Wireless Communications: Principles and Applications
Project Report

Li Yingzhe 5080309637
Dept. of Electronic Engineering
Shanghai Jiao Tong University, China
Email: fox9.1@sjtu.edu.cn

I. Introduction

The throughput scaling law for large-scale wireless ad hoc networks has been extensively studied since the seminal work of P. Gupta and P. R. Kumar [1]. They studied the random wireless network with n static nodes randomly located in the unit area and grouped into source-destination (S-D) pairs for transmission. Under the multi-hop relay algorithm, the network could achieve a per-node throughput of $\Theta(1/\sqrt{n \log n})$, with an average delay of $\Theta(\sqrt{n \log n})$. In addition, they also showed that when nodes are arbitrarily distributed, the maximum per-node throughput is $\Theta(1/\sqrt{n})$ under optimal conditions. The capacity performance has later been improved to $\Theta(1/\sqrt{n})$ by Franceschetti *et al.* [2] using percolation theory, even when the nodes are randomly located in the network area. After that, many other special cases of static wireless networks have been investigated. However, even under the optimal conditions, the capacity performance of the fixed network is still not very satisfying. This is mainly because in order to limit the interference from other nodes transmitting simultaneously, most transmissions should be carried out by using mulit-hop fashion. So much of the traffic is used to relay packets, and the per-node throughput could not be large enough.

Soon after P. Gupta and P. R. Kumar's seminal work on the capacity of wireless networks, M. Grossglauser and D. Tse introduced mobile nodes into the classical static network in [3]. In their work, the mobile nodes are allowed to transmit only when they are close to each other. They have shown that under the 2-hop relay algorithm, the mobile network could achieve a per-node throughput of $\Theta(1)$, which is much better than the classical multi-hop static network. However, this significant improvement of throughput capacity has been achieved at the cost of huge delay. Later, Neely *et al.* [4] re-evaluate this algorithm in a cellular structure, and show that the delay of the 2-hop mobile network could be $\Theta(n)$.

Since mobility could improve the capacity of wireless networks, various mobility models and the corresponding impact of mobility on both the capacity and delay performance of the network have been studied in later works . These include the i.i.d. mobility model [5], [6], [11]; Brownian mobility model [7], [8]; random way-point mobility model [9], [10]; and random walk mobility model [11], [12].

Besides studying on specific mobility models, some works have also studied the general mobility models, which could characterize certain amount of mobility models. Sharma *el al.* [13], studied a family of hybrid random walk models, in which the i.i.d. mobility model and the random walk mobility model are the two extreme cases. They also studied another family of mobility models called the discrete random direction models, in which the random way-point mobility model and the Brownian mobility model are the two extreme cases of this family of mobility models. They came up with the notion of critical delay, i.e., the least delay to be tolerated in order to achieve the same throughput scaling as the static network, and shown the different value of critical delay for the two family of mobility models.

The literatures mentioned above mainly focus on the scaling laws for capacity and delay in a single network. In recent years, the crowded frequency resource brings a new kind of software-defined radio into the focus of research. According to a report from the Federal Communication Commission (FCC) Spectrum Task Force [14], over 90 percent of the licensed spectrum remains idle at a given time and location. The software-defined radio called cognitive radio (CR) could perform spectrum sensing and access a wide range of licensed spectrum that are not being used by the licensed users. In the cognitive radio network, the primary uses have the higher priority to the spectrum, while the secondary users could access the spectrum opportunistically in order to limit the interference to the primary users. This new technology could greatly improve the spectrum efficiency and enable disparate radio communication devices onto the same platform.

Till now, many work have been done to investigate the capacity of large-scale static CRN, as well as the impact of CR users to the traditional licensed users. Shi *et al.* [15] constructed a homogeneous auxiliary network to estimate a lower bound on the capacity of a heterogeneous CRN. However their model requires the total band knowledge at any node and has not considered the interaction between the primary and secondary networks. Then in [16], Jeon *et al.* considered a licensed primary network and a cognitive secondary network coexisting in a planar area. By letting SUs opportunistically access the licensed spectrum without causing severe interference to PUs, they show that when the secondary network is denser than the primary network, both networks can simultaneously achieve the same throughput scaling law as a stand-alone network. In [17], Li and Dai further proved that even when the number of primary users n and the number of secondary users m are the same in order sense, the CRN could still achieve the $\Theta(\frac{1}{\sqrt{n}})$ and $\Theta(\frac{1}{\sqrt{m}})$ per-node throughput for the primary network and secondary network respectively. Later, in [18], Huang *et al.* characterized the conditions for the cognitive networks to achieve the same throughput and delay scaling as the stand-alone networks.

Those aforementioned results are obtained on a non-cooperative scheme without considering possible positive interactions between two coexisting networks, so the capacity and delay performance may not be further improved. Later, Gao *et al.* [19] proposed a supportive cognitive network, in which the secondary nodes may route packets for the primary network. In the scenario where both the primary users and secondary users are static, PUs can achieve near-optimal capacity performance, i.e. $\lambda_p = \Theta(1/\log n)$. While SUs may still achieve the same scaling laws as a stand-alone ad hoc network. Since mobility could improve the capacity scaling of ad-hoc wireless networks, thus it is natural to consider whether mobility would influence the performance of the CRN. Thus, in [19], the authors further considered the condition when the secondary users are set to be mobile. Under this scenario, the primary users could achieve the per-node throughput scaling of $\lambda_p = \Theta(1/\log n)$, and delay scaling of $\Theta(1)$ when SUs move according to the i.i.d. mobility model and $\Theta(1/S)$ under random walk mobility model, where S is the random walk step size. While the secondary users could achieve the per-node throughput scaling of $\lambda_s = \Theta(1)$. Such performance of capacity and delay scaling for both the primary and secondary users is theoretically optimal. However, this cooperative scheme requires that the number of supportive secondary users should at least $\Theta(n^2)$, which is much larger than the number of primary users, and thus casts a heavy burden on its implementation.

Motivated by the fact that various mobility models could significantly influence the capacity and delay scaling of the wireless networks, as well as the fact that cooperation among primary users and mobile secondary users could improve the performance of the CRN, but at the cost of large number of secondary users. Thus we want to study a CRN where secondary users would move under some general mobility model, i.e., such mobility model that could characterize a large amount of mobility models. And in order to fully utilize the mobility of secondary users and consequently decrease the number of secondary users, we want to come up with a scheme that could achieve the near-optimal performance of the primary users and secondary users but with much less secondary users. Such scheme would not only have the optimal scaling performance but also easier to implement due to the relative small number of secondary users.

So in our work, we should focus on such kind of CRN which satisfies our requirements: a static primary network and a mobile secondary network. We imitate the heterogeneous mobility of different transportation vehicles, and extend the step-wise roaming process to a more generalized hierarchical system. Our current main contributions are as follows:

- We develop a new kind of heterogeneous mobility model for the secondary cognitive network – the hierarchical hybrid mobility model. In this model, mobile nodes are divided into several layers according to their mobility. In particular, different layers correspond with different sizes of moving area – from global to regional. Different layer mobile nodes would have different speed, which is proportional to the square root of their moving area; while the mobile nodes within the same layer would have the same speed. Specifically, when the speed of each layer mobile nodes is maximum, then the mobile nodes are moving according to the i.i.d. mobility model within their moving area. While when the speed of each layer mobile nodes is minimum, it is equivalent for them to move according to the random walk mobility model within their moving area.
- We propose a new class of hierarchical relay algorithm, which originates from the Grossglauser-Tse 2-hop relay algorithm, to make a balanced use of relay and mobility. In this algorithm, several relays with different mobility are used through the routing process, and the packet approaches its destination in a step-wise fashion.
- With the help of hierarchical relay among different layers of SUs, this algorithm can help the primary network achieve $\Omega(n^{-\delta'})$ per-node throughput under optimal conditions. While the delay performance of primary network is $D_p = \Theta\left(h^2 n^{(1+\epsilon')/h}\right)$ under i.i.d. mobility model; $D_p = \Theta\left(h^2 n^{(1+\epsilon')/h} \log n\right)$ under random walk mobility model, where h denotes the number of different layers of the mobility of secondary users. In both cases, an optimal delay performance of $O(n^{\delta''})$ could be achieved.
- Under this hierarchical relay algorithm, the secondary network could achieve the capacity scaling of $\lambda_s = \Theta(h^{-1} n^{\epsilon'-\epsilon})$. While the delay performance of the secondary network is $D_{s,k_d} = \Theta\left(h(k_d-1)n^{(1+\epsilon')/h} + hn^{(h-k_d+1)(1+\epsilon')/h}\right)$ for the i.i.d. mobility model, where $k_d = 1, 2, ..., h$ denotes the layer of the destination SU. For the random walk mobility model, with random walk step size S, $D_{s,k_d} = \Theta\left(h(k_d-1)|logS|/S^2 + hn^{(h-k_d+1)(1+\epsilon')/h}\right)$, when $k_d = o(h)$, and $D_{s,k_d} = \Theta\left(hk_d|logS|/S^2\right)$ when $k_d = \Theta(h)$.
- By achieving the near-optimal performance for both the capacity and delay scaling of the primary network, the hierarchial relay algorithm only needs the number of SUs to be $O(n^{1+\epsilon})$, which is much less than that in [19], thus more easy to implement.

II. System Model

In this paper, we consider a static primary network and a multi-layer mobile secondary cognitive network coexisting in a planar unit square area. To be specific, the primary network consists of n static, randomly and evenly distributed PUs, which are grouped into S-D pairs one by one. The secondary network is divided into h layers according to different moving range. Each layer consists of $n^{1+\epsilon}$ mobile, randomly and evenly distributed cognitive SUs, where ϵ can be any small positive value. All these $m = hn^{1+\epsilon}$ SUs are also grouped into S-D pairs one by one.

