive overlaid networks. The delay scaling law for the primary tier can also be improved when then the secondary nodes are mobile. Meanwhile, the secondary tier can still achieve the same delay and throughput tradeoff as in a stand-alone network. Based on the fact that an opportunistic supportive secondary tier improves the performance of the primary tier, we make the following observation: The classic time-slotted multi-hop primary protocol [14] does not fully utilize the spatial/temporal resource such that a cognitive

secondary tier with denser nodes could explore the under-utilized segments to conduct nontrivial networking duties.

CHAPTER IV

COGNITIVE OVERLAID NETWORKS WITH A SMALL NUMBER OF NODES

In Chapter II and Chapter III, we discuss the asymptotic performance of CONs as the number of nodes approaches infinity. In this chapter[1], we consider a CON with a small number of nodes and try to gain some design insight from a different perspective. Specifically, we investigate the power and rate control schemes for multiple CR links in the same neighborhood, which operate over multiple channels (frequency bands) in the presence of PRs with a delay constraint imposed on data transmission. We first describe the system model. An efficient algorithm is then proposed to maximize the average sum-rate of the CR links over a finite time horizon under the constraints on the CR-to-PR interference and the average transmit power for each CR link. Finally, we compare the proposed algorithm with three heuristic algorithms.

A. System Model

For the convenience of description, in this chapter, we focus on the case where there are only two CR links in the same neighborhood. The proposed strategies can be easily extended to more general cases of multiple CR links. The two CR links operate over N channels, each of which is of the same bandwidth W and licensed to a primary network whose users communicate in a synchronous slotted manner. Assume that the availability of each channel for each of the two CR links follows the same rule at each time slot according to whether the channel is occupied by PRs or not. Over each time slot, each CR performs spectrum sensing individually to detect the existence of

[1]The work was submitted for publication to IEEE Transaction on Vehicular Technology and IEEE must be contacted if a party wishes to reuse the paper.

PR transmissions in its surroundings. Due to the limitation of hardware and sensing algorithms, the typical spectrum sensing time (e.g., 10 ms as shown in [42]) may not be negligible compared to one CR time slot (e.g., 10 ms specified in IEEE 802.22 [43]). In this case, it is difficult for the CRs to know the instantaneous activities of the PR transmissions. However, the CRs may obtain a delayed spectrum sensing result regarding the previous time slot. As such, in this chapter, we assume that the PR occupancy information would not be available for the CRs until the next time slot.

While the two CRs may have difficulties in obtaining the instantaneous PR occupancy information, we assume that it is relatively easier for them to obtain the channel gains of their own CR links in a real-time manner, since the channel estimation algorithm only needs to observe a certain number of pre-known training symbols while the spectrum sensing algorithm needs to observe a much-longer window over unknown signals. For example, the instantaneous channel gains for the CR links can be obtained by exploring the preamble of each time slot in IEEE 802.22 [43]. Similar assumptions can be found in related works [2] [5] [7] [8]. Assume that there exists a central node (which may be one of the two CR nodes) to coordinate the CR transmissions. At each time slot, the two CRs report the delayed PR occupancy information, the instantaneous channel gains for the CR links, and the remaining power budgets to the central node via a predetermined delay-free (i.e., with a negligible delay) control channel. The central node then determines the channel allocation and calculates the corresponding transmit power and data rate for each CR link. Afterwards, the resulting parameters are sent to the corresponding CR link for data transmission. Assuming that the CRs stay still or move slowly (such that the CRs will stay in the same neighborhood within one transmission period that is limited by a delay constraint T), we focus on designing an efficient CR transmission strategy to maximize the average sum-rate of the two CR links over T time blocks under the constraint of

the average transmit power for each CR link, while keeping the probability of CR-to-PR disturbance at each channel below a given level. Since the CR transmission strategy is highly related to the PR occupancy pattern, we next start with modeling the behavior of PRs.

1. Behavior of PRs

The PRs can arbitrarily access the N channels with an absolute priority. As such, depending on the occupancy of PRs, each channel has two states to the interests of CRs: BUSY or IDLE. For convenience, let us define an indicator function $I_{t,n}$ for the nth channel at time t as

$$I_{t,n} = \begin{cases} 0 & \text{if the } n\text{th channel is BUSY} \\ 1 & \text{if the } n\text{th channel is IDLE} \end{cases}. \tag{4.1}$$

Since the behavior of PRs is correlated in time, we assume that the evolution of each channel independently follows a two-state DTMC as shown in Fig. 13, where α is the transition probability from BUSY to IDLE and β is the one from IDLE to BUSY. We assume that $\alpha < 0.5$ and $\beta < 0.5$[2].

Define an indicator vector $\mathbf{I}_t = \{I_{t,1}, I_{t,2}, \ldots, I_{t,N}\}$, which will be used as a state variable in the DP formulation. Let $\mathcal{S} = \{\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_{2^N}\}$ denote the set of all N-dimensional vectors with binary components. The evolution of $\mathbf{I}_t$ can be described as a DTMC with transition matrix $\mathbf{P}$, whose ij-th element p_{ij} is defined as

$$p_{ij} = \text{Prob}\ \{\mathbf{I}_{t+1} = \mathbf{S}_j | \mathbf{I}_t = \mathbf{S}_i\},\ \mathbf{S}_i, \mathbf{S}_j \in \mathcal{S}. \tag{4.2}$$

[2]If a PR is idle in the current slot, it is more likely to be idle in the next slot, vice versa. This is based on the assumption that the slot length is less than the coherence time of the PR activity random process.

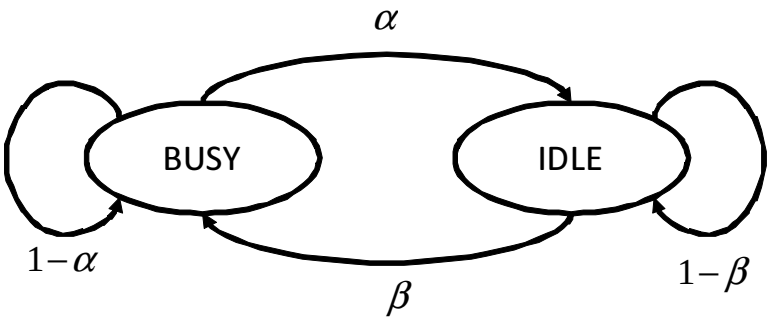

Fig. 13. The two-state DTMC model for the PR occupancy of each channel.

Given α and β, p_{ij} can be expressed as

$$p_{ij} = \prod_{l=1}^{N} \mu^l \tag{4.3}$$

where μ^l is given by

$$\mu^l = \begin{cases} \alpha & \text{if } s_i^l = 0, s_j^l = 1 \\ 1-\alpha & \text{if } s_i^l = 0, s_j^l = 0 \\ \beta & \text{if } s_i^l = 1, s_j^l = 0 \\ 1-\beta & \text{if } s_i^l = 1, s_j^l = 1 \end{cases} \tag{4.4}$$

with s_i^l and s_j^l the lth components of vectors $\mathbf{S}_i$ and $\mathbf{S}_j$, respectively.

Let M_t denote the number of idle channels at time slot t. Obviously, we have

$$M_t = \sum_{n=1}^{N} I_{t,n}. \tag{4.5}$$

Based on the above DTMC channel model, the evolution of M_t can also be described by a DTMC with transition matrix $\mathbf{Q}$, whose ij-th element q_{ij} is defined as

$$q_{ij} = \text{Prob}\ \{M_{t+1} = j | M_t = i\}. \tag{4.6}$$

Given α and β, q_{ij} can be expressed as

$$q_{ij} = \sum_{k=K_1}^{K_2} \begin{pmatrix} i \\ i-j+k \end{pmatrix} \alpha^k \beta^{i-j+k} (1-\alpha)^{N-i-k} (1-\beta)^{j-k} \tag{4.7}$$

where

$$K_1 = \begin{cases} 0 & \text{if } i \geq j \\ j-i & \text{otherwise} \end{cases} \tag{4.8}$$

and

$$K_2 = \begin{cases} N-i & \text{if } i+j \geq N \\ j & \text{otherwise} \end{cases} . \tag{4.9}$$

2. Power Mask Constraints

As being widely adopted in wireless standards (such as 802.15.4) over ISM bands to regulate interference power, power masks can be used to constrain the maximum transmit power of the CR such that the interference to PRs is kept below a certain level [7] [8]. Intuitively, we can set two different power mask values for the two different channel states such that we aggressively use high power when the channel is idle and we strictly limit the transmit power when the channel is BUSY. Specifically, the power mask at time t for the nth channel is given by

$$P_t^{\text{mask}} = \begin{cases} P^L & \text{if } I_{t,n} = 0 \\ P^H & \text{if } I_{t,n} = 1 \end{cases} \tag{4.10}$$

where $P^L \ll P^H$. However, in most practical scenarios, the CR has no access to the exact value of $I_{t,n}$ for the current time slot due to the delayed spectrum sensing output, where we assume that we only know the value for the previous time slot. In such a case, given the fact that $\alpha < 0.5$ and $\beta < 0.5$, i.e., the channel is more likely to stay in the same state as that in the previous time slot, we utilize the temporal

correlation of $I_{t,n}$ to set the power mask for the nth channel at time t as

$$P_t^{\text{mask}} = \begin{cases} P^L & \text{if } I_{t-1,n} = 0 \\ P^H & \text{if } I_{t-1,n} = 1 \end{cases}. \tag{4.11}$$

