Throughput and Delay Scaling in Supportive Two-Tier Networks

Long Gao, *Student Member, IEEE,* Rui Zhang, *Member, IEEE,*

Changchuan Yin, *Member, IEEE,* and Shuguang Cui, *Member, IEEE*

Abstract

Consider a wireless network 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 technologies. The primary tier consists of randomly distributed legacy nodes of density n, which have an absolute priority to access the spectrum. The secondary tier consists of randomly distributed cognitive nodes of density $m = n^\beta$ with $\beta > 2$, which can only access the spectrum opportunistically to limit the interference to the primary tier. Based on the assumption that the secondary tier is allowed to route the packets for the primary tier, we investigate the throughput and delay scaling laws of the two tiers in 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 the proposed protocols for the two tiers, we show that the primary tier can achieve a per-node throughput scaling of $\lambda_p(n) = \Theta(1/\log n)$ in the above two scenarios. In the associated delay analysis for the first scenario, we show that the primary tier can achieve a delay scaling of $D_p(n) = \Theta(\sqrt{n/\log n}\,\lambda_p(n))$ with $\lambda_p(n) = O(1/\log n)$. In the second scenario, with two mobility models considered for the secondary nodes: an i.i.d. mobility model and a random walk model, we show that the primary tier can achieve delay scaling laws of $\Theta(1)$ and $\Theta(1/S)$, respectively, where S is the random walk step size. The throughput and delay scaling laws for the secondary tier are also established, which are the same as those for a stand-alone network.

L. Gao, C. Yin, and S. Cui are with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, 77843, USA (Emails: lgao@tamu.edu, ccyin@tamu.edu, and cui@tamu.edu).

R. Zhang is with Institute for Infocomm Research, A*STAR, Singapore (Email: rzhang@i2r-star.edu.sg).

I. INTRODUCTION

The explosive growth of large-scale wireless applications motivates people to study the fundamental limits over wireless networks. Consider a randomly distributed wireless network with density n over a unit area, where the nodes are randomly grouped into one-to-one source-destination (S-D) pairs. Initiated by the seminal work in [1], the throughput scaling laws for such a network have been studied extensively in the literature [2]-[5]. For static networks, it is shown [1] that the traditional multi-hop transmission strategy can achieve a throughput scaling of $\Theta\left(1/\sqrt{n \log n}\right)$ [1] per S-D pair. Such a throughput scaling can be improved when the nodes are able to move. It is shown in [6] [7] that a per-node throughput scaling of $\Theta(1)$ is achievable in mobile networks by exploring two-hop transmission schemes. Unfortunately, the throughput improvement in mobile networks incurs a large packet delay [6] [7], which is another important performance metric in wireless networks. In particular, it is shown in [6] that the constant per-node throughput is achieved at the cost of a delay scaling of $\Theta(n)$. The delay-throughput tradeoffs for static and mobile networks have been investigated in [6]-[9]. Specifically, for static networks, it is shown in [7] that the optimal delay-throughput tradeoff is given by $D(n) = \Theta(n \lambda(n))$ for $\lambda(n) = O\left(1/\sqrt{n \log n}\right)$, where $D(n)$ and $\lambda(n)$ are the delay and throughput per S-D pair, respectively.

The aforementioned literature mainly focuses on the delay and throughput scaling laws for a single network. Recently, the emergence of cognitive radio networks motives people to extend the result from a single network to overlaid networks. Consider a licensed primary network 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. Based on such assumptions, it is shown in [10] [11] that both networks can achieve the same throughput and delay scaling laws as a stand-alone network. However, such results

---

[1]We use the following notations throughout this paper: 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 $n \to \infty$.

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 network still capable of keeping the same throughput and delay scaling laws as the case where no supportive actions are taken between the two networks.