A. Transmission Model

A typical wireless propagation channel is usually affected by path-loss, shadowing and fading effects. In this work, we ignore the influence of shadowing or fading for simplicity, and assume that the channel gain depends only on the distance of transmission. As such, the channel power gain $g(d)$ is given as

$$g(d) = d^{-\alpha}, \tag{1}$$

where d denotes the distance of transmission, $\alpha \geq 2$ represents the path-loss exponent.

We adopt the Gaussian channel model to regulate the transmission rate, which is a continuous function of the Signal to Interference plus Noise Ratio (SINR). Specifically, the data rate from a primary transmitter P_i to its receiver $P_{D(i)}$ is determined by:

$$R(P_i, P_{D(i)}) = log(1 + \frac{P_p g(\|P_i - P_{D(i)}\|)}{N_0 + I_p + I_{sp}}), \tag{2}$$

Here, P_p is the transmission power for the primary nodes, and N_0 is the ambient noise power. $\|\cdot\|$ denotes the Euclidian distance for two nodes in the unit area. Moreover, I_p is the sum interference from all the other concurrent primary transmitters to the receiver $P_{D(i)}$, I_{sp} is the sum interference from all the current secondary transmitters to $P_{D(i)}$. Suppose there are N_p and N_s simultaneous primary and secondary transmitters, then I_p is determined by:

$$I_p = P_p \sum_{k=1, k\neq i}^{N_p} g(\|P_k - P_{D(i)}\|), \tag{3}$$

Further if we set P_s to be the transmission power for the secondary nodes, and S_k $(1 \leq k \leq N_s)$ to be the secondary transmitters. Then I_{sp} is determined by:

$$I_{sp} = P_s \sum_{k=1}^{N_s} g(\|S_k - P_{D(i)}\|), \tag{4}$$

Similarly, the data rate from the secondary transmitter S_i to its receiver $S_{D(i)}$ is defined as:

$$R(S_i, S_{D(i)}) = log(1 + \frac{P_s g(\|S_i - S_{D(i)}\|)}{N_0 + I_s + I_{ps}}), \tag{5}$$

Here I_s is the sum interference from all the other simultaneous secondary transmitters, and I_{ps} represents the total interference from the concurrent primary transmitters.

We assume that all nodes, be it a PU or a SU, are allowed to transmit at the same power level, which means that P_p=P_s.

B. Cellular Structure

In this paper, we follow a cell-based algorithm, which corresponds with hierarchical mobility model to be presented later. So we first define h layers of cells.

We first partition the unit area into $N_{c,1} = n^{(1+\epsilon')/h}$ non-overlapping, rectangular cells, each covering an area of $n^{-(1+\epsilon')/h}$, where ϵ' can be any positive value that is smaller than ϵ. These cells are defined as 1st-layer cells. Similarly, we partition each 1st-layer cell into $N_{2,1} = n^{(1+\epsilon')/h}$ non-overlapping, rectangular 2nd-layer cells, each covering an area of $n^{-2(1+\epsilon')/h}$, thus the total number of 2nd-layer cells in the unit area should be $N_{c,2} = n^{2(1+\epsilon')/h}$, and so forth.

In the same way, we can obtain a series of cellular structure. For $1 \leq k \leq h$, we have $N_{c,k} = n^{k(1+\epsilon')/h}$ kth-layer cells, as shown in Figure 1.

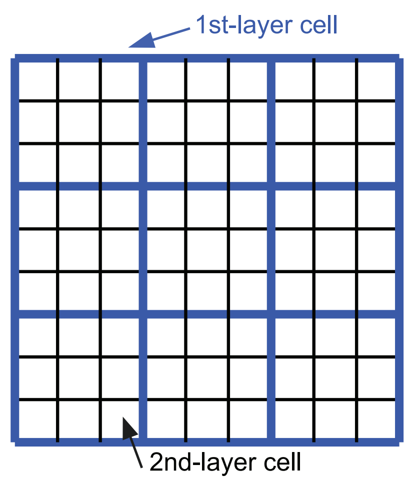

Fig. 1: Cellular Structure

In the previous part, we have assumed that all nodes transmit at the same power level, here we set the transmission range of these nodes to be one bottom-layer cell. In other words, any node can only communication with another node in the same hth-layer cell. In order to distinguish the special bottom-layer cell from other hierarchies, we give them an alias – ‘lattice’ in the following sections.

C. Mobility Model

The secondary cognitive network follows a special-designed heterogeneous mobility model. In this model, SUs are divided into h layers according to their mobility. Different layers correspond with different sizes of moving area. Here the moving area of a certain SU is defined as a rectangular region centered at its initial position.

In the previous subsection, we have defined a series of cells with different sizes. Here we associate them with our proposed mobility model. 1st-layer SUs are set to move around globally in the unit area, while others are assumed to move regionally within their moving area respectively. To be specific, a kth-layer SU $\mathcal{S}_{k,i}$, $k = 2, 3, ..., h$, is assumed to move around in a $(k-1)$th-layer cell.

In latter sections, we use $\mathcal{S}_{k,i} \in \mathcal{C}_{k',j}$ to denote $\mathcal{S}_{k,i}$ belongs to $\mathcal{C}_{k',j}$, if the moving area of a kth-layer SU $\mathcal{S}_{k,i}$ is inside a k'th-layer cell $\mathcal{C}_{k',j}$, $k = 2, 3, ..., h$, $k' = 1, 2, ..., k-1$. Similarly, if a PU $\mathcal{P}_i$ is inside an kth-layer cell $\mathcal{C}_{k,j}$, we say that $\mathcal{P}_i$ belongs to $\mathcal{C}_{k,j}$, $\mathcal{P}_i \in \mathcal{C}_{k,j}$, $k = 1, 2, ..., h$.

We have mainly studied two types of mobility models for the secondary users.

(1)**Two-dimensional i.i.d. mobility model:** All mobile SUs, irrespective of their layers, are supposed to move within their moving range according to the uniform independent and identical distribution. Within one time slot, all the SUs keep static, so that the transmission of all links keeps stable. While after this time slot, the position of all SUs are perfectly reshuffled, i.e., they become randomly, evenly and independently redistributed within their moving area, irrespective of their moving history, as shown in Figure 2.

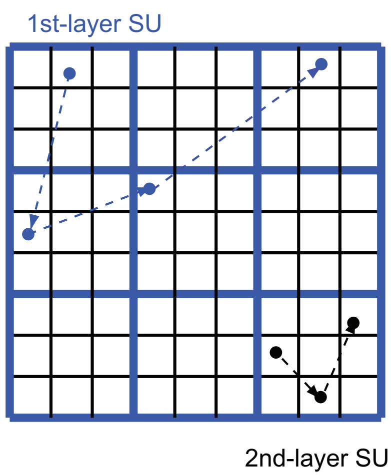

Fig. 2: Moving area of Secondary Nodes with i.i.d mobility model

(2) **Two-dimensional random walk mobility model [13]:** Pick a random kth-layer SU $\mathcal{S}_{k,i}$ which moves around in the $(k-1)$-th layer cell $\mathcal{C}_{k-1,j}$, where $k = 1, 2, ...h$. Divide each $\mathcal{C}_{k-1,j}$ into $1/S^2$ equal size random walk cells(RW-cells), and index each RW-cells by (i, j), where $i, j \in 0, ..., 1/S - 1$. Here we set the value of S to be $1 < \frac{1}{S} \leq n^{(1+\epsilon')/2h}$. When k ranges from 1 to h, all the $(k-1)$-th layer cells would choose the same value of S for simplicity. Here we call S the random walk step size.

Then we say $\mathcal{S}_{k,i}$ is moving according to the random walk mobility if within one time slot, it stays in one of the RW cells inside $\mathcal{C}_{k-1,j}$, denoted by (i, j). While in the next time slot, $\mathcal{S}_{k,i}$ would move to one of its 8 adjacent RW-cells or stays at the same RW-cell with equal probability. For simplicity, if $\mathcal{S}_{k,i}$ hits the boundary of $\mathcal{C}_{k-1,j}$, it will jump over the opposite edge.

We should note here that the RW-cell is to regulate the distance that each k-th layer SU could travel during one time slot. So the larger the RW-cell inside each $k-1$-th layer cell, the larger the speed that the k-th layer SU would have.

From the above definition, when $1/S$ approaches 1, the $\mathcal{S}_{k,i}$ will move according to the i.i.d. mobility model as defined previously. This corresponds to the largest speed that $\mathcal{S}_{k,i}$ will have. And when $1/S$ is set to be $n^{(1+\epsilon')/2h}$, the RW-cell is just the same as the k-th layer cell that $\mathcal{C}_{k-1,j}$ is divided into. So $\mathcal{S}_{k,i}$ would jump from one of the k-th layer cell to the adjacent or the same k-th layer cell in the next time slot, which represents the smallest speed that $\mathcal{S}_{k,i}$ has.

The random walk mobility of the 1st layer secondary user in the unit square is shown in Figure 3, and the similar condition could be applied to the k-th layer SU moving within the $(k-1)$-th layer cell, for $2 \leq k \leq h$.

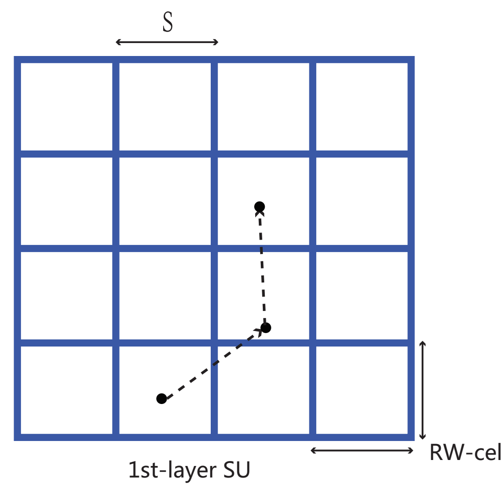

Fig. 3: Moving area of Secondary Nodes with random walk mobility model

D. Capacity and Delay

The per-node throughput of a S-D pair is defined as the data rate (in bits/time-slot) that each source node can transmit to its destination, which depends on the network density. For the primary network, we denote its per-node throughput by λ_p, while that of the secondary network is denoted by λ_s.