As such, we guarantee that the PRs who were previously using the channels are not interfered by the CR transmission in the current time slot. However, if a PR jumps into a specific channel, i.e., the value of the indicator function for the channel suddenly changes from 1 to 0 over slots $t-1$ and t, the CR transmission under the power mask given in (4.11) leads to intolerable interference to PR, which we call an outage in this channel. To measure this outage, we define the channel outage probability p_{ch} as the probability of CR transmission with power mask P^H when a PR jumps into the channel. Specifically, if at each time slot the CRs randomly select a portion of previously idle channels to transmit with power mask P^H and the remaining channels to transmit with power mask P^L, the outage probability p_{ch} for each of the channels is the same and can be expressed as

$$p_{\text{ch}} = \rho\beta \tag{4.12}$$

where ρ denotes the percentage of previously idle channels that are randomly selected to transmit with the high power mask. Therefore, a required target p_{ch} set by spectrum regulation bodies can be met by adjusting the parameter ρ.

3. Formulation of Sum-Rate Maximization

Given the knowledge of previous PR occupancy information and the current channel gains for the CR links, our objective is to maximize the average sum-rate of the two CR links over T time slots, while satisfying the constraint on the average transmit

power for each CR link and keeping p_{ch} below a threshold p_{th}, i.e.,

$$\begin{aligned} \max \quad & E\left\{\frac{1}{T}\sum_{t=0}^{T-1}\sum_{k=1}^{2}\sum_{n=1}^{N} c_{t,n}^{(k)}\right\} \\ \text{s. t.} \quad & \frac{1}{T}\sum_{t=0}^{T-1}\sum_{n=1}^{N} P_{t,n}^{(k)} \le P^{(k)}, \quad k=1,2, \quad t=0,\ldots,T-1, \\ & p_{\text{ch}} \le p_{\text{th}}, \end{aligned} \tag{4.13}$$

where $E\{\cdot\}$ denotes the expectation over the distributions of the CR channel gains and the PR occupancy pattern at each time slot; the decision variables $P_{t,n}^{(k)}$ denotes the transmit power on the nth channel for CR link k at time t; $P^{(k)}$ denotes the average transmit power budget for CR link k; and the data rate of the nth channel at slot t for CR k, $c_{t,n}^{(k)}$, is given as

$$c_{t,n}^{(k)} = W \log_2\left(1 + \frac{P_{t,n}^{(k)} H_{t,n}^{(k)}}{\sum_{j\neq k} P_{t,n}^{(j)} C_{t,n}^{(j,k)} + W\sigma^2}\right) \tag{4.14}$$

where $H_{t,n}^{(k)}$ is the channel gain for the nth channel at time t for CR link k, $C_{t,n}^{(j,k)}$ is the crosstalk factor from the transmitter of CR j to the receiver of CR k, and σ^2 is the spectral density of the white Gaussian noise-plus-interference. Note that we assume the noise-plus-interference process as a stationary Gaussian process given the possible large number of in-band interferers. Furthermore, to make the problem more tractable at this stage, we also assume that $H_{t,n}^{(k)}$s are the same across different n's at a given t and k, denoted as $H_t^{(k)}$, which is of independent exponential distribution with unit mean and variance over t and k. Furthermore, we assume that $\mathbf{I}_{t-1}$ and $H_t^{(k)}$s are independent of each other. Note that if needed, we can modify (4.14) to take into account other design considerations such as opportunistic scheduling among CRs.

The solution of the optimization problem (4.13) is not a set of values, but a power

(and accordingly rate) control strategy, which is a function of the delayed indicator vector $\mathbf{I}_{t-1}$, the current power budget $P_t^{(k)}$s for the two CR transmitters at time t, and the instantaneous channel gains $H_t^{(k)}$s at time t. Different realizations of the random processes governing the PR occupancy and the CR channel gains lead to different sets of transmission parameter values. Given the time correlation of the underlying random process of the PR occupancy, a key aspect of this problem is that the power and rate control strategy over time cannot be determined in isolation since we have to balance the current data rate and the expected future rate. The DP-based approach can optimally capture this kind of tradeoff [44-47], which is shown later in Section IV. C. In the following, let us start with the special MCST case: multiple CR links over a single time slot.

B. Power and Rate Control in MCST Case

In this section, we discuss the MCST case: two CR links with $T = 1$, where the problem given in (4.13) can be simplified as

$$\begin{aligned} \max \quad & E\left\{ \sum_{k=1}^{2} \sum_{n=1}^{N} c_{t,n}^{(k)} \right\} \\ \text{s. t.} \quad & \sum_{n=1}^{N} P_{t,n}^{(k)} \le P_t^{(k)}, \quad k = 1, 2, \\ & p_{\text{ch}} \le p_{\text{th}}, \end{aligned} \tag{4.15}$$

where $P_t^{(k)}$ denotes the transmit power budget for CR k at time t. Note that even without the second constraint, (4.15) is an open problem in the context of interference channels. Furthermore, it is difficult for a CR to obtain the exact values of the crosstalk factors from other CRs, which makes the problem even harder to solve. To make the problem tractable, we assume that the channels cannot be shared by

more than one CR, i.e., (4.15) is forced to have a frequency division multiple access (FDMA) solution. As shown later, such an assumption leads to small performance loss when the transmit power budgets $P^{(1)}$ and $P^{(2)}$ are small.

To satisfy the second constraint in (4.15), we first set ρ as

$$\rho = \frac{p_{\text{th}}}{\beta}. \tag{4.16}$$

Accordingly, we randomly select $L_t = \lfloor \rho M_{t-1} \rfloor$ channels from the previously idle channels to transmit data with power mask P^H, and send data over the remaining $N - L_t$ channels with power mask P^L, where $\lfloor \cdot \rfloor$ denotes the flooring operation. As such, the second constraint in (4.15) is satisfied and the optimization problem is transformed into (4.17) after we reorder the indices of the channels according to the assigned power mask values.

$$\begin{aligned} \max \quad & E\left\{ \sum_{k=1}^{2} \sum_{n=1}^{N} c_{t,n}^{(k)} \right\} \\ \text{s. t.} \quad & \sum_{n=1}^{N} P_{t,n}^{(k)} \le P_t^{(k)}, \quad k = 1, 2, \\ & P_{t,n}^{(k)} \le P^H, \quad n = 1, 2, \ldots, L_t \\ & P_{t,n}^{(k)} \le P^L, \quad n = L_t + 1, L_t + 2, \ldots, N, \\ & P_{t,n}^{(k)} P_{t,n}^{(l)} = 0, \forall k \ne l, n = 1, 2, \ldots, N, \end{aligned} \tag{4.17}$$

where the last constraint ensures that (4.17) has a FDMA solution. Unfortunately, (4.17) is still a combinatorial optimization problem that is NP-hard. However, if we know the channel allocation for each CR link, we can easily obtain the optimal $p_{t,n}^{(k)}$ by performing single-user power allocation for each CR over the assigned channels and calculate the corresponding $c_{t,n}^{(k)}$. In the following, we focus on obtaining a near-optimal FDMA channel allocation over the two CRs.

Let $N_{1,t}^{(k)}$ and $N_{2,t}^{(k)}$ denote the numbers of channels assigned to CR link k at time t for data transmissions with power mask P^H and power mask P^L, respectively. Based on the assumption that the channel gains are the same for a specific CR link at each time slot, if the optimal values of $N_{1,t}^{(k)}$s and $N_{2,t}^{(k)}$s for the two CR links are known, the optimal channel allocation can be achieved by arbitrarily setting $N_{1,t}^{(k)}$ channels with power mask P^H and $N_{2,t}^{(k)}$ channels with power mask P^L for CR link k, $k = 1, 2$. In order to obtain the optimal $N_{1,t}^{(k)}$ and $N_{2,t}^{(k)}$, we next cast a continuous FDMA bandwidth allocation problem to approximate (4.17):

$$\begin{aligned}
\max \quad & E\left\{\sum_{k=1}^{2}\sum_{i=1}^{2} W_{i,t}^{k}\log_2\left(1+\frac{P_{i,t}^{(k)}}{W_{i,t}^{(k)}\sigma^2}\right)\right\} \\
\text{s. t.} \quad & P_{1,t}^{(k)} + P_{2,t}^{(k)} \le P_t^{(k)}, \quad k = 1, 2, \\
& \sum_{k=1}^{2} W_{i,t}^{(k)} \le W_{i,t}, \quad i = 1, 2, \\
& \frac{P_{1,t}^{(k)}}{W_{1,t}^{(k)}} \le P^H, \quad k = 1, 2, \\
& \frac{P_{2,t}^{(k)}}{W_{2,t}^{(k)}} \le P^L, \quad k = 1, 2,
\end{aligned} \tag{4.18}$$

where $W_{1,t} = L_t W$ and $W_{2,t} = (N - L_t)W$; $W_{1,t}^{(k)}$ is the bandwidth under power mask P^H and $W_{2,t}^{(k)}$ is the bandwidth under power mask P^L, for CR link k; $P_{1,t}^{(k)}$ and $P_{2,t}^{(k)}$ are the total transmit power values allocated to $W_{1,t}^{(k)}$ and $W_{2,t}^{(k)}$, respectively, for CR link k. It can be shown that (4.18) is a convex optimization problem over $P_{i,t}^{(k)}$ and $W_{i,t}^{(k)}$, which can be easily solved by some existing algorithms (e.g., the interior point method [48]). After obtaining the optimal $W_{1,t}^{(k)}$ and $W_{2,t}^{(k)}$, we quantize them by setting $N_{1,t}^{(k)} = [W_{1,t}^{(k)}/W]$ and $N_{2,t}^{(k)} = [W_{2,t}^{(k)}/W]$, respectively, where $[\cdot]$ denotes the rounding operation. In this way, we can obtain near-optimal $N_{1,t}^{(k)}$s and $N_{2,t}^{(k)}$s (under