In this paper, we define a *supportive* two-tier network with a primary tier and a secondary tier as follows: The secondary tier is allowed to supportively relay the data packets for the primary tier in an opportunistic way (i.e., the secondary users only utilize empty spectrum holes in between primary transmissions even when they are relaying the primary packets), whereas the primary tier is only required to transport its own data. Let n and $m = n^\beta$ denote the node densities of the primary tier and the secondary tier, respectively. We investigate the throughput and delay scaling laws for such a supportive two-tier network with $\beta > 2$ in 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 primary tier can achieve a per-node throughput scaling of $\lambda_p(n) = \Theta(1/\log n)$ in the above two scenarios with a classic time-slotted multi-hop transmission protocol similar to the one in [1]. In the associated delay analysis for the first scenario, we show that the primary tier can achieve a delay scaling of $D_p(n) = \Theta\left(\sqrt{\frac{n}{\log n}} \lambda_p(n)\right)$ with $\lambda_p(n) = O(1/\log n)$. In the second scenario, with two mobility models considered for the secondary nodes: an i.i.d. mobility model and a random walk model, we show that the primary tier can achieve delay scaling laws of $\Theta(1)$ and $\Theta(1/S)$, respectively, where S is the random walk step size. The throughput and delay scaling laws for the secondary tier are also established, which are the same as those for 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 [1] 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.

The rest of the paper is organized as follows. The system model is described and the main results are summarized in Section II. The proposed protocols for the primary and secondary tiers are described in Section III. The delay and throughput scaling laws for the primary tier are derived in Section IV. The delay and throughput scaling laws for the secondary tier are studied in Section V. Finally, Section VI summarizes our conclusions.

II. System Model and Main Results

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

A. Network Model

The primary nodes are distributed according to a Poisson point process (PPP) of density n and randomly grouped into one-to-one source-destination (S-D) pairs. Likewise, the secondary nodes are distributed according to a PPP of density m and randomly grouped into S-D pairs. We assume that the density of the secondary tier is higher than that of the primary tier, i.e.,

$$m = n \tag{1}$$

where we consider the case with $\quad$ 2. The primary tier and the secondary tier share the same time, frequency, and space, but with different priorities to access the spectrum: The former one is the licensed

user of the spectrum and thus has a higher priority; and the latter one can only opportunistically access the spectrum to limit the resulting interference to the primary tier, even when it helps with relaying the primary packets.

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 channel power gain $g(r)$ is given as

$$g(r) = r^{-\alpha} \tag{2}$$

where r is the distance between the transmitter (TX) and the corresponding receiver (RX), and $\alpha > 2$ denotes the pathloss exponent.

The ambient noise is assumed to be additive white Gaussian noise (AWGN) with an average power N_0. During each time slot, we assume that each TX-RX pair utilizes a capacity-achieving scheme with the data rate of the ith primary TX-RX pair given by

$$R_p(i) = \log\left(1 + \frac{P_p(i)g(\|X_{p,tx}(i) - X_{p,rx}(i)\|)}{N_0 + I_p(i) + I_{sp}(i)}\right) \tag{3}$$

where the channel bandwidth is normalized to be unity for simplicity, $\|\cdot\|$ denotes the norm operation, $P_p(i)$ is the transmit power of the ith primary pair, $X_{p,tx}(i)$ and $X_{p,rx}(i)$ are the TX and RX locations of ith primary pair, respectively, $I_p(i)$ is the sum interference from all other primary TXs, $I_{sp}(i)$ is the sum interference from all the secondary TXs. Likewise, the data rate of the jth secondary TX-RX pair is given by

$$R_s(j) = \log\left(1 + \frac{P_s(j)g(\|X_{s,tx}(j) - X_{s,rx}(j)\|)}{N_0 + I_s(j) + I_{ps}(j)}\right) \tag{4}$$

where $P_s(j)$ is the transmit power of the jth secondary pair, $X_{s,tx}(j)$ and $X_{s,rx}(j)$ are the TX and RX locations of the jth secondary pair, respectively, $I_s(j)$ is the sum interference from all other secondary TXs to the RX of the jth secondary pair, and $I_{ps}(j)$ is the sum interference from all primary TXs.

B. Interaction Model

As shown in the previous work [10] [11], 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. 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 [10] [11].

C. 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 [6]: The secondary nodes are uniformly and randomly

---

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

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 [7] [8]: 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 [7] [8]. 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, which are different from the RW-cells defined above. In this paper, we only consider the case where the size of the RW-cell is greater than or equal to that of the primary cell.