The delay of S-D pair is defined as the average number of time-slots passed before the packet arrives at its destination after it leaves the source node, which also depends on the network density. For the primary network, we use D_p to denote its average delay. As for the secondary network, more complex analysis is needed for different S-D pairs. Later, we will show that the average delay of a secondary S-D pair depends largely on the mobility of its destination node. The bigger the moving range, the larger the delay. So we use $D_{s,k}$ to denote the average delay of a secondary S-D pair with a kth-layer destination.

Finally, we list the notations in Table I.

TABLE I: Definition of Symbols and Notations

Symbol	Definition
n	Number of primary users
m	Number of secondary users
h	Number of layers in the hierarchical structure
α	Path-loss exponent
$lattice$	The h-th layer (bottom layer) cell
$N_{c,i}$	Number of ith-layer cells in the unit area
$N_{i,j}$	Number of ith-layer cells in a jth-layer cell
$\mathcal{P}_i$	The ith primary user
$\mathcal{S}_{k,i}$	The ith secondary user of the kth-layer
$\mathcal{C}_{k,i}$	The ith cell of the kth-layer
$\mathcal{U}$	Unit area
n_k	Number of primary users in a kth-layer cell
$m_{i,k}$	Number of ith-layer secondary user in a kth-layer cell
λ_p	Per-node throughput of the primary network
λ_s	Per-node throughput of the secondary network
P_p	Transmission power for primary transmitters
P_s	Transmission power for secondary transmitters
D_p	Delay of primary S-D pairs
$D_{s,k}$	Delay for secondary S-D pairs with a kth-layer destination
$P_i \in C_{k,j}$	P_i is inside the k-th layer cell $C_{k,j}$
$S_{k,i} \in C_{k',j}$	The moving area of $S_{k,i}$ is inside k'-th layer cell $C_{k',j}$
$C_{k,i_k} \in C_{k',i_{k'}}$	The k-th layer cell C_{k,i_k} is inside the k'-th layer cell $C_{k',i_{k'}}$
S	Random walk step size

III. Hierarchical Relay Algorithm

In this section, we sketch a new hierarchical relay algorithm for the CRN, which makes best use of the proposed mobility model.

A. Primary Network Relay Algorithm

The primary algorithm seeks to take advantage of the secondary network extensively so as to achieve high capacity performance with low delay for PUs. And we assume that the secondary users could act as the relay for the primary packets, so the positive cooperation will exist between the primary nodes and secondary nodes.

Before going to details of the primary relay algorithm, we first introduce a lemma which guarantees the feasibility of each step of this algorithm.

Lemma 1: Pick a random lattice $\mathcal{C}_{h,i_h}$, assume:

$$\mathcal{C}_{h,i_h} \in \mathcal{C}_{h-1,i_{h-1}} \in ... \in \mathcal{C}_{1,i_1}.$$

Then at any moment, according to the proposed cellular structure and SU distribution, for any k satisfying $1 \le k \le h$, at least one k-th layer SU would reside in $\mathcal{C}_{h,i_h}$.

The above conditions are satisfied by all lattices with high probability.

Proof: In our cellular structure, $n^{1+\epsilon}$ SUs of each layer are randomly and evenly distributed into $N_{c,h} = n^{1+\epsilon'}$ lattices, with $0 < \epsilon' < \epsilon$. Pick a random lattice $\mathcal{C}_{h,i_h}$, as all the k-th layer cells are randomly distributed, so the probability that there are no k-th layer SUs inside it is given by:

$$\begin{aligned} & Pr\{m_{k,h}(i_h) = 0\} \\ & = (1 - n^{-(1+\epsilon')})^{n^{1+\epsilon}} \\ & \to e^{-n^{\epsilon-\epsilon'}} \\ & \to 0 \end{aligned}$$

as $n \to \infty$.

Since the lattice is randomly selected, so the above condition is satisfied by all the lattices. Thus at least one k-th layer SU could be found in any lattice, for $1 \le k \le n$.

This completes the proof. ■

Now we sketch the relay algorithm of the primary network.

First we define some special actions through the routing process in Table II.

TABLE II: Definition of Actions

Action	Definition
$\mathcal{T} \overset{B}{\Longrightarrow} \mathcal{R}$	Node $\mathcal{T}$ transmits packet B to node $\mathcal{R}$, and node $\mathcal{R}$ is in the same lattice as node $\mathcal{T}$ at the moment
$\mathcal{T} \xrightarrow{\mathcal{C}_{k,i}} \mathcal{C}_{k',j}$	Mobile node $\mathcal{T}$ moves around within kth-layer cell $\mathcal{C}_{k,i}$ until it arrives at k'th-layer cell $\mathcal{C}_{k',j}$, $k' > k$
$\mathcal{T} \overset{\mathcal{C}_{k,i}}{\rightleftharpoons} \mathcal{R}$	Mobile node $\mathcal{T}$ moves around within kth-layer cell $\mathcal{C}_{k,i}$ until it encounters mobile node $\mathcal{R}$ in a same lattice

Next, pick a random primary source node $\mathcal{P}_i$, whose corresponding destination node is $\mathcal{P}_j$. Assume the location of $\mathcal{P}_j$ as follows:

$$\mathcal{P}_j \in \mathcal{C}_{h,j_h} \in \mathcal{C}_{h-1,j_{h-1}} \in ... \in \mathcal{C}_{1,j_1}.$$

Then the primary S-D pair would use the secondary users of different layers to relay the primary packet B_p generated at $\mathcal{P}_i$. The relay algorithm is shown in Algorithm 1. And Figure 4 illustrates the complete routing path of the primary packet B_p.

Algorithm 1 Relay Algorithm for Primary Packet B_p

Input: The primary source node P_i and destiantion node P_j
Output: The h intermediate secondary relay nodes

1: $\mathcal{P}_i \overset{B_p}{\Longrightarrow} \mathcal{S}_{1,u_1}$
2: $\mathcal{S}_{1,u_1} \xrightarrow{\mathcal{U}} \mathcal{C}_{1,j_1}$
3: $\mathcal{S}_{1,u_1} \overset{B_p}{\Longrightarrow} \mathcal{S}_{2,u_2} \in \mathcal{C}_{1,j_1}$
4: **for** k=2 to $(h-1)$ **do**
5: $\quad \mathcal{S}_{k,u_k} \xrightarrow{\mathcal{C}_{k-1,j_{k-1}}} \mathcal{C}_{k,j_k}$
6: $\quad \mathcal{S}_{k,u_k} \overset{B_p}{\Longrightarrow} \mathcal{S}_{k+1,u_{k+1}} \in \mathcal{C}_{k,j_k}$
7: **end for**
8: $\mathcal{S}_{h,u_h} \xrightarrow{\mathcal{C}_{h-1,j_{h-1}}} \mathcal{C}_{h,j_h}$
9: $\mathcal{S}_{h,u_h} \overset{B_p}{\Longrightarrow} \mathcal{P}_j$

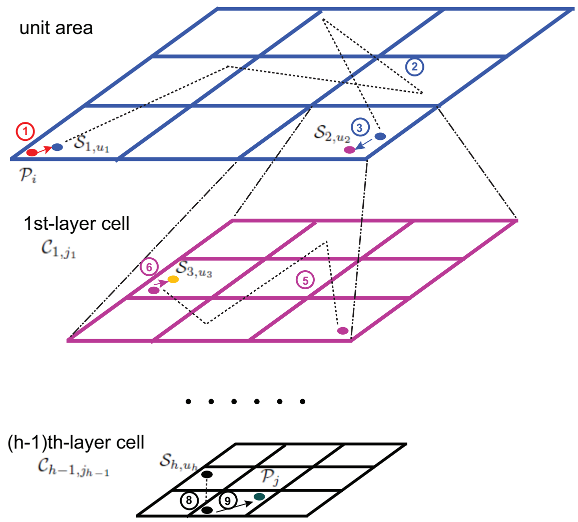

Fig. 4: Routing Path of Primary Packet B_p under h-layer Hierarchical Relay Algorithm

Algorithm 2 Relay Algorithm for Secondary Packet B_{s,k_d}

Input: The secondary source node $\mathcal{S}_{k_s,i}$ and destiantion node $\mathcal{S}_{k_d,j}$

Output: The k_d intermediate secondary relay nodes

$\mathcal{S}_{k_s,i} \overset{B_{s,k_d}}{\Longrightarrow} \mathcal{S}_{1,u_1}$

$\mathcal{S}_{1,u_1} \xrightarrow{\mathcal{U}} \mathcal{C}_{1,j_1}$

$\mathcal{S}_{1,u_1} \overset{B_{s,k_d}}{\Longrightarrow} \mathcal{S}_{2,u_2} \in \mathcal{C}_{1,j_1}$

for k=2 to (k_d-1) **do**

  $\mathcal{S}_{k,u_k} \xrightarrow{\mathcal{C}_{k-1,j_{k-1}}} \mathcal{C}_{k,j_k}$

  $\mathcal{S}_{k,u_k} \overset{B_{s,k_d}}{\Longrightarrow} \mathcal{S}_{k+1,u_{k+1}} \in \mathcal{C}_{k,j_k}$

end for

$\mathcal{S}_{k_d,u_{k_d}} \xleftrightarrow{\mathcal{C}_{k_d-1,j_{k_d-1}}} \mathcal{S}_{k_d,j}$

$\mathcal{S}_{k_d,u_{k_d}} \overset{B_{s,k_d}}{\Longrightarrow} \mathcal{S}_{k_d,j}$

Under this algorithm, a primary packet is relayed along SUs from the top layer to the bottom layer. In the meantime, this packet is approaching its destination in a step-wise fashion. This hierarchical algorithm takes best advantage of the hierarchical mobility model, and can greatly reduce the time needed for the roaming process. The detailed interpretation of this algorithm shall be presented in the next two sections.