FDMA constraint) for both CR links.

After we obtain the channel allocation for the two CR links, we calculate the optimal single-user power allocation for each CR link by solving the following problem for CR link k.

$$\begin{aligned} \max \quad & \sum_{n\in D_t^{(k)}} c_{t,n}^{(k)} \\ \text{s. t.} \quad & \sum_{n\in D_t^{(k)}} P_{t,n}^{(k)} \le P_t^{(k)}, \\ & P_{t,n}^{(k)} \le P^H, \quad n \in D_{1,t}^{(k)} \\ & P_{t,n}^{(k)} \le P^L, \quad n \in D_{2,t}^{(k)}, \end{aligned} \tag{4.19}$$

where $D_{1,t}^{(k)}$ and $D_{2,t}^{(k)}$ denote the sets of the channel indices assigned to CR link k at time t for data transmission with power mask P^H and power mask P^L, respectively; $D_t^{(k)} = D_{1,t}^{(k)} \bigcup D_{2,t}^{(k)}$. We see that (4.19) is a convex optimization problem, where the solution can be derived by solving the KKT conditions [48] and is given by

$$P_{t,n}^{(k)} = \left\{ \begin{array}{ll} \frac{P_t^{(k)}}{N_t^{(k)}}, & \text{if } P_t^{(k)} \le N_t^{(k)} P^L \text{ and } n \in D_t^{(k)} \\ \frac{P_t^{(k)} - N_{2,t}^{(k)} P^L}{N_{1,t}^{(k)}}, & \text{if } N_t^{(k)} P^L \le P_t^{(k)} \le N_{1,t}^{(k)} P^H + N_{2,t}^{(k)} P^L \text{ and } n \in D_{1,t}^{(k)} \\ P^H & \text{if } P_t^{(k)} \ge N_{1,t}^{(k)} P^H + N_{2,t}^{(k)} P^L \text{ and } n \in D_{1,t}^{(k)} \\ P^L, & \text{if } P_t^{(k)} \ge N_t^{(k)} P^L \text{ and } n \in D_{2,t}^{(k)} \end{array} \right\} \tag{4.20}$$

where $N_t^{(k)} = N_{1,t}^{(k)} + N_{2,t}^{(k)}$.

Note that the FDMA assumption and the rounding operation over $N_{i,t}^{(k)}$ make the proposed algorithm suboptimal. Next we evaluate the performance loss of the proposed algorithm due to these relaxations. Since there is no efficient algorithms (except for exhaustive search) to evaluate the exact objective value of (4.15), here

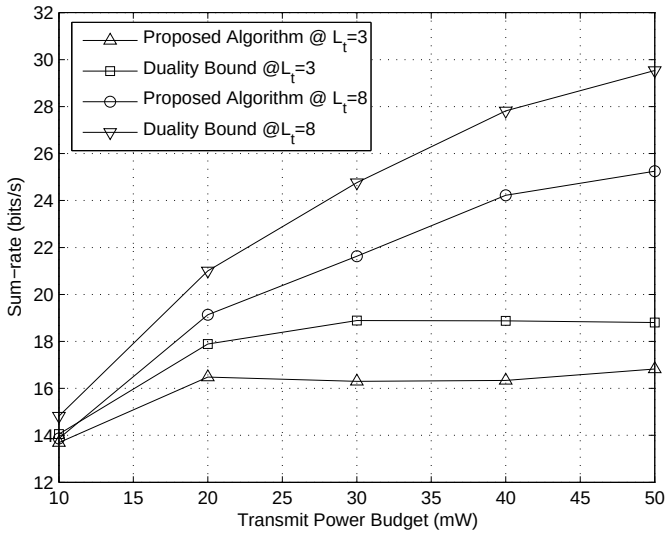

Fig. 14. The objective value of the one-snapshot optimization.

we take a performance upper bound, i.e., the optimal value of its dual problem that is always convex, as a reference. Assume that the two CRs have the same transmit power budget, i.e., $P_t^{(1)} = P_t^{(2)}$. In Fig. 14, we plot the achievable sum-rate of the proposed algorithm over different values of $P_t^{(k)}$ against the optimal objective value of the dual problem for (4.15), which can be evaluated by the dual decomposition method in [49]. The system parameters are set as in Table I and the crosstalk factors are of independent exponential distribution with unit mean and variance. In the figure, we see that the sum-rate increases with L_t given a fixed $P_t^{(k)}$, i.e., a larger p_{th} allows a higher sum-rate. For each L_t, the sum-rate increases with $P_t^{(k)}$ and will become eventually saturated when all channels transmit with the maximum allowable power mask. The performance of the proposed algorithm is close to the duality bound in the low transmit power regime. Across the whole regime, the performance gap is caused by the FDMA assumption, the rounding, and the duality gap between (4.15) and its dual problem.

Table I. System Parameters

Parameter	Value
P^H	10 mW
P^L	1 mW
σ^2	1 mW/Hz
N	10

As we discussed above, if we maximize the sum-rate for the two CR links over one time slot ($T = 1$), the optimal solution can be obtained by solving (4.15). However, when $T > 1$, we have to optimally capture the best tradeoff between the current data rate and the expected future rate, which can be solved as a DP problem. Next we discuss the multiple CR multiple time-slot (MCMT) case in more details.

C. Power and Rate Control in MCMT Case

In this section, we solve the MCMT case: two CR links and $T > 1$, and reformulate the optimization problem (4.13) following the general DP framework and discuss the solution. Let $B_t^{(k)}$ denote the power budget (which bears the same meaning as an energy budget when the time-slot length is fixed) at the beginning of time t for CR k, which evolves as

$$B_{t+1}^{(k)} = B_t^{(k)} - P_t^{(k)} \tag{4.21}$$

where $B_0^{(k)} = TP^{(k)}$ and $P_t^{(k)}$ is defined as

$$P_t^{(k)} = \sum_{n \in D_t^{(k)}} P_{t,n}^{(k)}. \tag{4.22}$$

We first define an immediate reward function $g_t(P_t^{(1)}, P_t^{(2)}, \mathbf{I}_{t-1}, H_t^{(1)}, H_t^{(2)})$, which provides a measure of the maximum sum-rate that the two CR links can achieve at the

current time slot without consideration of the future data rate, given the current power budgets $P_t^{(1)}$ and $P_t^{(2)}$, the current channel gains $H_t^{(1)}$ and $H_t^{(2)}$, and the previous PR occupancy vectors $\mathbf{I}_{t-1}$. We see that if we focus on FDMA solutions, $g_t(\cdot)$ is exactly the optimal objective value of (4.17). Afterwards, we define the reward function $J_t(B_t^{(1)}, B_t^{(2)}, \mathbf{I}_{t-1}, H_t^{(1)}, H_t^{(2)})$ at time t, which provides a measure of the expected sum-rate of the two CR links from time t to time $T-1$. According to the definition, the reward function $J_t(\cdot)$ at time t is the sum of the following two items: the immediate reward function $g_t(\cdot)$ and the expected value of the reward function $J_{t+1}(\cdot)$ over the distributions of $\mathbf{I}_t$, $H_{t+1}^{(1)}$, and $H_{t+1}^{(2)}$ at time $t+1$. Specifically, the function $J_t(\cdot)$ is given by

$$\begin{aligned} &J_t\left(B_t^{(1)}, B_t^{(2)}, \mathbf{I}_{t-1} = \mathbf{S}_i, H_t^{(1)}, H_t^{(2)}\right) = \\ &\max_{0 \le P_t^{(k)} \le B_t^{(k)}} \left\{ g_t\left(P_t^{(1)}, P_t^{(2)}, \mathbf{I}_{t-1} = \mathbf{S}_i, H_t^{(1)}, H_t^{(2)}\right) + \right. \\ &\left. \sum_{\mathbf{S}_j \in \mathcal{S}} p_{ij} \bar{J}_{t+1}\left(B_t^{(1)} - P_t^{(1)}, B_t^{(2)} - P_t^{(2)}, \mathbf{I}_t = \mathbf{S}_j\right) \right\} \end{aligned} \tag{4.23}$$

where $\bar{J}_{t+1}(\cdot)$ denotes the expected value of $J_{t+1}(\cdot)$ over the distributions of $H_{t+1}^{(1)}$ and $H_{t+1}^{(2)}$. Since we are only interested in the average sum-rate of the two CR links over T time slots, the reward function at slot T can be set as

$$J_T\left(B_T^{(1)}, B_T^{(2)}, \mathbf{I}_{T-1}, H_T^{(1)}, H_T^{(2)}\right) = 0. \tag{4.24}$$