D. Throughput and Delay

The *throughput per S-D pair* (per-node throughput) is defined as the average data rate that each source node can transmit to its chosen destination as in [10] [11], which is asymptotically determined by the network density. Besides, the *sum throughput* is defined as the product between the throughput per S-D pair and the number of S-D pairs in the network. In the following, 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; and we use $T_p(n)$ and $T_s(m)$ to denote the sum throughputs for the primary tier and the secondary tier, respectively.

The delay of a primary packet is defined as the average number of primary time slots that it takes to reach the primary destination node after the departure from the primary source node. Similarly, we define the delay of a secondary packet as the average number of secondary time slots for the packet to travel from the secondary source node to the secondary destination node. We use $D_p(n)$ and $D_s(m)$

to denote packet delays for the primary tier and the secondary tier, respectively. For simplicity, we use a fluid model [7] for the delay analysis, in which we divide each time slot to multiple packet slots and the size of the data packets can be scaled down with the increase of network density.

E. Main Results

We summarize the main results in terms of the throughput and delay scaling laws for the supportive two-tier network here. 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 $\lambda_p(n) = \Theta(1/\log n)$ and a delay scaling of $D_p(n) = \Theta\left(\sqrt{\dfrac{n}{\log n}}\lambda_p(n)\right)$ for $\lambda_p(n) = O(1/\log n)$.

It is shown that the secondary tier can achieve a per-node throughput scaling of $\lambda_s(m) = \Theta\left(\dfrac{1}{\sqrt{m\log m}}\right)$ and a delay scaling of $D_s(m) = \Theta(\sqrt{m}\lambda_s(m))$; for $\lambda_s(m) = O\left(\dfrac{1}{\sqrt{m\log m}}\right)$.

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 $\lambda_p(n) = \Theta(1/\log n)$, 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 $\Theta\left(m^2 S \log\dfrac{1}{S}\right)$ with the i.i.d. mobility model and the RW mobility model, respectively.

III. 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 [10] [11], while the secondary tier chooses its protocol according to the given primary transmission scheme. In the following, 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 $n \to \infty$.

A. The Primary Protocol

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

i) 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 1* for details), we have $a_p(n) \geq \frac{P}{2} \log n = n$ such that each cell has at least one primary node w.h.p..

ii) 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 paper.

iii) 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 [10] [11]. 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.

iv) 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 $P a_p^{\frac{\alpha}{2}}(n)$, where P is a constant.
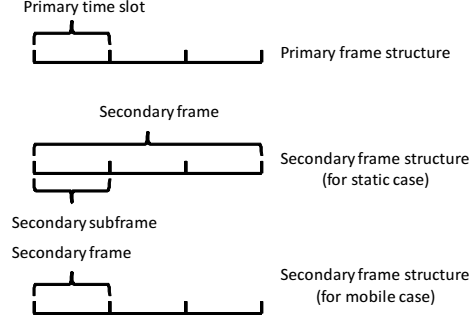
Fig. 1.   Frame relationship between the two tiers.

v) 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 IV.

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 [1]. When the secondary tier is present as shown in Section IV, the primary tier can achieve a better throughput scaling law with the aid of the secondary tier.

B. The 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 2* in Section IV, 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. 1. 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. 2. Only the secondary TXs in an active secondary cell outside all the preservation regions can transmit data

---

[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.

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 $P a_s^2 (m)$.

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

$$N = r \frac{m}{\log m} :$$

(5)

From *Lemma 1* in Section IV, we can guarantee that there are more than N secondary nodes in each primary cell w.h.p. when 2. When $1 < < 2$, 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 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. 2, where the collection region is located between
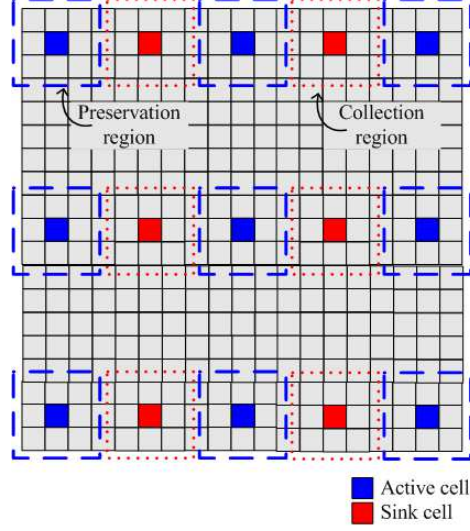
Fig. 2. Preservation regions and collection regions.