B. Secondary Network Relay Algorithm

The secondary network relay algorithm deals with transmission requests from secondary S-D pairs, and this algorithm also utilize the hierarchical cooperation among different layers of secondary users.

Pick a random secondary source node $\mathcal{S}_{k_s,i}$, whose corresponding destination node is $\mathcal{S}_{k_d,j}$, here we call the packet from $\mathcal{S}_{k_s,i}$ to $\mathcal{S}_{k_d,j}$, denoted by B_{s,k_d}, a k_dth-layer secondary packet.

Assume the location of $\mathcal{S}_{k_d,j}$ as follows:

$$\mathcal{S}_{k_d,j} \in \mathcal{C}_{k_d-1,j_{k_d-1}} \in \mathcal{C}_{k_d-2,j_{k_d-2}} \in ... \in \mathcal{C}_{1,j_1}.$$

Then the routing process of this k_dth-layer secondary packet B_{s,k_d} is given in Algorithm 2.

Under this algorithm, a secondary packet is relayed along SUs from the top layer to the hierarchy of its destination SU. The larger mobility its destination node possesses, the less relays it needs. This packet is not relayed further down to the lowest layer because the time needed for two nodes to meet at the same cell only depends on the one with larger moving area, and thus more relay can not further reduce the roaming time.

IV. Scheduling Schemes for the Hierarchical Relay Algorithm

After we have defined the hierarchical relay algorithm, we would build up the scheduling scheme for the CRN. The scheduling scheme would choose which primary or secondary S-D pair to be activated in each time slot, and the transmission range of each successful transmission is restricted to be within a lattice in our scheme. Specifically, the primary scheduling scheme is designed to transmit the primary packets, while the secondary scheduling scheme is to transmit the secondary packets.

In order ensure the equal opportunity for all the lattices to be active and limit the interference among concurrent transmissions, a 9-TDMA scheme is adopted which is similar to that in [2]: Divide all the lattices into 9 subsets according to a 3×3 pattern, as shown in Figure 5, then the lattices in different subset would be activated with a round-robin fashion during a certain time period.

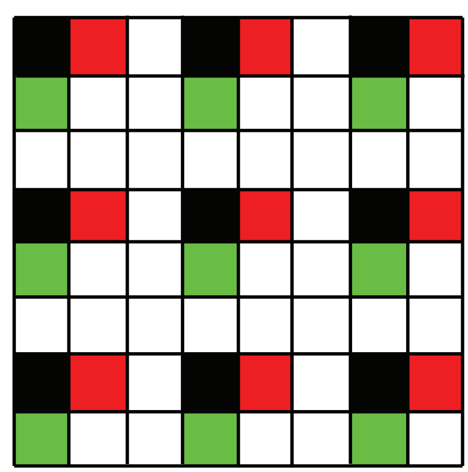

Fig. 5: 9-TDMA Subsets of Lattices

In our scheme, the primary time slot and secondary time slot has the same length. Since the secondary users would act as the relay node for the primary packets, thus in order to guarantee the transmission opportunity for both the primary packets and secondary packets, we divide the secondary time slot into two equal subslots. The first subslot is used to relay the primary packets and the second subslot is used to deliver the secondary packets. During each primary time slot and secondary subslot, a 9-TDMA scheme would be adopted for successful transmission.

In the following, the scheduling scheme for the primary network and secondary network and established respectively.

A. Primary Scheduling Scheme

From the hierarchical relay algorithm, a primary packet is relayed by h different layer secondary users, so the primary scheduling scheme is associated with $h+1$ phases and each phase consumes one time slot. If a lattice is active in a certain time period, at most one node in it is allowed to transmit during this period. Note that when two or more relay SUs satisfy the condition to be scheduled, one of them would be chosen randomly and relay the primary packet during first secondary subslot. Then the specific phases of the primary scheduling scheme are as follow:

Phase 1: During the active period of lattice, randomly select a source PU $\mathcal{P}_i$ in the lattice. Let $\mathcal{P}_i$ transmit a primary packet B_p to a random 1st-layer relay SU $\mathcal{S}_{1,u_1}$ in the same lattice;

For $k = 2, 3, ..., h$,

Phase k: During the active period of each lattice, randomly select a $(k-1)$th-layer SU $\mathcal{S}_{k-1,u_{k-1}}$ in the lattice. If $\mathcal{S}_{k-1,u_{k-1}}$ contains a primary packet B_p whose destination P_j belongs to the same $(k-1)$-layer cell as the designated lattice, let $\mathcal{S}_{k-1,u_{k-1}}$ relay B_p to a random kth-layer relay SU $\mathcal{S}_{k,u_k}$ in the same lattice. Otherwise, re-select another 1st-layer SU in the lattice until all $(k-1)$th-layer SUs have been tried;

Phase h+1: During the active period of each lattice, randomly select a destination PU P_j in the lattice. If there is an hth-layer relay SU $\mathcal{S}_{h,u_h}$ in the same lattice carrying a primary packet B_p which is destined to P_j, then let P_j receive B_p from $\mathcal{S}_{h,u_h}$. Note that if two or more relay SUs match this condition, we choose one of them randomly. Otherwise, re-select another destination PU in the lattice until all destination PUs have been tried;

B. Secondary Scheduling Scheme

First, we define the *preservation region* so as to keep the interference from the secondary users to the primary users, which has the similar idea to [19]. The definition of the preservation region is a square that contains 9 lattices, with the active primary transmitter or receiver at the center cell, as shown in Figure 6.

Only the secondary transmitters outside any preservation region can they transmit the secondary packets. Otherwise, it should buffer the packet until it is outside the preservation region. From the definition of preservation region, the secondary users could be scheduled only when they are outside the current preservation regions. In the following section, we would prove that this constraint will not degrade the performance of the secondary users when n approaches infinity.

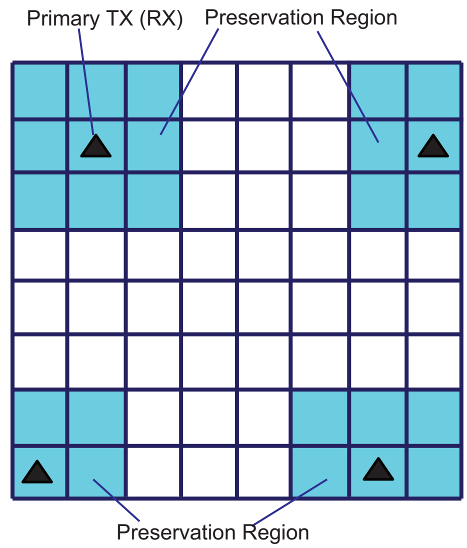

Fig. 6: Illustration of the preservation region

Since the secondary packet would be relayed at most h times, so the secondary scheduling scheme would consume $h+1$ phases. Then the specific phases of the primary scheduling scheme are as follow, note that when two or more SUs satisfy the condition to be scheduled, one of them would be chosen randomly during second secondary subslot:

Phase 1: During the active period of each lattice, randomly select a source SU $\mathcal{S}_{k_s,i}$ in the lattice. Let $\mathcal{S}_{k_s,i}$ transmit a secondary packet B_{s,k_d} to a random 1st-layer relay SU $\mathcal{S}_{1,u_1}$ in the same lattice;

For $k = 2, 3, ...h$,

Phase k: During the active period of each lattice, randomly select a $(k-1)$th-layer SU $\mathcal{S}_{k-1,u_{k-1}}$ in the lattice. If $\mathcal{S}_{k-1,u_{k-1}}$ contains a $(k-1)$th-layer secondary packet $B_{s,k-1}$ whose destination $\mathcal{S}_{k-1,j_{k-1}}$ happens to be in the same lattice at this moment, let $\mathcal{S}_{k-1,u_{k-1}}$ relay $B_{s,k-1}$ to $\mathcal{S}_{k-1,j_{k-1}}$; or if $\mathcal{S}_{k-1,u_{k-1}}$ contains a k'th-layer secondary packet $B_{s,k'}$ $(k \le k' \le h)$ whose destination $\mathcal{S}_{k',j_{k'}}$ belongs to the same $(k-1)$th-layer cell as the designated lattice, let $\mathcal{S}_{k-1,u_{k-1}}$ relay $B_{s,k'}$ to a random kth-layer relay SU $\mathcal{S}_{k,u_k}$ in the same lattice.

Phase h+1: During the active period of each lattice, randomly select an hth-layer destination SU $\mathcal{S}_{h,j_h}$ in the lattice. If there is another hth-layer relay SU $\mathcal{S}_{h,u_h}$ in the same lattice carrying a secondary packet $B_{s,h}$ which is destined to $\mathcal{S}_{h,j_h}$, then let $\mathcal{S}_{h,j_h}$ receive $B_{s,h}$ from $\mathcal{S}_{h,u_h}$.

In the following two sections, we would calculate the capacity and delay performance for both the primary network and secondary network, respectively. In addition, both the condition that SUs are moving according to i.i.d. mobility model and random walk mobility model are considered.

V. Capacity and Delay Performance for the Primary network

In this section, we would first evaluate the capacity scaling of the primary network, and then study the delay performance for the primary network when the secondary users are moving according to the i.i.d. mobility model and random walk mobility model respectively.

A. Capacity Performance

In this part, we would first calculate the upper bound for the number of primary users within each lattice, and then prove all the transmitters could support a constant data rate during each phase. Finally we would give the throughput of the primary network.

***Lemma** 2 (Ji et al. [20]):* Assume x nodes are placed into y equal-sized areas randomly, evenly and independently. Let $Z(x,y)$ be the random variable that counts the maximum number of nodes in any area. Then with high probability,

$$Z(x,y) = \begin{cases} \Theta(\frac{x}{y}), & \text{if } x \gg y\log y, \\ \Theta(\log y), & \text{if } x = cy\log y \text{ for some constant } c, \\ \Theta(\frac{\log y}{\log \frac{y \log y}{x}}), & \text{if } \frac{y}{polylog(y)} \le x \ll y\log y, \\ \Theta(\frac{\log y}{\log \frac{y}{x}}), & \text{if } x < \frac{y}{\log y}. \end{cases} \tag{6}$$

***Lemma** 3:* According to the h-layer cellular structure, there are at most $\Theta(1)$ PUs in any lattice with high probability.