By proceeding backward in time from slot $T-1$ to 0, we can finally obtain $J_0(\cdot)$ for all $\mathbf{S}_i \in \mathbf{S}$. The expected value of $\bar{J}_0(\cdot)$ over the distribution of $\mathbf{I}_{-1}$ can be expressed as

$$\bar{\bar{J}}_0\left(B_0^{(1)}, B_0^{(2)}\right) = \frac{1}{T} \sum_{\mathbf{S}_i \in \mathcal{S}} \eta_i \bar{J}_0\left(B_0^{(1)}, B_0^{(2)}, \mathbf{I}_{-1} = \mathbf{S}_i\right) \tag{4.25}$$

where $[\eta_0, \eta_1, \ldots, \eta_{2^N}]$ is the initial distribution of the DTMC with transition matrix

$\mathbf{P}$ given in (4.3). We see that $\bar{\bar{J}}_0(\cdot)$ is the objective value of (4.13). Furthermore, if $P_{t,n}^{(k)*} = \mu_t^*(B_t^{(1)}, B_t^{(2)}, \mathbf{I}_{t-1}, H_t^{(1)}, H_t^{(2)})$ maximizes the righthand side of (4.23) for each set of $\{B_t^{(1)}, B_t^{(2)}, \mathbf{I}_{t-1}, H_t^{(1)}, H_t^{(2)}\}$, the optimal solution for (4.13) is thus obtained.

To obtain the optimal power control policy $\mu = \{\mu_0^*, \mu_1^*, \ldots, \mu_{T-1}^*\}$ and the corresponding objective value $\bar{\bar{J}}_0(\cdot)$, we have to solve T subproblems given by the dynamic recursion. Once we work out the solution, we store the power and rate control policy as a look-up table. The size of the look-up table is of $O(N2^N T^3 L_1^2 L_2^2)$, where L_1 is the total number of quantized levels for the average transmit power budget and L_2 is the total number of quantized levels for the channel gain. During the system operation, the central control node can figure out the optimal transmit power and rate from the table according to real-time system parameters. In this sense, the real-time computation complexity is low. Since the proposed algorithm in this section allows both variable channel allocation and variable transmit power budget at each time slot, we call it the VCA-VPB algorithm.

We now summarize how the delayed spectrum sensing information is used in the VCA-VPB algorithm as follows: Given the previous channel state (busy or idle) for each channel, we set the power mask according to (4.11) for the current time slot, which may cause a PR outage. As such, we add a PR outage constraint in the sum-rate optimization problem given in (4.13), which thus implicitly takes the previous channel state information into account. Furthermore, the previous channel state information is also used in the DP recursion (4.23) to calculate the expected sum-rate in the future time slots.

A typical DP-based algorithm is usually designed to deal with the case with finite T. In the asymptotic case when T and N approach infinity, the two-state Markov chain for the channel occupancy converges to the stationary state at most of time. In the stationary state, we assume that each channel stays busy with probability

π_0 and idle with probability π_1. Let M denote the number of channels stay idle at the stationary state. We have $\frac{M}{N} = \frac{1}{N}\sum_{n=1}^{N} I_n \rightarrow \pi_1$ according to weak law of large numbers (WLLN), i.e., $M \rightarrow \pi_1 N$, as $N \rightarrow \infty$. As such, the number of the idle channels does not change over time and the power allocation only depends on the instantaneous channel gains for the CR links. We conjecture that the optimal power allocation strategy is somehow like water-filling over channel gain distribution as given in [50].

D. Heuristic Algorithms

To evaluate the performance of the proposed VCA-VPB algorithm, we provide three reference suboptimal heuristic algorithms defined as follows.

1. FCA-FPB Algorithm

In this algorithm, we fix the channel allocation for the CRs, each of them taking $N/2$ channels, and fix the power budget for the two CR transmitters at each time slot, i.e., set $P_t^{(k)} = P^{(k)}$, $t = 0, \ldots, T-1$, for CR link k. We call this heuristic decomposition algorithm as the FCA-FPB algorithm. In particular, we fix the channel allocation over the two CRs by setting $D_t^{(1)} = \{1, \cdots, N/2\}$ and $D_t^{(2)} = \{N/2+1, \cdots, N\}$. The original problem (4.13) can be decomposed into $2T$ subproblems in the form of (4.19) and thus can be solved by (4.20).

2. VCA-FPB Algorithm

In this algorithm, we fix the power budget for each CR by setting $P_t^{(k)} = P^{(k)}$, $t = 0, \ldots, T-1$, for CR link k. However, we allow variable channel allocation over the two CR links at each time slot. We call this heuristic decomposition approach

as the VCA-FPB algorithm. The original problem (4.13) can be decomposed into T subproblems in the form of (4.15), which can be solved efficiently by the proposed algorithm in Section III. B.

3. FCA-VPB Algorithm

In this algorithm, we fix the channel allocation for the CRs, each of them taking $N/2$ channels, but allow variable transmit power budget at each time slot. We call this heuristic decomposition algorithm as the FCA-VPB algorithm. In particular, we fix the channel allocation over the two CRs by setting $D_t^{(1)} = \{1, \cdots, N/2\}$ and $D_t^{(2)} = \{N/2+1, \cdots, N\}$. The original problem (4.13) can be decomposed into two subproblems for the two CR links. For CR link k, the subproblem can be expressed as

$$\begin{aligned} \max \quad & E\left\{\frac{1}{T}\sum_{t=0}^{T-1}\sum_{n\in D_t^{(k)}} c_{t,n}^{(k)}\right\} \\ \text{s. t.} \quad & \frac{1}{T}\sum_{t=0}^{T-1}\sum_{n\in D_t^{(k)}} P_{t,n}(k) \le P^{(k)}, \\ & P_{t,n}^{(k)} \le P^H, \quad n \in D_{1,t}^{(k)} \\ & P_{t,n}^{(k)} \le P^L, \quad n \in D_{2,t}^{(k)}. \end{aligned} \tag{4.26}$$

This problem is solved with a similar DP-based method to that for the VCA-VPB algorithm.

E. Numerical Results

In this section, with the two-CR case, we evaluate the achievable average sum-rate of the proposed VCA-VPB algorithm against three heuristic algorithms defined in Section IV. D. In the simulation, without loss of generality, we set $W = 1$. Other

system parameters are given in Table I. In addition, we set $P^{(1)} = P^{(2)} = 10$ mW and discretize the total transmit power over time at a resolution of 5 mW.

The performance comparison over different values of T in terms of the average sum-rate is shown in Fig. 15. In the simulation, we set $\alpha = 0.01$, $\beta = 0.1$, and $p_{\text{th}} = 0.15$. In this case, we have $\rho = 1$, which implies that all the previously idle channels are used for the CR links to transmit with power mask P^H. In Fig. 15, the theoretical result of the VCA-VPB algorithm is obtained according to (4.25) by solving the DP recursion given in (4.23). The simulation result of the VCA-VPB algorithm is obtained as follows: A realization of the sequence $\mathbf{I}_t$, $t = -1, 0, \ldots, T-1$, is first generated according to the underlying Markov chain; two realizations of the sequences $H_t^{(1)}$ and $H_t^{(2)}$ are then generated according to their distributions, respectively; for each time t, given $B_t^{(1)}$, $B_t^{(2)}$, $\mathbf{I}_{k-1}$, $H_t^{(1)}$, and $H_t^{(2)}$, $P_{t,n}^{(k)}$ is obtained from a pre-stored look-up table generated according to the quantized result of the VCA-VPB algorithm; knowing the transmit power $P_{t,n}^{(k)}$, the corresponding data rate $c_{t,n}^{(k)}$ is determined according to (4.14); finally, the average sum-rate is evaluated by averaging the sum of $c_{t,n}^{(k)}$s over all T time slots during the simulation run. In the simulation, 1000 such runs are performed to average out the impact of a particular realization of the sequences $\mathbf{I}_t$, $H_t^{(1)}$, and $H_t^{(2)}$. The theoretical result of the FCA-VPB is obtained in a similar way to that for the VCA-VPB algorithm. We see that the theoretical and simulation results of the VCA-VPB algorithm are almost the same, both of which are roughly 15% better than that of the reference schemes. The performance gain of the VCA-VPB algorithm against the FCA-VPB algorithm and the VCA-FPB algorithm is due to the fact that we explore both the multi-user diversity and the flexibility of power allocation over time via the DP approach. The FCA-FPB algorithm performs the worst among these four algorithms as expected.