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., $P a_p^{\overline{2}}(n)$.

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. 1. 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 [13]

$$\tau = \min\{t : s(t) \le e^{-1}\} \tag{6}$$

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

$$s(t) = \min\{s : p_{(x,y);(u,v)}(t) \ge (1-s)\pi_{(u,v)},$$
$$\text{for all } x,y,u,v \in \{1,2,\ldots,\sqrt{\frac{p}{n_s}}\}\} \tag{7}$$

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 $\delta = \Theta(1=S)$ [13].

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 (6). The packet received at time slot t is considered to be 'type k' and stored in the kth queue, if $\lfloor \frac{t}{64} \rfloor \mod Q = k$, 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 $k = \left\lfloor \frac{t}{64} \right\rfloor \bmod Q$ 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., $P a_p^{\bar{2}}(n)$.

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 [7] to transmit secondary packets with power of $P a_s^{\bar{2}}(m)$ within the secondary tier.

IV. 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.

A. 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 1: 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 I.

Lemma 2: 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 ρ 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 $\rho = \frac{\Theta(ma_p(n))}{\Theta(ma_p(n)) + \Theta(na_p(n))}$ from *Lemma 1*, which approaches one as $\frac{n}{m} \to 1$. This completes the proof. ∎

Lemma 3: With the protocols given in Section III, 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 II.

Lemma 4: With the protocols given in Section III, 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 II.

Based on *Lemmas 1-4*, we have the following theorem.

Theorem 1: With the protocols given in Section III, the primary tier can achieve the following throughput per S-D pair and sum throughput w.h.p. when $\frac{n}{m} \geq 2$:

$$\lambda_p(n) = \Theta\left(\frac{1}{na_p(n)}\right) \tag{8}$$

and

$$T_p(n) = \frac{1}{a_p(n)};$$ (9)

where $a_p(n)\sqrt{\log n / n} = \omega(1)$ and $a_p(n) = o(1)$.

Proof: From *Lemma 3* and *Lemma 4*, 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 1*, 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 2*) by splitting data into $\Theta(\sqrt{m/\log m})$ secondary data paths, each of them at a rate of $\Theta(\frac{1}{m} \cdot \frac{1}{a_s(m)})$. Set $\sqrt{a_s(m)} = \frac{na_p(n)}{m\sqrt{\log m}}$, which satisfies $a_s(m)\sqrt{2 \log m/m} = \omega$. As such, each primary source node achieves a throughput scaling law of $\Theta(1/na_p(n))$. Since the total number of primary nodes in the unit square is of $\Theta(n)$ w.h.p., we have $T_p(n) = \Theta(n\lambda_p(n)) = \Theta(1/a_p(n))$ w.h.p.. This completes the proof. ∎

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

$$\lambda_p(n) = \Theta\left(\frac{1}{\sqrt{\log n}}\right)$$ (10)

and

$$T_p(n) = \Theta\left(\sqrt{\frac{n}{\log n}}\right):$$ (11)

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 2*. 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 (31) 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

$$L_p = 64L_s: \tag{12}$$

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:

$$D_p(n) = \frac{3}{64}D_s(m) + C = \left(\frac{1}{a_s(m)} \right) \tag{13}$$

where the secondary-tier delay $D_s(m)$ is later derived in (31), 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 (13) 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 IV, we derive the relationship between $a_p(n)$ and $a_s(m)$ in our supportive two-tier setup as

$$a_s(m) = \frac{n^2 a_p^2(n)}{m \log m} \tag{14}$$

where we have $a_s(m) = 2 \log m = m$ when $a_p(n) = \frac{p}{2} \log n = n$.

Substituting (14) into (13), we have the following theorem.

Theorem 2: According to the proposed protocols in Section III, the primary tier can achieve the following delay w.h.p. when ≥ 2.

$$D_p(n) = \Theta\left(\frac{m \, \log m}{n a_p(n)}\right) = \Theta\left(\frac{n \, \log n}{n a_p(n)}\right)^! : \tag{15}$$

Delay-Throughput Tradeoff

Combining the results in (8) and (15), the delay-throughput tradeoff for the primary tier is given by the following theorem.