Proof: In our h-layer cellular structure, n PUs are randomly and evenly distributed into $N_{c,2} = n^{1+\epsilon'}$ lattices. According to the last condition in Lemma 2, the maximum number of PUs in any lattice should be

$$\max(n_2) = \Theta(\frac{1+\epsilon'}{\epsilon'}) = \Theta(1). \tag{7}$$

■

***Lemma** 4:* During all the phases of the primary scheduling scheme, each transmitter within a lattice could support a constant data rate .

Proof: In order to prove that every transmitter in every phase of the scheduling scheme could support a constant data rate, we divide the whole routing process into three parts: input, relay and output. The input process corresponds to the Phase 1 of the scheduling scheme, and the output process corresponds to the Phase $h+1$ of the scheduling process, and the relay processes would correspond to the intermediate Phase k, where $2 \le k \le h$. Next we would prove all the transmitters in each of the process could support a constant data rate.

First consider the input process, during which the primary transmitters would transmit the primary packet to the 1st layer SU in the same lattice. Since a 9-TDMA scheme is adopted, so the data rate of the primary transmitter could be given as:

$$R(P_i, S_{1,i_1}) = \frac{1}{9} log(1 + \frac{P_p g(\|P_i - S_{1,i_1}\|)}{N_0 + I_p + I_{sp}}), \tag{8}$$

Where P_i is the primary transmitter and S_{1,u_1} is the 1st-layer secondary relay node. $\frac{1}{9}$ is introduced by dividing the primary time slot into 9 TDMA subslots. Since we have restricted the transmission range to be within a lattice, so the transmission power for the primary transmitter is $P_p = Pl^\alpha$, where P is a constant and l is the side length of a lattice.

Since the transmission from the source primary node to the 1st-layer relay SU is within a lattice, so $\|P_i - S_{1,i_1}\| \le \sqrt{2}l$, thus

$$\begin{aligned} P_p g(\|P_i - S_{1,i_1}\|) &= Pl^\alpha(\|P_i - S_{1,i_1}\|)^{-\alpha} \\ &\ge Pl^\alpha(\sqrt{2}l)^{-\alpha} \\ &= P(\sqrt{2})^{-\alpha} \end{aligned} \tag{9}$$

Then we should consider I_p, which is the sum interference from the concurrent primary transmitters. Since the 9-TDMA scheme is adopted, so from Fig 7, there would be at most 8 concurrent primary transmitters with a distance of at least $2l$ from S_{1,i_1}, and 16 primary transmitters with a distance at least $5l$ and so on. So the I_p is bounded by:

$$\begin{aligned} I_p &= \sum_{k=1,k\neq i}^{N_p} P_p g(\|P_k - S_{1,i_i}\|) \\ &\le \sum_{k=1,k\neq i}^{\infty} 8k[(3k-1)l]^{-\alpha} Pl^\alpha \\ &= 8P \sum_{k=1,k\neq i}^{\infty} \frac{k}{(3k-1)^\alpha} \\ &= R_1 \end{aligned} \tag{10}$$

Here R_1 is a constant. Since the path-loss exponent α is larger than 2, so the infinite series of equation 10 converges to a finite number.

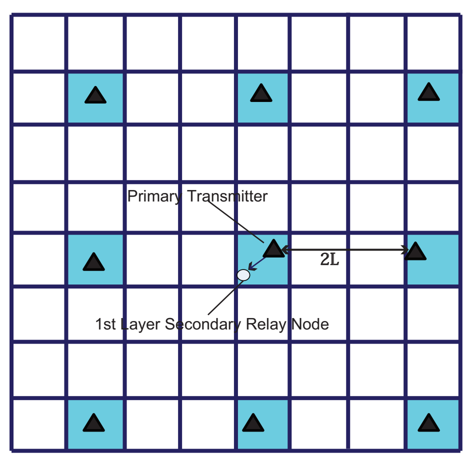

Fig. 7: Interference from the concurrent primary TXs

Next we should bound the sum interference from the concurrent secondary transmitters. In the input process, the preservation region is set around every primary transmitter, and all the secondary nodes that fall into any preservation region are not allowed to transmit the secondary packets. So a minimum distance of l could be guaranteed from all the concurrent secondary transmitters to S_{1,i_1}. And since the secondary network also employs the 9-TDMA scheme, thus I_{sp} is bounded by:

$$\begin{aligned} I_{sp} &= \sum_{k=1,k\neq i}^{N_s} P_s g(\|S_k - S_{1,i_i}\|) \\ &\leq \sum_{k=1,k\neq i}^{\infty} 8k[(3k-2)l]^{-\alpha} P l^{\alpha} \\ &= 8P \sum_{k=1,k\neq i}^{\infty} \frac{k}{(3k-2)^{\alpha}} \\ &= R_2 \end{aligned} \tag{11}$$

Here R_2 is a constant and the reason that equation (11) converges is the same as equation (10).

Combining all the three terms together, we could show that the transmission rate for the input process is bounded by

$$\begin{aligned} R(P_i, S_{1,i_1}) &= \frac{1}{9} log(1 + \frac{P_p g(\|P_i - S_{1,i_1}\|)}{N_0 + I_p + I_{sp}}) \\ &\geq \frac{1}{9} log(1 + \frac{P(\sqrt{2})^{-\alpha}}{R_1 + R^2}) \\ &= C_1 \end{aligned} \tag{12}$$

As for the relay processes, the transmission rate for the secondary relay nodes are regulated by:

$$R(S_{i,u_i}, S_{i+1,u_{i+1}}) = \frac{1}{9} log(1 + \frac{P_s g(\|S_{i,u_i} - S_{i+1,u_{i+1}}\|)}{N_0 + I_s + I_{ps}}), \tag{13}$$

Here, for $1 \leq i \leq h$, S_{i,u_i} is the intermediate relay node for the primary packet. According to the primary scheduling scheme, there is no primary TXs transmitting during the relay processes, so the term I_{ps} equals to 0. And since the secondary network would also employ the 9-TDMA scheme to transmit the primary packets, so similar to equation (10), the sum interference from all the concurrent secondary transmitters to $S_{i+1,u_{i+1}}$ is also bounded by a constant. Thus, the transmission rate for the relay process is also a constant C_2.

Finally, we would consider the output process, during which the h-th layer relay SU would transmit the packet to the primary destination node. The transmission rate is regulated by:

$$R(S_{h,u_h}, P_j) = \frac{1}{9} log(1 + \frac{P_s g(\|S_{h,u_h} - P_j\|)}{N_0 + I_s + I_{ps}}), \tag{14}$$

During this phase, the I_{ps} is again 0 for the same reason as the relay process. In addition, since we have set the preservation region around the primary receivers, so a minimum distance of l could be guaranteed for all the concurrent secondary transmitters to P_j. And as the 9-TDMA scheme is chosen, thus this could lead to a finite I_s. Consequently, the transmission rate in the output process is also a constant C_3.

Since during the input, relay and output process, all the transmitters could support a constant data rate, that finishes the proof. ■

According to the primary scheduling scheme and the results of Lemma 3 10, we can derive the following theorem that counts the throughput of the primary network.

***Theorem** 1:* Under the generalized h-layer hierarchical relay algorithm, the primary network can achieve the following per-node throughput with high probability:

$$\lambda_p = \Theta\left(\frac{1}{h}\right), \tag{15}$$

Proof: During the primary scheduling phases, since every transmitter could transmit the packet with a constant rate, thus we assume that any node could transmit at a rate of R bits per time-slot. Then since each scheduling phase would consume $\frac{1}{h+1}$ fraction of the complete scheduling cycle, thus the per-node throughput during each time slot is degraded by $\frac{1}{h+1}$. And according to Lemma 3, the number of source PUs in any lattice does not exceed $\Theta(1)$ with high probability. Thus the per-node throughput for the primary S-D pair is of $\Theta(\frac{R}{1\cdot(h+1)}) = \Theta(\frac{1}{h})$. ■

B. Delay Performance

In this part, we would give the delay performance for the primary network under the condition that the secondary nodes are moving according to the i.i.d. mobility model and random walk mobility model respectively.

1) The i.i.d mobility model:

In this part, we would evaluate the delay performance of the primary network when the secondary nodes are moving under the i.i.d. mobility model.

***Lemma** 5:* Let a node $\mathcal{T}$ move around in y equal-sized cells according to the i.i.d. mobility model. The position of this node changes once every time-slot. Then the average number of time-slots t it takes for this node to arrive at a certain cell $\mathcal{C}$ from its original position should be

$$E(t) = y. \tag{16}$$

Proof: First we calculate the probability that node $\mathcal{T}$ arrives at cell $\mathcal{C}$ for the first time at the tth time-slot,

$$p(t) = \frac{1}{y}(\frac{y-1}{y})^{t-1}.$$

Then we can calculate the expected value of t by performing an integration,

$$\begin{aligned}E(t) &= \sum_{i=1}^{+\infty} i \times p(i) \\ &= \sum_{i=1}^{+\infty}\sum_{j=i}^{+\infty} p(j) \\ &= \sum_{i=1}^{+\infty} \frac{1}{y}(\frac{y-1}{y})^{i-1}\frac{1}{1-\frac{y-1}{y}} \\ &= \sum_{i=1}^{+\infty}(\frac{y-1}{y})^{i-1} \\ &= y.\end{aligned}$$

■

***Corollary** 1:* Let two nodes $\mathcal{T}$ and $\mathcal{R}$ move around in y equal-sized cells according to the i.i.d. mobility model. The position of these two nodes change once every time-slot. Then the average number of time-slots t' it takes for the two nodes to encounter in a same cell should follow

$$E(t') = y. \tag{17}$$

Proof: Since the nodes are moving according to the i.i.d. mobility model, thus the probability that the two nodes encounter in a same cell is

$$p = C_y^1(\frac{1}{y})^2. \tag{18}$$

So the probability that the two nodes first encounter in the same cell at the t-th time slot is given by:

$$p(t) = \frac{1}{y}(\frac{y-1}{y})^{t-1}.$$

Then the following proof is the same to Lemma 5. ■

By using the Lemma 5, we can calculate the delay performance of the primary network.