In Fig. 16, we plot the performance of the VCA-VPB algorithm against that

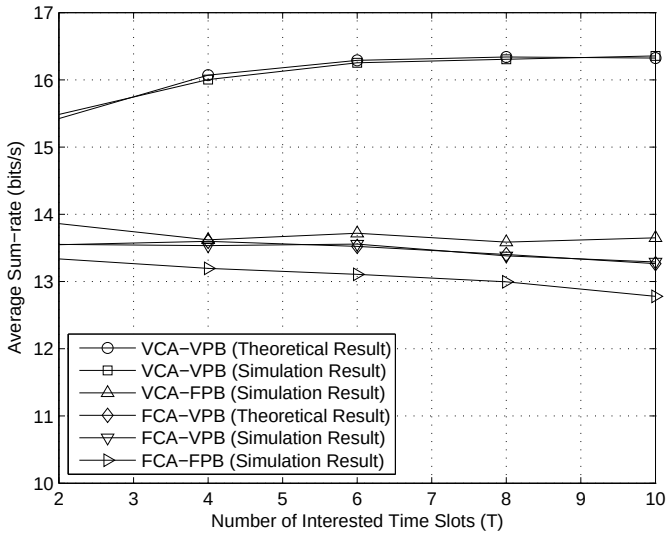

Fig. 15. The average sum-rate of the VCA-VPB algorithm ($\alpha = 0.01$ and $\beta = 0.1$).

of the reference schemes over different values of T with $\alpha = 0.01$, $\beta = 0.2$, and $p_{\text{th}} = 0.15$, which leads to $\rho = 0.75$ in this case. We see that the theoretical and simulation results of the VCA-VPB algorithm are still close to each other, which are much better than those of the reference schemes. However, the average rate obtained by the VCA-VPB algorithm is less than that in Fig. 15. This is due to the fact that we now can only use 75% of the previously idle channels with power mask P^H at each time slot and the average number of idle channels at each time slot decreases due to the fast variation of PR behavior (with large β value).

In Fig. 17, we give a specific realization of the PR behavior in terms of the number of previously idle channels M_{t-1} and channel gains $H_t^{(1)}$ and $H_t^{(2)}$ for the two CR links. We see that M_{t-1} decreases over t based on the underlying DTMC channel model with this specific realization. Accordingly, in Fig. 18, we plot the individual power budget $P_t^{(k)}$ and the sum-power budget $P_t^{(1)} + P_t^{(2)}$ obtained by the VCA-VPB algorithm. We see that the power assigned to the two CR links varies over different

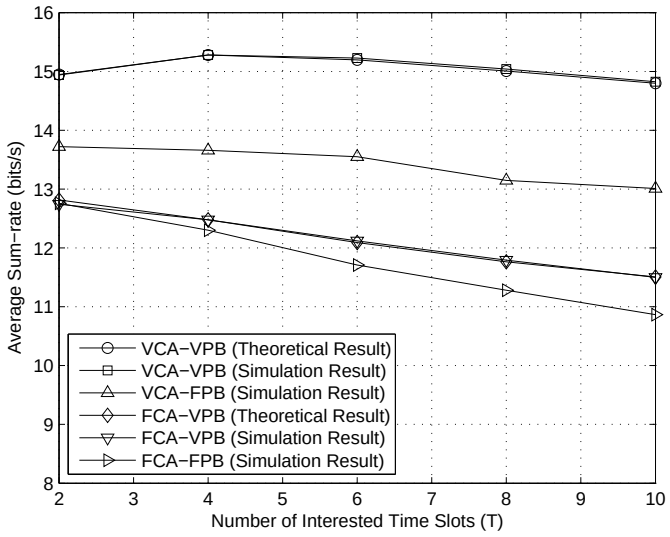

Fig. 16. The average sum-rate of the VCA-VPB algorithm ($\alpha = 0.01$ and $\beta = 0.2$).

time slots. Roughly speaking, the overall power allocation decreases over time. Some data points (e.g., $t = 7$) may violate this tendency due to the variation of the channel gains over different time slots.

In Fig. 19, to illustrate the effect of delayed spectrum sensing result we plot the performance of the VCA-VPB algorithm with delayed spectrum sensing against that of the modified VCA-VPB algorithm with ideal spectrum sensing (without spectrum sensing delay) over different values of T. The modified VCA-VPB algorithm with ideal spectrum sensing is described as follows. We replace the delayed spectrum sensing information with the instantaneous one in (23) and use a similar DP-based approach as the one in the VCA-VPB algorithm to maximize the average sum-rate of the two CR links; the only difference from the VCA-VPB algorithm is that we set the power mask for each channel according to (10) based on the instantaneous channel occupancy information such that there are no channel outages for the PRs. We see that the performance of the VCA-VPB algorithm is the same as that of the

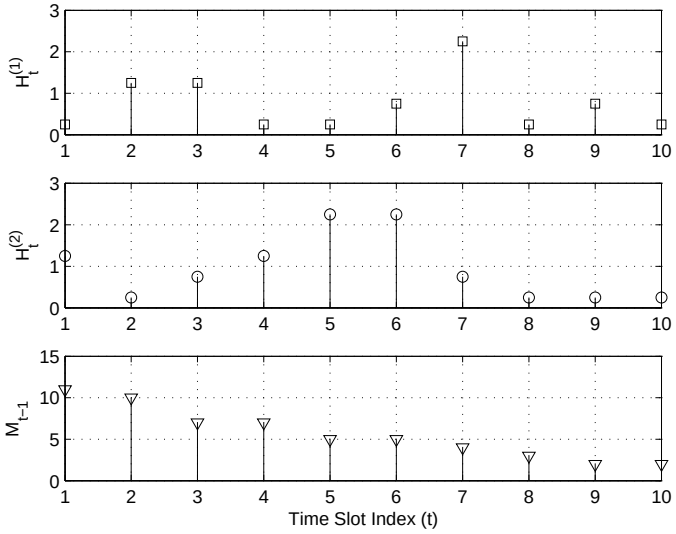

Fig. 17. A realization of the PR behavior and channel gains ($\alpha = 0.01$ and $\beta = 0.2$).

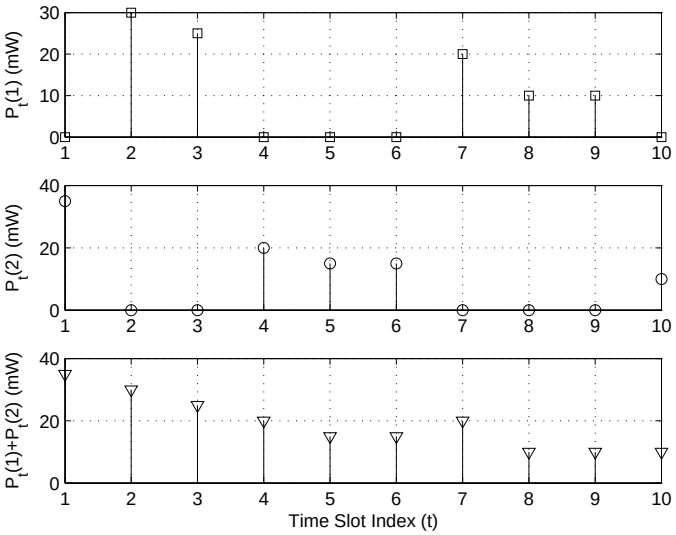

Fig. 18. The power allocation of the VCA-VPB algorithm ($\alpha = 0.01$ and $\beta = 0.2$).

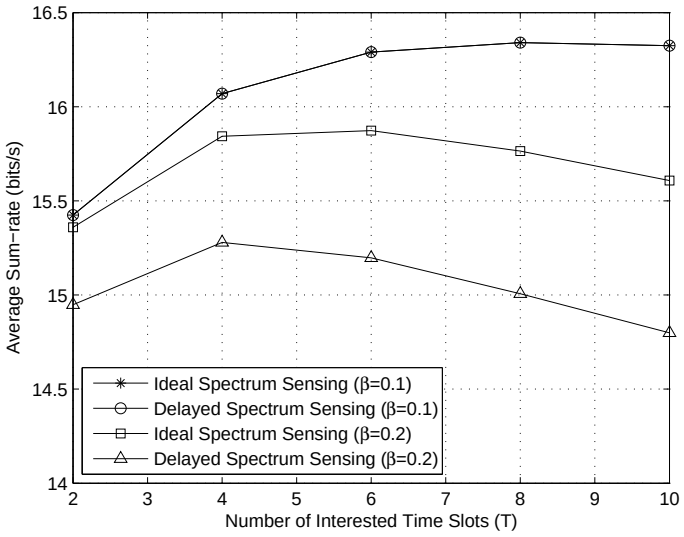

Fig. 19. The impact of delayed spectrum sensing ($\alpha = 0.01$).

modified VCA-VPB algorithm when $\beta = 0.1$. This is due to the fact that $\beta < p_{\text{th}}$ in this case such that we can use all previously idle channels with power mask P^H at each time slot, i.e., all the channels are fully utilized as in the modified VCA-VPB algorithm. However, the VCA-VPB algorithm could still cause 10% PR outage while the modified VCA-VPB does not. When $\beta = 0.2$, there is a performance gap between the case with delayed spectrum sensing and the case with ideal spectrum sensing, which is caused by the different utilization of the previously idle channels.