Theorem 3: With the protocols given in Section III, the delay-throughput tradeoff in the primary tier is given by

$$D_p(n) = \Theta\left(n \, \log n \, \lambda_p(n)\right) \quad \text{for} \quad \lambda_p(n) = O\left(\frac{1}{\log n}\right) : \tag{16}$$

B. The Scenario with Mobile Secondary Nodes

Throughput Analysis

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

Lemma 5: With the protocols given in Section III, 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 III.

Lemma 6: With the protocols given in Section III, 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 III.

Based on *Lemmas 1-2* and *Lemmas 5-6*, we have the following theorem.

Theorem 4: With the protocols given in Section III, the primary tier can achieve the following throughput per S-D pair and sum throughput w.h.p.:

$$\lambda_p(n) = \Theta\left(\frac{1}{n a_p(n)}\right) \tag{17}$$

and

$$T_p(n) = \Theta\left(\frac{1}{a_p(n)}\right) ; \tag{18}$$

when $a_p(n) \overset{p}{\ge} 2 \log n = n$ and $a_p(n) = o(1)$.

Proof: From *Lemma 5* and *Lemma 6*, 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 $(na_p(n))$ w.h.p. as shown in *Lemma 1*, the maximum throughput per S-D pair is of $(K = (na_p(n))) = (1 = (na_p(n)))$ w.h.p.. Next, we show that with the proposed protocols, the above maximum throughput scaling is achievable. In the proposed protocols, from *Lemma 2* 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 2*. 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 $(pK = (na_p(n))) = (1 = (na_p(n)))$ and thus a achievable sum throughput of $(1 = a_p)$ for the primary tier w.h.p.. This completes the proof. ∎

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

$$_p(n) = \frac{1}{\log n} \tag{19}$$

and

$$T_p(n) = \frac{n}{\log n} : \tag{20}$$

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 II.C.

1) The i.i.d. Mobility Model: We have the following theorem regarding the delay of the primary tier.

Theorem 5: With the protocols given in Section III, the primary tier can achieve the following delay w.h.p. when $\alpha \geq 2$:

$$D_p(n) = \Theta(1):\tag{21}$$

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. 3 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, 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

$$q = 1 - (1 - a_p(n))^M;\tag{22}$$

$$\approx 1 - e^{-M \cdot a_p(n)};$$

$$\to 1; \text{ as } n \to 1; \text{ for } \alpha \geq 2;$$

where $f \asymp g$ means that f and g have the same limit when $n \to 1$, $M = \Theta(m a_p(n))$ 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 [12]

$$E\{W_1\} = 64 \frac{1-p}{q-p} \to 64; \text{ as } n \to 1; \tag{23}$$

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 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

$$L = n + 1 \tag{24}$$

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 $n \to 1$. Thus, the relay-queues at all secondary nodes are stable over time for each given n, which completes the proof. ∎

2) The RW Mobility Model: For the RW model, we have the following theorem regarding the delay of the primary tier.

Theorem 6: With the protocols given in Section III, the primary tier can achieve the following delay w.h.p. when $\beta \geq 2$:

$$D_p(n) = \Theta\left(\frac{1}{S}\right) = O\left(\frac{1}{a_p(n)}\right) \tag{25}$$
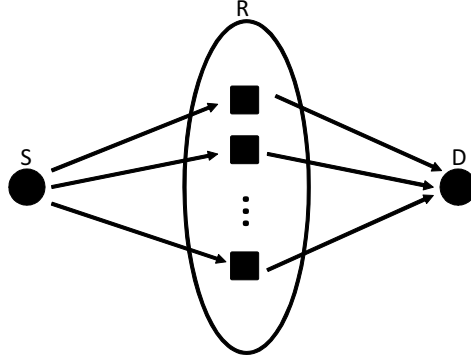
where $S \asymp a_p(n)$.

Fig. 3.   Illustration of the virtual relay node R.

