Throughput and Delay Scaling of Cognitive Radio Networks with Heterogeneous Mobile Users

Pengyuan Du¹, Mario Gerla¹, and Xinbing Wang²

¹ Department of Computer Science, University of California, Los Angeles, USA
² School of Electronic, Info. & Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China

Abstract. We study the throughput and delay scaling laws of cognitive radio networks (CRN) with primary and secondary users operating at the same time, space and sharing the spectrum. Both the primary and secondary users are initially randomly and uniformly distributed, then move according to a General Heterogeneous Speed-restricted Mobility (GHSM) model. However, the primary users have higher priority to access the spectrum while the secondary users should access opportunistically. In GHSM model, we define \( (h+1) \) heterogeneous moving patterns using a universal set \( \mathcal{T} = \{ T_i \mid 0 \leq i \leq h, A_i = n^{-\frac{1}{\beta}} \} \), where \( A_i \) determines the moving area of each pattern. The set of primary (secondary) moving patterns \( \mathcal{T}^{(p)} (\mathcal{T}^{(s)}) \) is a subset of \( \mathcal{T} \) which is randomly and independently selected. We assign \( n \) nodes, where \( \beta > 1 \) primary (secondary) nodes to each moving pattern, and their initial positions are subject to a poisson point process. Additionally, in GHSM model with \( \mathcal{T}^{(p)} \sim \mathcal{T}^{(s)} = \Theta^3(h) = \Theta(\log n) \). By proposing a cooperative routing strategy, we fully utilize the mobility heterogeneity of primary and secondary users to achieve near-optimal throughput and delay performance of order \( \Theta(poly \log n) \) when \( \gamma_0 \geq \beta \). In other cases, our transmission scheme shows advantages over [7] in delay performance of the primary network and over [13] in delay of the secondary network.

Keywords: Cognitive Radio Networks, Throughput Capacity, Delay, Heterogeneous Mobility, Cooperative Routing

1 Introduction

Recently, there has been more stress over the already-crowded radio spectrum since wireless applications demand ever more bandwidth. Driven by such demands, many researches have been conducted on cognitive radio networks (CRN) because of its efficient usage of spectrum. Initiated by Gupta and Kumar’s work [5], the fundamental performance scaling laws of CRN raised great interests in the networking research community. Vu et al. considered the throughput scaling law for a single-hop cognitive radio network and obtained a linear scaling law for secondary network in [10] and [11]. In [6], Jeon et al. studied the throughput scaling of a cognitive network under general environment and showed that both primary and secondary networks can achieve the same throughput scaling law as a stand-alone wireless network while the secondary network may suffer from a finite outage probability. Yin et al. in [14] developed the throughput scaling laws under a similar assumption and adopted transmission protocols that could guarantee zero outage probability for secondary network.

In contrast to the above-mentioned static scenarios, capacity and delay performance in wireless mobile networks is yet another topic that has been explored. In [4], the author showed that the mobile network could achieve the optimal throughput of \( \Theta(1) \) under the 2-hop relay algorithm at the cost of \( \Theta(n) \) delay per packet. Other mobility models have also been studied subsequently, including the i.i.d. mobility model [8], random way-point mobility model [9], random walk mobility model [15], and restricted mobility model [3][12].

Moreover, recent research showed that mobility in CRN could bring even more benefits. In particular, the movement secondary users would facilitate possible cooperations between the two coexisting networks.

³ The following asymptotic notations are used throughout this paper. Given non-negative functions \( f(n) \) and \( g(n) \):
1. \( f(n) = \omega(g(n)) \) means that \( \lim_{n \to \infty} \frac{g(n)}{f(n)} = 0 \).
2. \( f(n) = o(g(n)) \) means that \( g(n) = \omega(f(n)) \).
3. \( f(n) = O(g(n)) \) means that there exists a constant \( c_1 \) and integer \( N \) such that \( f(n) \leq c_1 g(n) \) for \( n > N \).
4. \( f(n) = \Theta(g(n)) \) means that for two constants \( 0 < c_2 < c_3 \), \( c_2 g(n) \leq f(n) \leq c_3 g(n) \) for sufficiently large \( n \).
5. \( f(n) \sim g(n) \) means that \( \lim_{n \to \infty} \frac{f(n)}{g(n)} = 1 \).
and improve their performance scaling laws. In [2], Gao et al. proposed a supportive two-tier network where secondary users are willing to relay packets for the primary users to improve throughput and delay scaling of the primary network. Then, Wang et al. [13] derived a cooperation scheme which achieves near-optimal capacity and delay scaling for the static primary network, but with less supportive mobile secondary users. This is achieved by dividing the secondary users into \( h \) different layers and the mobile secondary users of different layers are associated with different moving areas. However, in order to regulate the moving area of different layers, this network model requires a strict cell partition scheme which arbitrarily partitioned the whole network into \( h \) different layer cells. Li et al utilize a similar approach to improve capacity scalings for secondary network in a more general and flexible hierarchical mobility model [7].