F. Summary

In this chapter, we proposed a power and rate control scheme based on DP approaches for multiple CR links operating over multiple channels. The proposed strategy explores the multi-user diversity and the time correlation of the PR behavior. Analytical and simulation results show that the proposed algorithm leads to a significant performance improvement against heuristic approaches in terms of the average sum-rate

over a finite time horizon.

CHAPTER V

CONCLUSIONS

In this dissertation, the fundamental limits of large CONs are first investigated in terms of the asymptotic throughput and packet delay performance when the numbers of nodes approach infinity. The following two types of CONs are considered: 1) selfish CONs, in which neither the primary tier nor the secondary tier is willing to route the packets for the other, and 2) supportive CONs, in which the secondary tier is willing to route the packets for the primary tier while the primary tier does not. In addition, we study the throughput and delay of a CON with a small number of nodes. In this chapter, the dissertation contributions are first summarized. Several research directions for future work are then discussed.

A. Summary of Dissertation Contributions

In Chapter II, the throughput and delay scaling laws for selfish CONs are investigated, which is summarized as follows

- For the primary tier, it is shown that the throughput per S-D pair is $\lambda_p(n) = \Theta(\sqrt{\frac{1}{n \log n}})$ *w.h.p.* and the sum throughput is $T_p(n) = \Theta(\sqrt{\frac{n}{\log n}})$ *w.h.p.*. These results are the same as those in a stand-alone ad hoc wireless network considered in [14]. Following the fluid model [21], we give the delay-throughput tradeoff for the primary tier as $D_p(n) = \Theta(n\lambda_p(n))$ for $\lambda_p(n) = O(\frac{1}{\sqrt{n \log n}})$, which is the optimal delay-throughput tradeoff for a stand-alone wireless ad hoc network established in [21].

- For the secondary tier, it is shown that the throughput per S-D pair is $\lambda_s(m) = \Theta(\sqrt{\frac{1}{m \log m}})$ *w.h.p.* and the sum throughput is $T_s(m) = \Theta(\sqrt{\frac{m}{\log m}})$ *w.h.p.*.

Although due to the presence of the preservation regions, the secondary packets seemingly experience larger delays compared with that of the primary tier, we show that the delay-throughput tradeoff for the secondary tier is the same as that in the primary tier, i.e., $D_s(m) = \Theta(m\lambda_s(m))$ for $\lambda_s(m) = O(\frac{1}{\sqrt{m \log m}})$.

In Chapter III, the throughput and delay scaling laws for supportive CONs are investigated, which is summarized as follows

Case I: The primary and secondary nodes are all static.

- It is shown that the primary tier can achieve a per-node throughput scaling of $\lambda_p(n) = \Theta\left(1/\log n\right)$ and a delay scaling of $D_p(n) = \Theta\left(\sqrt{n^\beta \log n}\lambda_p(n)\right)$ for $\lambda_p(n) = O\left(1/\log n\right)$.
- It is shown that the secondary tier can achieve a per-node throughput scaling of $\lambda_s(m) = \Theta\left(\frac{1}{\sqrt{m \log m}}\right)$ and a delay scaling of $D_s(m) = \Theta(m\lambda_s(m))$ for $\lambda_s(m) = O\left(\frac{1}{\sqrt{m \log m}}\right)$.

Case II: The primary nodes are static and the secondary nodes are mobile.

- It is shown that the primary tier can achieve a per-node throughput scaling of $\lambda_p(n) = \Theta\left(1/\log n\right)$, and delay scaling laws of $\Theta(1)$ and $\Theta(1/S)$ with the i.i.d. mobility model and the RW mobility model, respectively.
- It is shown that the secondary tier can achieve a per-node throughput scaling of $\lambda_p(n) = \Theta(1)$, and delay scaling laws of $\Theta(m)$ and $\Theta\left(m^2 S \log \frac{1}{S}\right)$ with the i.i.d. mobility model and the RW mobility model, respectively.

We see that if the secondary tier is static, the throughput scaling law of the primary tier could be improved by sacrificing the delay performance; if the secondary tier is mobile, both the throughput and delay scaling laws of the primary tier could be

improved with the aid of the secondary tier. Furthermore, the secondary tier can achieve the same throughput and delay scaling laws as a stand-alone network at the same time.

In Chapter IV, the throughput and delay performance of a CON with a small number of nodes is investigated. We proposed a power and rate control scheme based on dynamic programming for multiple CR links operating over multiple channels. The proposed strategy explores the multi-user diversity and the time correlation of the PR behavior. Analytical and simulation results show that the proposed algorithm leads to a significant performance improvement against heuristic approaches in terms of the average sum-rate over a finite time horizon.

B. Future Work

We discuss the following possible extensions to the work presented in this dissertation.

The throughput and delay scaling laws of supportive CONs in Chapter III are based on the assumption that $\gamma \geq 2$, which means that the density of the secondary tier is much denser than that of the primary tier. When $1 < \gamma < 2$, the proposed protocols need to be modified. In our future work, we will investigate the throughput and delay scaling laws of CONs in this regime. In addition, we assume that in the mobile case, the random walk size is larger than the primary cell size. In our future work, we will extend our results to the situation where the primary cell size is larger than the random walk size.

In this dissertation, we discussed the case where both the primary and secondary tiers are static for selfish CONs. For supportive CONs, we discuss the cases where the primary tier is static and the secondary tier is either static or mobile. In our future work, we will consider the case where both the primary and secondary tiers

are mobile. Other possible extensions are following: Given the constraint over the primary throughput constant-factor degradation, how to maximize the throughput of the secondary network; Considering more complicated delay models, such as the constant-size-packet model, explore the delay-throughput tradeoff for the secondary network.

For the CON with a small number of nodes, in this dissertation we made the problem tractable by assuming a FDMA-structured solution and flat-fading channels. In our future work, we will try to address this problem under more general setups.

REFERENCES

[1] Federal Communications Commission, Spectrum Policy Task Force, Rep. ET Docket no. 02-135, Nov. 2002.

[2] G. Bansal, M. J. Hossain, and V. K. Bhargava, "Optimal and suboptimal power allocation schemes for OFDM-based cognitive radio systems" *IEEE Transactions on Wireless Communications*, vol. 7, no. 11, pp. 4710-4718, Nov. 2008.

[3] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2053-2071, May 2008.

[4] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE Journal on Selected Arears in Communications*, vol. 23, no. 2, pp. 201-220, Feb. 2005.

[5] M. H. Islam, Y. Liang, and A. T. Hoang, "Joint power control and beamforming for cognitive radio networks", *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, Jul. 2008.

[6] Z. Quan, S. Cui, and A. H. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 28-40, Feb. 2008.

[7] F. Wang, M. Krunz, and S. Cui, "Price-based spectrum management in cognitive radio networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 74-87, Feb. 2008.

[8] H. B. Salameh, M. Krunz, and O. Younis, "MAC protocol for opportunistic cognitive radio networks with soft guarantees," *IEEE Transaction on Mobile Computing*, accepted.

[9] R. Zhang and Y. Liang, "Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks," *IEEE Journal on Selected Arears in Signal Processing*, vol. 2, no. 1, pp. 88-102, Feb. 2008.

[10] L. Zhang, Y. Liang, and Y. Xin, "Joint beamforming and power allocation for multiple access channels in cognitive radio networks," *IEEE Journal on Selected Arears in Communications*, vol. 26, no. 1, pp. 38-51, Jan. 2008.

[11] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in Ad Hoc network: a POMDP framework," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 589-600, Apr. 2007.

[12] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 1813-1827, May 2006.

[13] T. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572-584, Sept. 1979.

[14] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388-404, Mar. 2000.

[15] N. Bansal and Z. Liu, "Capacity, delay and mobility in wireless ad-hoc networks," *Proceedings of Infocom*, vol. 2, pp. 1553-1563, Mar. 2003.

[16] S. N. Diggavi, M. Grosslauser, and D. Tse, "Even one-dimensional mobility increases ad hoc wireless capacity," *IEEE Transactions on Information Theory*, vol. 51, no. 11, pp. 3947-3954, Nov. 2005.

[17] M. Franceschetti, O. Dousse, D. Tse, and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory," *IEEE Transactions on Information Theory*, vol. 53, no. 3, pp. 1009-1018, Mar. 2007.

[18] M. Grossglauser and D. Tse, "Mobilty increases the capacity of ad-hoc wireless networks," *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, pp. 477-486, Aug. 2002.