***Theorem** 2:* Under the generalized h-layer hierarchical relay algorithm, when the secondary users are moving according to the i.i.d mobility model, the primary network can achieve the following average delay with high probability:

$$D_p = \Theta\left(h^2 n^{(1+\epsilon')/h}\right), \tag{19}$$

Proof: Here we shall analyze the routing process of a random primary packet B_p. In the proposed algorithm, the delay is caused by two kind of events: hopping and roaming. However, each hop consumes only 1 time-slot, which is much shorter than each roaming step, $D_h \ll D_r$, so we just ignore the hopping delay and mainly focus on the roaming delay.

Specifically, we follow the 3 steps below to evaluate the average delay:

- Along its journey, B_p is relayed by h intermediate secondary nodes, and thus experiences h roaming periods;
- During each roaming period, the relay SU is assumed to move around among $n^{(1+\epsilon')/h}$ equal-sized cells, and thus causes an average roaming delay of $n^{(1+\epsilon')/h}$;
- Finally, considering that hops among relay SUs of different layers can only take place in one corresponding primary scheduling phase, we need to add an multiplier of $(h+1)$ to each average roaming delay, i.e., $D_r = (h+1)n^{(1+\epsilon')/h}$.

Sum these up, we derive that the primary network can achieve an average delay performance of

$$\begin{aligned}D_p &= hD_r \\ &= h(h+1)n^{(1+\epsilon')/h} \\ &= \Theta\left(h^2 n^{(1+\epsilon')/h}\right).\end{aligned} \tag{20}$$

■

2) The random walk mobility model:

In this part, the delay performance when the secondary users are moving according to the random walk mobility model would be derived.

First, we would introduce the definition of first hitting time and first return time, which would be used to calculate the delay performance. Consider a random walk on a 2-d torus with random walk step size S, which means the torus is equally divided into $\frac{1}{S} \times \frac{1}{S}$ random walk cells. Then the state of the random walk would correspond to a Markov chain $x(t)$, which take values in X. Here X is defined as $X = \{(x,y) : x, y = 0, 1, ..., \frac{1}{S} - 1$. According to [21], we get the following definitions:

***Definition** 1:* (First hitting time) The first hitting of an arbitrary state $x \subset X$ is defined by:

$$\tau_x = inf\{t \geq 0 : x(t) = x\} \tag{21}$$

***Definition** 2:* (First return time) The first return of an arbitrary state $x \subset X$ is defined by:

$$\tau_x^+ = inf\{t \geq 1 : x(t) = x\}, \tag{22}$$

where x(0)=x.

Since the secondary users are uniformly distributed in their corresponding moving area, thus we could refer to the following lemma from [8] that provides the expectation for the first hitting time and first return time of a single state.

***Lemma** 6:* Consider the random walk on a 2-D torus with step size S, and x is an arbitrary state on this torus, namely an arbitrary random walk cell in the torus. Then the expectation of the first hitting time to enter state x is:

$$E(\tau_x) = \Theta(|logS|/S^2), \tag{23}$$

***Lemma** 7:* Consider a random walk on a 2-D torus with step size S, and x is the initial state of the random walk. Then the expectation of the first return time to re-enter state x is:

$$E(\tau_x^+) = \Theta(1/S^2). \tag{24}$$

From the given lemmas regarding the first hitting time and first return time, the delay of performance of the primary network could be derived as follow:

***Theorem** 3:* When all the secondary users are moving according to the random walk mobility model, with a common random walk step size S within their own moving area. Then under the generalized h-layer hierarchical relay algorithm, the primary network can achieve the following average delay with high probability:

$$D_p = \Theta(h^2 log|S|/S^2), \tag{25}$$

Proof: Suppose the routing path of the primary packet B_p is

$$\mathcal{P}_i \Rightarrow \mathcal{S}_{1,u_1} \Rightarrow \mathcal{S}_{2,u_2} \Rightarrow ... \ \Rightarrow \mathcal{S}_{h,u_h} \Rightarrow \mathcal{P}_j,$$

Where $\mathcal{P}_i$ and $\mathcal{P}_j$ are the primary source node and primary destination node respectively, while $\mathcal{S}_{k,u_k}$, k=1,2,...h are the intermediate secondary relay nodes.

Further assume the location of $\mathcal{P}_j$ satisfies:

$$\mathcal{P}_j \in \mathcal{C}_{h,i_h} \in \mathcal{C}_{h-1,i_{h-1}} \in ... \ \in \mathcal{C}_{1,i_1}.$$

Next we would analyze the roaming process of the secondary mobile relays. For $k = 1, 2, ...h-1$, consider the secondary user $\mathcal{S}_{k,u_k}$ which moves within the $(k-1)$-th layer cell $\mathcal{C}_{k-1,i_{k-1}}$. Until $\mathcal{S}_{k,u_k}$ has entered $\mathcal{C}_{k,i_k}$, i.e., the k-th layer cell that the primary destination node is located, it will relay the packet to the $(k+1)$-th layer secondary user $\mathcal{S}_{k+1,u_{k+1}}$ in the same lattice.

Suppose $\mathcal{C}_{k-1,i_{k-1}}$ is divided into $\frac{1}{S^2}$ random walk cells, where $1 < 1/S^2 < n^{(1+\epsilon')/h}$, and x_0 is the random walk cell that contains $\mathcal{C}_{k,i_k}$. If we denote the probability p to be p=Pr($\mathcal{S}_{k,u_k}$ enters $\mathcal{C}_{k,i_k}$ | $\mathcal{S}_{k,u_k}$ enters x_0), τ_{x_0} and $\tau_{x_0}^j$ to be the first hitting time and j-th return time of x_0. Then the average delay for $\mathcal{S}_{k,u_k}$ to enter $\mathcal{C}_{k,i_k}$, denoted by T_k, should be:

$$\begin{aligned} T_k &= \tau_{x_0}p + (\tau_{x_0} + \tau_{x_0}^1)(1-p)p + ... \\ &\quad + (\tau_{x_0} + \tau_{x_0}^1 + ... + \tau_{x_0}^j)(1-p)^j p + ... \end{aligned} \tag{26}$$

It could be derived that $p = \frac{n^{-(1+\epsilon')/h}}{S^2}$, $E(\tau_{x_0}) = \Theta(|logS|/S^2)$ and $E(\tau_x^j) = \Theta(1/S^2)$. Thus,

$$\begin{aligned} E(T_k) &= E[\tau_{x_0}p + (\tau_{x_0} + \tau_{x_0}^1)(1-p)p + ... \\ &\quad + (\tau_{x_0} + \tau_{x_0}^1 + ... + \tau_{x_0}^j)(1-p)^j p + ...] \\ &= E(\tau_{x_0}) + E(\tau_{x_0}^1)\frac{1-p}{p} \\ &= \Theta(|logS|/S^2) + \Theta(1/S^2)S^2 n^{(1+\epsilon')/h} - \Theta(1/S^2) \\ &= \Theta(|logS|/S^2). \end{aligned} \tag{27}$$

The final step, i.e., the h-th layer SU $\mathcal{S}_{h,u_h}$ transmit the packet to the primary destination node P_j, would also incur the same delay as equation (27). So the primary packet would relay h times and every relay step would take $\frac{1}{h+1}$ fraction of the primary scheduling scheme, thus the average delay for the primary network is:

$$\begin{aligned} D_p &= \Theta(h(h+1)log|S|/S^2) \\ &= \Theta(h^2 log|S|/S^2) \end{aligned} \tag{28}$$

■

VI. Capacity and Delay Performance for the Secondary network

In this section, we would use the similar method as the previous section to evaluate the capacity and delay performance of the secondary network. The secondary network is different from the primary network because the secondary users should access the spectrum opportunistically, i.e., only when the secondary user is outside the current preservation regions can it transmit the secondary packet. Thus, before calculating the capacity and delay scaling of the secondary network, a critical step is to ensure that all the secondary users would have the opportunity to transmit the secondary packets, as shown in the following lemma.

***Lemma** 8:* With high probability, a randomly chosen secondary user is outside the preservation of the concurrent primary transmitters or receivers.

Proof: Consider a randomly chosen secondary user $S_{k,i}$, and denote the probability that $S_{k,i}$ is outside any current preservation region by P. It could be seen that P also equals to the probability that no primary transmitter or receiver resides in 9 lattices that centered at $S_{k,i}$. Since all the primary users are randomly deployed, thus the lower bound of P could be calculated as follow:

$$\begin{aligned} P &\geq (1 - 9n^{-(1+\epsilon')})^n \\ &= \{[1 - 9n^{-(1+\epsilon')}]^{-\frac{1}{9n^{(1+\epsilon')}}}\}^{9n^{-(1+\epsilon')}n} \\ &\to e^{\frac{9n}{n^{1+\epsilon'}}} \\ &\to 1. \end{aligned} \tag{29}$$

Equation 29 holds when n approaches ∞. Consequently, the randomly chosen SU should be outside the preservation regions with high probability. ■

Since we have guaranteed the transmission opportunity for the secondary users, the next step is to calculate the capacity and delay performance of the secondary network.

A. Capacity Performance

Similar to section V-A, first we use the following lemma to estimate the upper bound of secondary users in the lattice.

***Lemma** 9:* According to the h-layer cellular structure, there are at most $\Theta(hn^{\epsilon-\epsilon'})$ SUs in any lattice with high probability.