Proof: Like the proof in the i.i.d. mobility case, we fix a primary S-D pair and consider the (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, each secondary node maintains $Q =$ 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 (m) small ones, and the packet that arrives at time t is stored in the kth queue, where $k = \lfloor \frac{t}{64} \rfloor \bmod$. 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) $\lfloor \frac{t}{64} \rfloor \bmod$ $= k$. Consider 64 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) $\lfloor \frac{t}{64} \rfloor \bmod$ $= k$. Let q

denote the probability that event b) occurs during one observation period, which can be expressed as

$$q = 1 - \left(1 - \prod_{i \in I} q_0 P_{(x_i, y_i)(x_d, y_d)}(t_d)\right) ; \qquad (26)$$

$$\geq 1 - (1 - q_0(1 - e^{-1})S)^M ;$$

$$\geq 1 - e^{-q_0(1 - e^{-1})SM} ;$$

$$\to 1; \text{ as } n \to 1 ; \text{ for } \geq 2;$$

where 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 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 $64(\geq 1)$; 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

$$E\{W_2\} = 64 \cdot \frac{1-p}{q-p} \cdot 64 = \Theta\left(\frac{1}{S}\right); \qquad (27)$$

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

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 (24) 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 (8) and (25).

$$D_p(n) = O\left(\frac{n}{\lambda_p(n)}\right); \text{ for } \lambda_p(n) = O\left(\frac{1}{\log n}\right): \qquad (28)$$

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 [7] for a static stand-alone network. Note that the above throughput and delay analysis is based on the assumption $\alpha \geq 2$, and we leave the case with $1 < \alpha < 2$ in our future work.

V. THROUGHPUT AND DELAY ANALYSIS FOR THE SECONDARY TIER

A. 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 [11], which are summarized by the following theorems.

Theorem 7: With the secondary protocol defined in Section III, the secondary tier can achieve the following throughput per S-D pair and sum throughput w.h.p.:

$$\lambda_s(m) = \Theta\left(\frac{1}{m} \cdot \frac{1}{\sqrt{a_s(m)}}\right) \tag{29}$$

and

$$T_s(m) = \Theta\left(\sqrt{\frac{1}{a_s(m)}}\right); \tag{30}$$

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 8: With the secondary protocol defined in Section III, the packet delay is given by

$$D_s(m) = \Theta\left(\sqrt{\frac{1}{a_s(m)}}\right): \tag{31}$$

Delay-Throughput Tradeoff

Combining the results in (29) and (31), the delay-throughput tradeoff for the secondary tier is given by the following theorem.

Theorem 9: With the secondary protocol defined in Section III, the delay-throughput tradeoff is

$$D_s(m) = \Omega(m \lambda_s(m)); \text{ for } \lambda_s(m) = O\left(\frac{1}{m \log m}\right): \tag{32}$$

For detailed proofs of the above theorems, please refer to [11].

B. 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 5* 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 [6] [7] 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 10: With the protocols given in Section III, the secondary tier can achieve the following throughput per S-D pair and sum throughput w.h.p.:

$$\lambda_s(m) = \Theta(1) \tag{33}$$

and

$$T_s(m) = \Theta(m): \tag{34}$$

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

$$D_s(m) = \Theta(m): \tag{35}$$

Theorem 12: With the protocols given in Section III, the secondary tier can achieve the following delay w.h.p. based on the RW model:

$$D_s(m) = \Theta\left(m^2 S \log\frac{1}{S}\right): \tag{36}$$

Note that (36) is a generalized result for $S \leq 1=m$. When $S = 1=m$, the delay $D_s(m) = \Theta(m \log m)$ is the same as that in [7].

VI. CONCLUSION

In this paper, we studied the throughput and delay scaling laws for a supportive two-tier network, 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 [1] 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.