[19] A. Jovicic, P. Viswanath, and S. R. Kulkarni, "Upper bounds to transport capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 50, no. 11, pp. 2555-2565, Nov. 2004.

[20] S. R. Kullkarni and P. Viswanath, "A deterministic approach to throughput scaling in wireless networks," *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1041-1049, Jun. 2004.

[21] A. El Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Optimal throughput-delay scaling in wireless networks—Part I: The fluid model," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2568-2592, Jun. 2006.

[22] A. El Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Optimal throughput-delay scaling in wireless networks—Part II: Constant-size packets," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5111-5116, Jun. 2006.

[23] X. Lin and N. B. Shroff, "The fundamental capacity-delay tradeoff in large mobile ad hoc networks," in *Third Annual Mediterranean Ad Hoc Networking Workshop*,

Bodrum, Turkey, Jun. 2004.

[24] X. Lin, G. Sharma, R. Mazumdar, and N. B. Shroff, "Degenerate delay-capacity trade-offs in ad hoc networks with Brownian mobility," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2777-2784, Jun. 2006.

[25] M. J. Neely and E. Modiano, "Capacity and delay tradeoffs for ad hoc mobile networks," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1917-1937, Jun. 2005.

[26] A. Ozgur and O. Lvque, "Throughput-delay trade-off for hierarchical cooperation in ad hoc wireless networks," Preprint. Jan. 2008. [Online]. Available: http://arxiv.org/pdf/0802.2013v1.

[27] G. Sharma, R. Mazumdar, and N. B. Shroff, "Delay and capacity trade-offs in mobile ad hoc networks: A global perspective," *IEEE/ACM Transactions on Networking*, vol. 15, no. 5, pp. 981-992, Oct. 2007.

[28] S. Toumpis and A. J. Goldsmith, "Large wireless networks under fading, mobility, and delay constraints," in *Proceedings of Infocom*, Hong Kong, Mar. 2004.

[29] O. Lvque and İ. E. Telatar, "Information-theoretic upper bounds on the capacity of large extended ad hoc wireless networks," *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 858-865, Mar. 2005.

[30] A. Ozgur, O. Lvque, and D. N. C. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3549-3572, Oct. 2007.

[31] L. Xie and P. R. Kumar, "A network information theory for wireless communication: Scaling laws and optimal operation," *IEEE Transactions on Information*

Theory, vol. 50, no. 5, pp. 748-767, Feb. 2004.

[32] L. Xie and P. R. Kumar, "On the path-loss attenuation regime for positive cost and linear scaling of transport capacity in wireless networks," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2318-2328, Jun. 2006.

[33] F. Xue and P. R. Kumar, *Scaling Laws for Ad Hoc Wireless Networks: An Information Theoretic Approach*. Now Publishers, Delft, The Netherlands, 2006.

[34] M. Vu, N. Devroye, M. Sharif, and V. Tarokh, "Scaling laws of cognitive networks," in *Proceedings of CrownCom*, pp. 2-8, Aug. 2007.

[35] M. Vu and V. Tarokh, "Scaling laws of single-hop cognitive networks," *IEEE Transactons on Wireless Communnications*, vol. 8, no. 8, pp. 4089-4097, Aug. 2009.

[36] S. -W. Jeon, N. Devroye, M. Vu, S. -Y. Chung, and V. Tarokh, "Cognitive networks achieve throughput scaling of a homogeneous network," Preprint. Jul. 2009. [Online]. Available: http://arxiv.org/pdf/0801.0938v2.

[37] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*, Cambridge, U.K.: Cambridge Univ. Press, 2005.

[38] F. Xue and P. R. Kumar, "The transport capacity of wireless networks under fading channels," *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 834-847, Mar. 2005.

[39] L. Ying, S. Yang, and R. Srikant, "Optimal delay-throughput trade-offs in mobile ad hoc networks," *IEEE Transactions on Information Theory*, vol. 54, no. 9, Sept. 2008.

[40] D. Aldous and J. Fill, "Reversible Markov chain and random walks on graph," [online]. Available: http://www.stat.berkeley.edu/users/aldous/RWG/book.html.

[41] H. Daduna, *Queueing Networks with Discrete Time Scale*, New York, Springer, 2001.

[42] Y. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, no.4, pp. 13261337, Apr. 2008.

[43] C. R. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. J. Shellhammer, and W. Caldwell, "IEEE 802.22: the first cognitive radio wiereless regional area network standard," *IEEE Communications Magazine*, vol. 47, no. 1, pp. 130-138, Jan. 2009.

[44] A. Fu, E. Modiano, and J. N. Tsitsiklis, "Optimal transmission scheduling over a fading channel with energy and deadline constraints," *IEEE Transactions on Wireless Communications*, vol. 5, no. 3, pp. 630-640, Mar. 2006.

[45] A. Fu, E. Modiano, and J. N. Tsitsiklis, "Optimal energy allocation and admission control for communications satellites," *IEEE/ACM Transactions on Networking*, vol. 11, no. 3, pp. 488-500, Jun. 2003.

[46] I. Bettesh and S. Shamai, "Optimal power and rate control for minimal average delay: the single-user case," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4115-4141, Sept. 2006.

[47] F. Zhang and S. T. Chanson, "Improving communication energy efficiency in wireless networks powered by renewable energy sources," *IEEE Transactions on*

Vehicular Technology, vol. 54, no. 6, pp. 2125-2136, Nov. 2005.

[48] S. Boyd and L. Vandenberghe, *Convex Optimization*, U.K., Cambridge University Press, 2005.

[49] R. Cendrillon, W. Yu, M. Moonen, J. Verlinden, and T. Bostoen, "Optimal multiuser spectrum balancing for digital subscriber line," *IEEE Transactions on Communications*, vol. 54, no. 5, pp. 922-933, May 2006.

[50] A. Goldsmith and P. Varaiya, "Capacity of fading channel with channel side information," *IEEE Transactions on Information Theory*, vol. 43, no. 11, pp. 1986-1992, Nov. 1997.

APPENDIX A

PROOF OF LEMMA 11

Proof. Let n_p denote the number of the primary nodes in a particular primary cell, which is a Poisson random variable with parameter $\mu = na_p$. By the Chernoff bound, the probability that a particular primary cell has no more than $\varepsilon_1\mu$ primary nodes is given by

$$\begin{aligned} P(n_p \le \varepsilon_1\mu) &\le \frac{e^{-\mu}(e\mu)^{\varepsilon_1\mu}}{(\varepsilon_1\mu)^{\varepsilon_1\mu}} \\ &= \frac{e^{-na_p}(ena_p)^{\varepsilon_1 na_p}}{(\varepsilon_1 na_p)^{\varepsilon_1 na_p}} \\ &= e^{-na_p(1-\varepsilon_1(1-\log\varepsilon_1))} \\ &= e^{-na_p(1-\frac{(1+\log\lambda)}{\lambda})} \\ &\le e^{-na_p} \end{aligned} \tag{A.1}$$

where $0 < \varepsilon_1 < 1$, $\lambda = 1/\varepsilon_1 > 1$, and we use the fact that $\log\lambda \le \lambda - 1$. Let A denote the event that at least one primary cell has no more than $\varepsilon_1 na_p$ primary nodes. By the union bound, we have

$$P(A) \le \frac{1}{a_p} e^{-na_p} \to 0 \tag{A.2}$$

as $n \to \infty$. Therefore, each primary cell has more than $\varepsilon_1 na_p$ primary nodes w.h.p.. Furthermore, given $\varepsilon_2 > e$, we have

$$\begin{aligned} P(n_p \ge \varepsilon_2\mu) &\le \frac{e^{-\mu}(e\mu)^{\varepsilon_2\mu}}{(\varepsilon_2\mu)^{\varepsilon_2\mu}} \\ &= \frac{e^{-na_p}(ena_p)^{\varepsilon_2 na_p}}{(\varepsilon_2 na_p)^{\varepsilon_2 na_p}} \\ &= e^{-na_p}\left(\frac{e}{\varepsilon_2}\right)^{\varepsilon_2 na_p} \end{aligned} \tag{A.3}$$

Let B denote the event that at least one primary cell has no less than $\varepsilon_2 n a_p$ primary nodes. By the union bound, we have

$$P(B) \leq \frac{1}{a_p} e^{-na_p} \left(\frac{e}{\varepsilon_2} \right)^{\varepsilon_2 n a_p} \to 0 \tag{A.4}$$

as $n \to \infty$. Thus, each primary cell has less than $\varepsilon_2 n a_p$ primary nodes w.h.p.. Combining (A.2) and (A.4) completes the proof for the case of primary nodes. The proof for the case of secondary nodes follows a similar way with n replaced by m. $\square$

APPENDIX B

PROOF OF LEMMA 13 AND LEMMA 14

Proof of Lemma 13. Assume that at a given moment, there are K_p active primary cells. The rate of the ith active primary cell is given by

$$R_p(i) = \frac{1}{64} \log \left(1 + \frac{P_p(i) g \left(||X_{p,tx} - X_{p,rx}|| \right)}{N_0 + I_p(i) + I_{sp}(i)} \right) \tag{B.1}$$

where $\frac{1}{64}$ denotes the rate loss due to the 64-TDMA transmission of primary cells. In the surrounding of the ith primary cell, there are 8 primary interferers with a distance of at least $6\sqrt{a_p}$ and 16 primary interferers with a distance of at least $13\sqrt{a_p}$, and so on. As such, the $I_p(i)$ is upper-bounded by

$$\begin{aligned} I_p(i) &= \sum_{k=1, k\neq i}^{K_p} P_p g(||X_{p,tx}(k) - X_{p,rx}(i)||) \\ &< P \sum_{t=1}^{\infty} 8t(7t-1)^{-\kappa} \triangleq A. \end{aligned} \tag{B.2}$$

Next, we discuss the interference $I_{sp}(i)$ from secondary transmitting interferers to the ith primary RX. We consider the following two case:

Case I : The secondary tier transmits either the primary packets to the next secondary relay nodes or transmits the secondary packets to the next hop, i.e., in the first or secondary subframes.