This lemma could be easily derived using the result of Lemma 2. Next we can prove the following lemma that can guarantee a constant data rate for the secondary transmitters or relay nodes.

***Lemma** 10:* During all the phases of the secondary scheduling scheme, each secondary transmitter or relay node within a lattice could support a constant data rate.

Proof: Similar to the method of lemma 10, we can also divide the secondary scheduling scheme into input, output and relay processes. During each phase, the data rate of the secondary transmitter S_i to its corresponding relay or receiver $S_{D(i)}$ is regulated by

$$R(S_i, S_{D(i)}) = \frac{1}{9}log(1 + \frac{P_s g(\|S_i - S_{D(i)}\|)}{N_0 + I_s + I_{ps}}), \quad (30)$$

Since the secondary network also employ a 9-TDMA scheme, thus I_s could be bounded by a constant. Since the secondary transmitters are outside the preservation region, thus I_{ps} can also be bounded by a constant. Thus, the secondary transmitter could support a constant data rate during all the phases of the secondary scheduling sheme. ■

According to the secondary scheduling scheme and the results of former lemmas, we can derive the following theorem that counts the throughput of the secondary network.

***Theorem** 4:* Under the generalized h-layer hierarchical relay algorithm, the secondary network can achieve the following per-node throughput with high probability:

$$\lambda_s = \Theta\left(h^{-1} n^{\epsilon' - \epsilon}\right). \quad (31)$$

Proof: During the secondary scheduling phases, we assume that any node could transmit at a rate of R bits per time-slot. The per-node throughput during each time slot is degraded by $\frac{1}{h+1}$ since each scheduling phase would consume $\frac{1}{h+1}$ fraction of the complete scheduling cycle. Then according to Lemma 9, the number of source SUs in any lattice does not exceed $\Theta(hn^{\epsilon-\epsilon'})$ with high probability. Thus the per-node throughput for the primary S-D pair is of $\lambda_s = \Theta\left(h^{-1} n^{\epsilon'-\epsilon}\right)$. ■

B. Delay Performance

In this part, the delay performance of the secondary network is derived. From the secondary scheduling scheme, because different layer destination SU would experience different number of relays, thus the delay performance of the secondary users should depend on which layer the destination SU is. Similar to section V-B, we would consider the case that secondary users are moving according to the i.i.d. mobility model and random walk mobility model separately.

1) The i.i.d mobility model:

***Theorem** 5:* Under the generalized h-layer hierarchical relay algorithm, when the secondary users are moving according to the i.i.d. mobility model, the CRN can achieve the following average delay with high probability:

$$D_{s,k_d} = \Theta\left(h(k_d - 1)n^{(1+\epsilon')/h} + hn^{(h-k_d+1)(1+\epsilon')/h}\right), \quad (32)$$

for the secondary network, where k_d is the layer of the destination SU, here $k_d = 1, 2, ..., h$.

Proof: The delay performance of the secondary network is somewhat different from that of the primary network, because secondary packets with different destinations should follow different routing processes. We can divide the routing process of a k_dth-layer secondary packet B_{s,k_d} 2 parts:

In the first part, B_{s,k_d} is relayed along intermediate SUs in the same way as a primary packet, from a 1st-layer SU to a k_dth-layer SU. Along this journey, B_{s,k_d} experiences $(k_d - 1)$ roaming periods. So this part causes an average delay of $(h+1)(k_d - 1)n^{(1+\epsilon')/h}$.

In the second part, B_{s,k_d} is carried by a k_dth-layer relay SU $\mathcal{S}_{k_d,u_{k_d}}$. $\mathcal{S}_{k_d,u_{k_d}}$ moves around among $N_{h,k_d-1} = n^{(h-k_d+1)(1+\epsilon')/h}$ lattices until it encounters $\mathcal{S}_{k_d,j}$ in a same lattice. According to corollary 1, this part is accompanied by an average delay of $(h+1)n^{(h-k_d+1)(1+\epsilon')/h}$.

Add the two parts up, we derive that a k_dth-layer secondary packet can achieve an average delay performance of

$$\begin{aligned} D_{s,k_d} &= (h+1)(k_d-1)n^{(1+\epsilon')/h} + (h+1)n^{(h-k_d+1)(1+\epsilon')/h} \\ &= \Theta\left(h(k_d-1)n^{(1+\epsilon')/h} + hn^{(h-k_d+1)(1+\epsilon')/h}\right). \end{aligned} \quad (33)$$

This finishes the proof. ■

2) The random walk mobility model:

***Theorem** 6:* Under the generalized h-layer hierarchical relay algorithm, when the secondary users are moving according to the random walk mobility model with random walk step size S, the CRN can achieve the following average delay with high probability:

$$D_{s,k_d} = \Theta\left(h(k_d - 1)|logS|/S^2 + hn^{(h-k_d+1)(1+\epsilon')/h}\right), \quad (34)$$

for the secondary network, when $k_d = o(h)$, and

$$D_{s,k_d} = \Theta\left(hk_d|logS|/S^2\right). \quad (35)$$

for the secondary network when $k_d = \Theta(h)$.

Proof: For a random k_dth-layer secondary packet B_{s,k_d}, we can divide the routing process of B_{s,k_d} into 2 parts:

In the first part, B_{s,k_d} is relayed along intermediate SUs in the same way as a primary packet, from a 1st-layer SU to a k_dth-layer SU. Along this journey, B_{s,k_d} experiences $(k_d - 1)$ roaming periods. Recall the result of equation (27), each roaming step would consume an average delay of $\Theta(|logS|/S^2)$. So this part causes an average delay of $D_1 = (h+1)(k_d - 1)|logS|/S^2$.

In the second part, the packet B_{s,k_d} is carried by $\mathcal{S}_{k_d,u_{k_d}}$, while $\mathcal{S}_{k_d,u_{k_d}}$ and the secondary destination node $\mathcal{S}_{k_d,j}$ both move within the $k_d - 1$-th layer cell $\mathcal{C}_{k_d-1,j_{k_d-1}}$, according to the random walk mobility. And $\mathcal{S}_{k_d,u_{k_d}}$ would relay the packet to $\mathcal{S}_{k_d,j}$ when they have come into the same lattice. So the delay of the second part should be the time for $\mathcal{S}_{k_d,u_{k_d}}$ and $\mathcal{S}_{k_d,j}$ to meet in the same lattice, when they both move according to random walk on a $\frac{1}{S} \times \frac{1}{S}$ 2-D torus. We denote this delay by D_2.

To make $\mathcal{S}_{k_d,u_{k_d}}$ and $\mathcal{S}_{k_d,j}$ encounter in the same lattice, they need to come into the same random walk cell first. If they come within the same lattice, $\mathcal{S}_{k_d,u_{k_d}}$ would relay the packet to $\mathcal{S}_{k_d,j}$. However, if they are in the same random walk cell but not the same lattice, they need to continue roaming in $\mathcal{C}_{k_d-1,j_{k_d-1}}$ until they come to another same k_d-th layer cell. At this time, if they are in the same lattice, they could relay the packet, or the previous process would continue.

Thus, the delay for this process should be:

$$D_2 = \tau_s p + (\tau_s + \tau_{I_1})(1-p)p + ... + (\tau_s + \tau_{I_1} + ... + \tau_{I_j})(1-p)^j p + ... \quad (36)$$

In equation 36, τ_s denotes the time for the two nodes to enter in a same k_d-th layer cell, both start from two random positions. In addition, τ_{I_j} denotes the j-th successive inter-meeting time of $\mathcal{S}_{k_d,u_{k_d}}$ and $\mathcal{S}_{k_d,j}$ to enter into a same random walk cell. And p is the probability that the two SUs come into the same lattice when they move into the same random walk cell.

For a random walk on the 2-D torus, the expectation of τ_s is on the same order as the first hitting time for an arbitrary state. Thus from lemma 6, $E(\tau_s) = \Theta\left(|logS|/S^2\right)$. While from lemma 7, $E(\tau_{I_j}) = \Theta\left(1/S^2\right)$, for $j = 1, 2, ...$

Moreover, since the area of $\mathcal{C}_{k_d-1,j_{k_d-1}}$ is $n^{-(k_d-1)(1+\epsilon^{'})/h}$, so the area of the random walk cell in $\mathcal{C}_{k_d-1,j_{k_d-1}}$ is $n^{-(k_d-1)(1+\epsilon^{'})/h}/\frac{1}{S^2}$. Consequently, the number of lattices within a random walk cell is $N = S^2 n^{-(k_d-1-h)(1+\epsilon^{'})/h}$, and the probability that the two SUs enter the same lattice when the enter the same random walk cell is:

$$\begin{aligned} p &= C_N^1 (\frac{1}{N})^2 \\ &= \frac{1}{N} \\ &= \frac{n^{(k_d-1-h)(1+\epsilon^{'})/h}}{S^2} \end{aligned} \quad (37)$$

.

So if we take the expectation on both sides of equation 36, we could get the expectation of D_2:

$$\begin{aligned} E(D_2) = E(\tau_s)&[p + (1-p)p + ... + (1-p)^i + ...] + \\ &E(\tau_{I_j})[(1-p)p + 2(1-p)^2 p + ... + j(1-p)^j p + ...] \\ &= E(\tau_s) + \frac{1-p}{p} E(\tau_{I_j}). \end{aligned} \quad (38)$$

Substitute the values of each term into equation 38, we would derive that:

$$\begin{aligned} E(D_2) &= E(\tau_s) + \frac{1-p}{p} E(\tau_{I_j}) \\ &= \Theta\left(log|S|/S^2\right) + \Theta\left(1/S^2\right) S^2 n^{-(k_d-1-h)(1+\epsilon^{'})/h} \\ &- \Theta\left(1/S^2\right). \\ &= \begin{cases} \Theta\left(n^{(h+1-k_d)(1+\epsilon^{'})/h}\right), & \text{if } k_d = o(h), \\ \Theta\left(|logS|/S^2\right), & \text{if } k_d = \Theta(h). \end{cases} \end{aligned} \quad (39)$$

Moreover, since the hop among $\mathcal{S}_{k_d,u_{k_d}}$ and $\mathcal{S}_{k_d,j}$ takes 1 of the (h+1) phases of the scheduling scheme, so we would multiply $(h+1)$ to D_2 to get the delay of the second part.