APPENDIX I

Proof of *Lemma 1*

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 ϵ_1 primary nodes is given by

$$
\begin{aligned}
P(n_p \leq \epsilon_1 \lambda) &\leq \frac{e^{-\lambda}(e\lambda)^{\epsilon_1 \lambda}}{(\epsilon_1 \lambda)^{\epsilon_1 \lambda}} \\
&= \frac{e^{-na_p}(ena_p)^{\epsilon_1 na_p}}{(\epsilon_1 na_p)^{\epsilon_1 na_p}} \\
&= e^{na_p(1-\epsilon_1(1-\log \epsilon_1))} \\
&= e^{na_p(1-\frac{(1+\log \delta)}{\delta})} \\
&\leq e^{-na_p}
\end{aligned}
\tag{37}
$$

where $0 < \epsilon_1 < 1$, $\delta = 1/\epsilon_1 > 1$, and we use the fact that $\log \delta \leq \delta - 1$. Let A denote the event that at least one primary cell has no more than $\epsilon_1 na_p$ primary nodes. By the union bound, we have

$$
P(A) \leq \frac{1}{a_p} e^{-na_p} \to 0
\tag{38}
$$

as $n \to \infty$. Therefore, each primary cell has more than $\epsilon_1 na_p$ primary nodes w.h.p.. Furthermore, given $\epsilon_2 > e$, we have

$$
\begin{aligned}
P(n_p \geq \epsilon_2 \lambda) &\leq \frac{e^{-\lambda}(e\lambda)^{\epsilon_2 \lambda}}{(\epsilon_2 \lambda)^{\epsilon_2 \lambda}} \\
&= \frac{e^{-na_p}(ena_p)^{\epsilon_2 na_p}}{(\epsilon_2 na_p)^{\epsilon_2 na_p}} \\
&= e^{-na_p}\left(\frac{e}{\epsilon_2}\right)^{\epsilon_2 na_p}
\end{aligned}
\tag{39}
$$

Let B denote the event that at least one primary cell has no less than $\epsilon_2 na_p$ primary nodes. By the union bound, we have

$$
P(B) \leq \frac{1}{a_p} e^{-na_p}\left(\frac{e}{\epsilon_2}\right)^{\epsilon_2 na_p} \to 0
\tag{40}
$$

as $n \to \infty$. Thus, each primary cell has less than $\epsilon_2 na_p$ primary nodes w.h.p.. Combining (38) and (40) 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. $\blacksquare$

APPENDIX II

Proofs of *Lemma 3* and *Lemma 4*

Proof: [Proof of Lemma 3] Assume that at a given moment, there are K_p active primary cells. The rate of the ith active primary cell is given by

$$R_p(i) = \frac{1}{64} \log \left(1 + \frac{P_p(i) g(\|X_{p,tx} - X_{p,rx}\|)}{N_0 + I_p(i) + I_{sp}(i)}\right) \tag{41}$$

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 $6^p \overline{a_p}$ and 16 primary interferers with a distance of at least $13^p \overline{a_p}$, and so on. As such, the $I_p(i)$ is upper-bounded by

$$I_p(i) = \sum_{k=1;k \neq i}^{K_p} P_p g(\|X_{p,tx}(k) - X_{p,rx}(i)\|) \tag{42}$$
$$< P \sum_{t=1}^{\infty} 8t(7t-1)^{-\alpha}, A:$$

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

$$I_{sp}^I(i) = \sum_{k=1;k \neq i}^{K_s} P_s g(\|X_{s,tx}(k) - X_{p,rx}(i)\|) \tag{43}$$
$$< P \sum_{t=1}^{\infty} 8t(7t-6)^{-\alpha}, B:$$

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 $2^p \overline{a_p}$ and 4 secondary

interferers with a distance of at least $9^{\rho}\sqrt{a_p}$, and so on. Then, $I_{sp}(i)$ is upper-bounded by

$$
I_{sp}^{II}(i) = \sum_{k=1;k\ne i}^{K_p} P_p g(\|X_{s,tx}(k) - X_{p,rx}(i)\|) \tag{44}
$$
$$
< P \sum_{t=1}^{d} 2t(7t-5)^{-\alpha} \triangleq C:
$$

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

$$
R_p(i) > \frac{1}{64} \log\left(1 + \frac{P(\sqrt{5})^{-\alpha}}{N_0 + 2P\sum_{t=1}^{P} 8t(7t-6)^{-\alpha}}\right): \tag{45}
$$

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

Proof: [Proof of Lemma 4] 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 III

Proofs of *Lemma 5* and *Lemma 6*

Proof: [Proof of Lemma 5] 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

$$
R_p(i) = \frac{1}{64} \log\left(1 + \frac{P_p(i)g(\|X_{p,tx} - X_{p,rx}\|)}{N_0 + I_p(i) + I_{sp}(i)}\right) \tag{46}
$$