Case II : The secondary tier delivers the data packets to the primary destination nodes, i.e., in the third secondary subframe.

In Case I, assume that there are K_s active secondary cells, which means that the number of the active secondary TXs is also K_s. Since a minimum distance $\sqrt{a_s}$ can

be guaranteed from all secondary transmitting interferers to the primary RXs in the preservation regions, $I_{sp}(i)$ is upper-bounded by

$$\begin{aligned} I_{sp}^{I}(i) &= \sum_{k=1,k\neq i}^{K_s} P_s g(||X_{s,tx}(k) - X_{p,rx}(i)||) \\ &< P\sum_{t=1}^{\infty} 8t(7t-6)^{-\kappa} \triangleq B. \end{aligned} \tag{B.3}$$

In Case II, there are K_p collection regions and thus K_p active secondary TXs. In the surrounding of the ith primary cell, there are 2 secondary interferers with a distance of at least $2\sqrt{a_p}$ and 4 secondary interferers with a distance of at least $9\sqrt{a_p}$, and so on. Then, $I_{sp}(i)$ is upper-bounded by

$$\begin{aligned} I_{sp}^{II}(i) &= \sum_{k=1,k\neq i}^{K_p} P_p g(||X_{s,tx}(k) - X_{p,rx}(i)||) \\ &< P\sum_{t=1}^{\infty} 2t(7t-5)^{-\kappa} \triangleq C. \end{aligned} \tag{B.4}$$

Given $B > A$ and $B > C$, we have

$$R_p(i) > \frac{1}{64}\log\left(1 + \frac{P(\sqrt{5})^{-\kappa}}{N_0 + 2P\sum_{t=1}^{\infty} 8t(7t-6)^{-\kappa}}\right). \tag{B.5}$$

Since $\sum_{t=1}^{\infty} 8t(7t-6)^{-\kappa}$ converges to a constant for $\kappa > 2$, there exists a constant $K_1 > 0$ such that $R_p(i) > K_1$. This completes the proof. □

Proof of Lemma 14. The proof is similar to that for *Lemma* 3. When a primary RX receives packets from its surrounding secondary nodes, it suffers from two interference terms from all active primary TXs and all active secondary TXs, either of which can be upper-bounded by a constant independent of n and m. Thus there is a constant rate K_2, at which the secondary tier can deliver packets to the intended primary destination node. □

APPENDIX C

PROOF OF LEMMA 15 AND LEMMA 16

Proof of Lemma 15. Assume that at a given moment, there are K_p active primary cells. The supported rate of the ith active primary cell is given by

$$R_p(i) = \frac{1}{64} \log \left(1 + \frac{P_p(i) g \left(||X_{p,tx} - X_{p,rx}|| \right)}{N_0 + I_p(i) + I_{sp}(i)} \right) \tag{C.1}$$

where $\frac{1}{64}$ denotes the rate loss due to the 64-TDMA transmission of primary cells. In the surrounding of the ith primary cell, there are 8 primary interferers with a distance of at least $6\sqrt{a_p}$ and 16 primary interferers with a distance of at least $13\sqrt{a_p}$, and so on. As such, the $I_p(i)$ is upper-bounded by

$$\begin{aligned} I_p(i) &= \sum_{k=1, k\neq i}^{K_p} P_p g(||X_{p,tx}(k) - X_{p,rx}(i)||) \\ &< P \sum_{t=1}^{\infty} 8t(7t-1)^{-\kappa} \triangleq A. \end{aligned} \tag{C.2}$$

Next, we discuss the interference $I_{sp}(i)$ from secondary transmitting interferers to the ith primary RX. According to the proposed secondary protocol, the secondary nodes are divided into two classes: Class I and Class II, which operate over the switched timing relationships with the odd and the even time slots. Without the loss of generality, we consider the interference $I_{sp}(i)$ from secondary transmitting interferers to the ith primary RX at the odd time slots. Assume that there are K_s active secondary cells, which means that the number of the active secondary TXs of Class I is K_s. Since a minimum distance $\sqrt{a_s}$ can be guaranteed from all secondary transmitting interferers of Class I to the primary RXs in the preservation regions, the

interference from the active secondary TXs of Class I, $I_{sp}^{I}(i)$, is upper-bounded by

$$\begin{aligned} I_{sp}^{I}(i) &= \sum_{k=1,k\neq i}^{K_s} P_s g(||X_{s,tx}(k) - X_{p,rx}(i)||) \\ &< P\sum_{t=1}^{\infty} 8t(7t-6)^{-\kappa} \triangleq B. \end{aligned} \tag{C.3}$$

Furthermore, there are K_p collection regions, which means that the number of the active secondary TXs of Class II is K_p. Since a minimum distance $2\sqrt{a_p}$ can be guaranteed from all secondary transmitting interferes of Class II to the primary RXs in the preservation regions, the interference from the active secondary TXs of Class II, $I_{sp}^{II}(i)$, is upper-bounded by

$$\begin{aligned} I_{sp}^{II}(i) &= \sum_{k=1,k\neq i}^{K_p} P_p g(||X_{p,tx}(k) - X_{p,rx}(i)||) \\ &< P\sum_{t=1}^{\infty} 8t(7t-5)^{-\kappa} \triangleq C. \end{aligned} \tag{C.4}$$

Given $B > A$ and $B > C$, we have

$$\begin{aligned} R_p(i) &= \frac{1}{64}\log\left(1 + \frac{P_p(i)g\left(||X_{p,tx} - X_{p,rx}||\right)}{N_0 + I_p(i) + I_{sp}^{I}(i) + I_{sp}^{II}(i)}\right) \\ &> \frac{1}{64}\log\left(1 + \frac{P(\sqrt{5})^{-\kappa}}{N_0 + 3P\sum_{t=1}^{\infty} 8t(7t-6)^{-\kappa}}\right). \end{aligned} \tag{C.5}$$

Since $\sum_{t=1}^{\infty} 8t(7t-6)^{-\kappa}$ converges to a constant for $\kappa > 2$, there exists a constant $K_3 > 0$ such that $R_p(i) > K_3$. This completes the proof. □

Proof of Lemma 16. The proof is similar to that for *Lemma* 5. When a primary RX receives packets from its surrounding secondary nodes, it suffers from three interference terms from all active primary TXs, all active secondary TXs of Class I, and all active secondary TXs of Class II, each of which can be upper-bounded by a constant independent of n and m. Thus, there is a constant rate K_4, at which the secondary

tier can deliver packets to the intended primary destination node. □

APPENDIX D

DERIVATION OF (3.10)

We know that given $a_p(n) \geq \sqrt{2}\gamma \log n/n$, the maximum throughput per S-D pair for the primary tier is $\Theta\left(\frac{1}{na_p(n)}\right)$. Since a primary packet is divided into N segments and then routed by N parallel S-D paths within the secondary tier, the supported rate for each secondary S-D pair is required to be $\Theta\left(\frac{1}{Nna_p(n)}\right) = \Theta\left(\frac{\sqrt{\log m}}{\sqrt{m}na_p(n)}\right)$. As such, based on (3.25), the corresponding secondary cell size $a_s(m)$ needs to be set as

$$a_s(m) = \frac{n^2 a_p^2(n)}{m \log m}$$

where we have $a_s(m) \geq 2 \log m/m$ when $a_p(n) \geq \sqrt{2}\gamma \log n/n$.

VITA

Long Gao received his B.S. degree in communication engineering from Beijing Jiaotong University, Beijing, China, in 2003, and his M.S. degree in telecommunication engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2006. He graduated with his Ph.D. in electrical engineering from Texas A&M University in December, 2009. His current research interests include cognitive radios, network information theory, and cross-layer optimization. Long Gao can be contacted at the following address:

Department of Electrical and Computer Engineering
c/o Dr. Shuguang Cui
Texas A&M University
College Station, TX, 77843-3128

The typist for this dissertation was Long Gao.