Summing D_1 and $(h+1)D_2$ together, the delay for the secondary network could be derived.

This completes the proof. ■

VII. Optimal Condition

Now that we have obtained the expression of the capacity and delay performance under the generalized h-layer hierarchical relay algorithm, we shall compute the optimal condition of the CRN by assigning an appropriate number of hierarchies to our model. Since the intuition for dividing the secondary users into several layers is to utilize the hierarchical relay of the SUs to reduce the delay while maintaining a good capacity performance, so we choose the optimal number of layer $h = \log_2 n$. And from this choice, we can get the following theorem:

Theorem 7: With $h = \log_2 n$, the primary network may achieve a per-node throughput of

$$\lambda_p = \Omega(n^{-\delta'}), \quad (40)$$

with an average delay of

$$D_p = O(n^{\delta''}), \quad (41)$$

where δ' and δ'' can be any positive values.

Proof: Let $h = \log_2 n$, then according to *Equation* (31), the capacity of the primary network should be

$$\lambda_p = \Theta\left(\frac{1}{h}\right) = \Omega\left(n^{-\delta'}\right).$$

On the other hand, according to Equation (19), the corresponding delay under i.i.d. mobility model should be

$$\begin{aligned} D_p &= \Theta\left(h^2 n^{(1+\epsilon')/h}\right) \\ &= \Theta\left((\log_2 n)^2 \times 2^{1+\epsilon'}\right) \\ &= O\left(n^{\delta''}\right). \end{aligned}$$

When the SUs are moving according to the random walk mobility model, we consider the condition for choosing the minimum random walk step size, i.e., $\frac{1}{S^2} = n^{(1+\epsilon^{'})/h}$, then the delay performance is:

$$\begin{aligned} D_p &= \Theta\left(h^2 |logS|/S^2\right) \\ &= \Theta\left(h^2 \times \frac{1+\epsilon^{'}}{h} \times \log(n) \times n^{(1+\epsilon^{'})/h}\right) \\ &= \Theta((\log n)^2) \\ &= O\left(n^{\delta''}\right). \end{aligned}$$

■

Theorem 8: With $h = \log_2 n$, the secondary network could achieve a per-node throughput of

$$\lambda_s = \Omega(n^{-\epsilon''}), \quad (42)$$

with an average delay of

$$D_s = O(n^{1+\epsilon'''}), \quad (43)$$

when $k_d = o(h)$, and

$$D_s = O(n^{\delta''}), \quad (44)$$

when $k_d = \Theta(h)$.

Here δ'' can be any positive value, and $\epsilon''' > \epsilon' > 0$,$\epsilon'' > \epsilon - \epsilon' > 0$.

VIII. Discussion

After deriving all the results regarding our CRN, we would have following discussion over the hierarchical relay algorithm: (1) The choice of $h = \log_2 n$ is indeed optimal. Since if we choose h to be a constant, then the layer of the secondary users would be constant, so from Equation (19), it can be noticed that we cannot achieve an arbitrary small delay performance for the primary users. And if we choose $h = \Omega(\log_2 n)$, this choice may reduce the capacity of the primary network according to equation (31). Consequently, $h = \log_2 n$ is a reasonable number layer, which could reduce the delay while maintaining a good capacity performance.

(2) From the optimal condition, we can see that for the primary network, the hierarchical relay algorithm could help the primary users achieving a near-optimal performance both in the capacity and delay performance. This suggest that the hierarchical cooperation could significantly increase the overall performance of the primary network.

(3) From the optimal condition of the secondary network, the secondary network could also achieve a near-optimal capacity performance, as long as ϵ and ϵ' are chosen to be close enough. As to the delay performance, we can notice that there is gap between the condition when the layer of destination k_d is chosen to $o(h)$ and when k_d is $\Theta(h)$. This gap is mainly introduced by the last step of the relay process for the secondary packets. Since we have constrained the transmission range to be within one lattice, so the smaller k_d is, the larger number of lattices would reside in the $(k_d - 1)$-th layer cell where the destination SU resides. Thus, the probability for the k_d-th layer relay node to encounter with the destination node will decrease, which could lead to a larger delay. As to the case when k_d is larger $\Theta(h)$, the delay performance of the secondary users could also be near-optimal.

(4) Compared to the results of [16], which achieves the same results as the stand-alone network for both the primary users and secondary users, our CRN could have better capacity and delay performance. The increase in capacity is mainly due to the cooperation between the PU and SU, as well as the choice for the transmission range, i.e., the area of one lattice. And the low delay is mainly due to the hierarchical relay of the secondary users, so that the packet could approach the destination in a step-wise fashion.

(5) Compared to the results of [19], which requires the number of SUs to be $\Theta(n^2)$, our CRN could achieve the near-optimal performance with less supportive SUs, i.e., $\Theta(n^{1+\epsilon})$. And this improvement is crucial for deploying the cognitive radio network which make the structure easier to implement.

(6) Possible future work: 1. In our network model, all the SUs are equally and strictly divided into h-layers, which aims to regulate their speed or moving range. This regulation is intuitive, as in our real world, nearly all the mobile nodes would have different speed, or moving area. Such as the difference between airplane, cars, bicycles and pedestrians. However, our assumptions on partitioning the network into these regularly placed layered cells may not be representative enough. So the next step we could study the situation where all the nodes would possess different moving speed or moving area, which could be follow a certain distribution. Such case could characterize our mobility model in this work, but of course more technically challenging. 2. In cognitive radio network, a very critical consideration is how to control the interference from the secondary users to primary users, so that both network could achieve a reasonable performance. Up to know, most works considering the capacity and delay scaling solve this problem by defining the preservation regions, which has the similar idea with our work. However, is the preservation region the only way to solve this problem, is there any more effective way to solve this problem? So this question might be another point consider. Maybe we can introduce some other elements like base station to the CRN, so the relay could be assisted by the base stations.

References

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

[2] M. Franceschetti, O. Dousse, D. N. C. 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, 2007.

[3] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Transactions on Networking*, vol. 10, pp. 477-486, 2002.

[4] 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, 2005.

[5] S. Toumpis and A. J. Goldsmith, "Large Wireless Networks under Fading, Mobility, and Delay Constraints," *in Proceedings of IEEE INFOCOM, 2004*, pp. 609-619.

[6] X. Lin and N. B. Shroff, "The fundamental capacity-delay tradeoff in large mobile wireless networks," *Presented at the Third Annual Mediterranean Ad Hoc Networking Workshop Bodrum, Turkey, 2004.*

[7] A. E. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-delay trade-off in wireless networks," *in Proceedings of IEEE INFOCOM, 2004*, pp. 464-475.

[8] X. Lin, G. Sharma, R. Mazumdar, and N. Shroff, "Degenerate delay-capacity trade-offs in ad hoc networks with Brownian mobility," *IEEE Transactions on Information Theory*, vol. 14, no. SI, pp. 2777-2784, June 2006.

[9] G. Sharma and R. Mazumdar, "Scaling laws for capacity and delay in wireless ad hoc networks with random mobility," *in Proceedings of IEEE ICC, 2004*, vol. 7, pp. 3869-3873.

[10] G. Sharma and R. Mazumdar, "Delay and capacity tradeoff in wireless ad hoc networks with random way-point mobility," *Dept. Elect. Comput. Eng., Purdue Univ., West Lafayette, IN, 2005.*

[11] L. Ying, S. Yang, R. Srikant, "Optimal Delay-Throughput Trade-Offs in Mobile Ad Hoc Networks," *IEEE Transactions on Information Theory*, vol. 54, no. 9, pp. 4119-4143, 2008.

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

[13] G. Sharma, R. Mazumdar, and N. B. Shroff, "Delay and Capacity Trade-offs in Mobile Ad Hoc Networks: A Global Perspective," *IEEE Transactions on Networking*, vol. 15, no. 5, pp. 981-992, October 2007.

[14] Federal Communications Commission Spectrum Policy Task Force, "Report of the Spectrum Efficiency Working Group," FCC, Tech. Rep., 2002.

[15] Y. Shi, C. Jiang, Y.T. Hou, and S. Kompella, "On Capacity Scaling Law of Cognitive Radio Ad Hoc Networks," *submitted to IEEE Transactions on Mobile Computing*, 2007.

[16] S. Jeon, N. Devroye, M. Vu, S. Chung, and V. Tarokh, "Cognitive Networks Achieve Throughput Scaling of A Homogeneous Network," *submitted to IEEE Transactions on Information Theory*, 2009. Available at http://arxiv.org/abs/0801.0938.

[17] C. Li and H. Dai, "On the throughput scaling of cognitive radio ad hoc networks," in IEEE INFOCOM Mini conference, 2011.

[18] Wentao Huang and Xinbing Wang, "Throughput and Delay Scaling of General Cognitive Networks," *in Proceedings of IEEE INFOCOM, 2011*.

[19] L. Gao, R. Zhang, C. Yin, and S. Cui, "Throughput and Delay Scaling in Supportive Two-tier Networks," *IEEE Transactions on Networking*, pp. 2833 - 2837, 2009. Available at http://arxiv.org/abs/0905.3407.

[20] M. Ji, Z. Wang, H. R. Sadjadpour, and J. J. Garcia-Luna-Aceves, "The Capacity of Ad Hoc Networks with Heterogeneous Traffic Using Cooperation," *in Proceddings of IEEE INFOCOM, 2010*.

[21] D. Aldous and J. Fill, Reversible Markov chains and random walks on graphs, Monograph in preparation, 2002 [Online]. Available: http:// stat-www.berkeley.edu/users/aldous/RWG/book.html