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 $6^{\rho}\sqrt{a_p}$ and 16 primary interferers with a distance of at least $13^{\rho}\sqrt{a_p}$, and so on. As such, the $I_p(i)$ is upper-bounded by

$$
I_p(i) = \sum_{k=1;k\ne i}^{K_p} P_p g(\|X_{p,tx}(k) - X_{p,rx}(i)\|) \tag{47}
$$
$$
< P \sum_{t=1}^{d} 8t(7t-1)^{-\alpha} \triangleq A:
$$

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 $\rho \overline{a_s}$ 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

$$
\begin{aligned}
I_{sp}^{I}(i) &= \sum_{k=1,k\neq i}^{K_s} P_s g(\|X_{s,tx}(k) - X_{p,rx}(i)\|) \\
&< P \sum_{t=1}^{\Delta} 8t(7t-6)^{-\alpha}, \quad B:
\end{aligned}
\tag{48}
$$

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 $2\rho \overline{a_p}$ 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

$$
\begin{aligned}
I_{sp}^{II}(i) &= \sum_{k=1,k\neq i}^{K_p} P_p g(\|X_{p,tx}(k) - X_{p,rx}(i)\|) \\
&< P \sum_{t=1}^{\Delta} 8t(7t-5)^{-\alpha}, \quad C:
\end{aligned}
\tag{49}
$$

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

$$
\begin{aligned}
R_p(i) &= \frac{1}{64}\log\left(1+\frac{P_p(i)g(\|X_{p,tx}-X_{p,rx}\|)}{N_0+I_p(i)+I_{sp}^{I}(i)+I_{sp}^{II}(i)}\right) \\
&> \frac{1}{64}\log\left(1+\frac{P\rho^{-\alpha}(-5)}{N_0+3P\sum_{t=1}^{\Delta}8t(7t-6)^{-\alpha}}\right):
\end{aligned}
\tag{50}
$$

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

Proof: [Proof of Lemma 6] 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 IV

Derivation of (14)

We know that given $a_p(n)^{\frac{p}{2}} \log n = n$, the maximum throughput per S-D pair for the primary tier is $\frac{1}{n a_p(n)}$. 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 $\frac{1}{N n a_p(n)} = \frac{\sqrt{\log m}}{\sqrt{m} n a_p(n)}$. As such, based on (29), 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 $a_s(m)^{\frac{p}{2}} \log m = m$ when $a_p(n)^{\frac{p}{2}} \log n = n$.

REFERENCES

[1] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, pp. 388-404, Mar. 2000.

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

[3] A. Josan, M. Liu, D. L. Neuhoff, and S. S. Pradhan, "Throughput scaling in random wireless networks: a non-hierarchical multipath routing strategy," Preprint. Oct. 2007. [Online]. Available: http://arxiv.org/pdf/0710.1626.

[4] A. Ozgur, O. Leveque, and D. 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.

[5] M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad hoc wireless network," *IEEE/ACM Transaction on Networking*, vol. 10, no. 4, pp. 477-486, Aug. 2002.

[6] 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. 1917-1936, June 2005.

[7] A. E. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Optimal throughput-delay scaling in wirless networks–part I: the fluid model," *IEEE Transaction on Information Theory*, vol. 52, no. 6, pp. 2568-2592, June 2006.

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

[9] N. Bansal and Z. Liu, "Capacity, delay and mobility in wireless ad-hoc networks," *IEEE INFOCOM 2003*, vol. 2, pp. 1553-1563, Mar. 2003.

[10] S. Jeon, N. Devroye, M. Vu, S. Chung, and V. Tarokh, "Cognitive networks achieve throughput scaling of a homogeneous network," preprint. Jan. 2008. [Online]. Available: http://arxiv.org/pdf/0801:0938.

[11] C. Yin, L. Gao, and S. Cui, "Scaling laws for overlaid wireless networks: a cognitive radio network vs. a primary network," preprint. May 2008. [Online]. Available: http://arxiv.org/pdf/0805:1209.

[12] H. Daduna, *Queueing Networks with Discrete Time Scale*, Springer, 2001.

[13] 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.