Based on these works, we go one step further by extending the network model to a more general situation where both primary and secondary users possess different moving ability. This is motivated by the fact that even though the primary users have priority to access the spectrum resources, their moving pattern should not be any different compared to the secondary users. Therefore we first define a universal set \( \mathcal{T} \), which includes \((h + 1)\) types of moving pattern. For each moving pattern, a moving area of \( n^{-\chi} \) is determined where \( \chi \) is a random variable following the discrete uniform distribution with \( h + 1 \) different values, ranging from 0 to \( \chi_0 \). Then we construct the set of primary moving patterns \( \mathcal{T}^{(p)} \) by choosing randomly and independently from the universal set with probability \( \frac{1}{2} \). The left elements in \( \mathcal{T} \) form set \( \mathcal{T}^{(s)} \). We assign \( n^{\beta} \) nodes for each primary(secondary) moving pattern. The key questions we want to explore under this General Heterogeneous Speed-restricted Mobility (GHSM) model include: (1) with more primary nodes in the network, how to exploit the mobility of primary and secondary users to support their transmissions? (2) whether or not we can still achieve near-optimal capacity and delay performance? (3) we divide all moving patterns into two sets, so there may exist moving ability gaps among primary or secondary users. What impacts does such phenomenon have on the throughput and delay scaling laws?

Our main contributions are as follows:

– We present a GHSM model with both mobile primary users and mobile secondary users, and moving abilities of all users are determined in a general and representative way.
– A cooperative routing scheme is proposed for primary transmissions which guarantees the primary users better delay performance compared with the scheme in [7].
– As for the secondary network, we show that the moving ability gap among secondary users will not degrade throughput and delay performance in order sense, i.e., the performance of secondary network is no worse than the performance in [7] [13].

The rest of this paper is organized as follows. In Section 2 we introduce the network model and definitions. In Section 3 we present the routing and scheduling schemes. In Section 4 and 5, we analyze the throughput capacity and delay of primary and secondary networks, respectively. In Chapter 6, we discuss the insights of our results. Finally we conclude this paper in Section 7.

2 System Model

Throughout this paper we denote the probability of an event \( E \) as \( \mathbb{P}(E) \) and we mainly deal with events which take place with high probability (w.h.p.), or with probability 1 as the node density tends to infinity.

2.1 Network Geometry

We consider the network extension to be a unit torus where primary nodes and secondary nodes co-exist. In the mobile cognitive network, both the primary and secondary networks consist of mobile users. The primary nodes are randomly distributed according to Poisson Point Process (P.P.P.) of density \( N = (h_p + 1)n \), while the secondary nodes are of density \( M = (h_s + 1)m = (h_s + 1)n^{\beta} \), with \( h_p + h_s = h = \Theta(\log n) \). These nodes move according to the heterogeneous speed-restricted mobility model which will be introduced later. All nodes are randomly grouped into source-destination (S-D) pairs.

2.2 Communication Model

For the wireless channel in this work, we assume the channel gain depends only on the distance between the transmitter and its receiver. Therefore the normalized channel power gain \( g(d) \) is given by

\[
g(d) = d^{-\delta},
\]
where $d$ denotes the distance between the transmitter (Tx) and its receiver (Rx) and $\delta > 2$ is the path loss exponent.

To determine the transmission rate of each network, we assume that each transmission deploy a scheme that can achieve the additive white Gaussian noise (AWGN) channel capacity. For a given signal to interference and noise ratio (SINR), this capacity is given by the well known formula $R = \log(1+\text{SINR})$ bps/Hz assuming the additive interference is also white, Gaussian, and independent from the noise and signal.

We now characterize the rates achieved by the primary and secondary transmit pairs. Suppose that $N_p$ primary pairs and $N_s$ secondary pairs communicate simultaneously. Before proceeding with a detailed description, let us define the notations used in the paper, given by Table 1. Thus the $i$-th primary Tx-Rx pair can communicate at a rate of

$$R_p^i = \log(1 + \frac{P_{p}^i g(\|X_{p,tx}^i - X_{p,rx}^i\|)}{N_0 + I_{p}^i + I_{sp}^i}),$$

(2)

where $\|\cdot\|$ denotes the Euclidean norm of a vector. Similarly, the $j$-th secondary pair can communicate at a rate of

$$R_s^j = \log(1 + \frac{P_{s}^j g(\|X_{s,tx}^j - X_{s,rx}^j\|)}{N_0 + I_{s}^j + I_{ps}^j}).$$

(3)

### 2.3 Mobility Model

The primary and secondary users would move under a GHSM model modified based on the HSRM model in [7]. At the beginning, all users are uniformly and randomly distributed over the network. Then they would move within their own circular area centered at the initial position, according to a i.i.d. mobility model. The radius $R$ denotes the restricted speed of different users, and the node positions would be totally reshuffled in each moving area from one time slot to another.

The moving area of primary and secondary mobile users is set to be $n^{-\chi}$, where $\chi$ is a variable following the discrete uniform distribution of $h + 1$ different values: $\chi = \frac{\chi_0}{h}, \frac{2\chi_0}{h}, ..., \frac{(h-1)\chi_0}{h}, \chi_0$. Here $\chi_0$ is a random positive value and $h = \Theta(\log n)$. Each moving area $A_i = n^{-\chi}$ characterizes an $i$-th type of moving pattern $T_i$, and $\mathcal{T} = \{T_i|0 \leq i \leq h\}$. Different from the HSRM in [7], we have both primary and secondary mobile users. Specifically, we randomly choose $h_p$ types of moving pattern from $\mathcal{T}$ which form $\mathcal{T}^{(p)}$ and assign $n$ primary nodes for each moving pattern $T_i \in \mathcal{T}^{(p)}$. The other $h_s$ moving patterns form $\mathcal{T}^{(s)}$ and we assign $n^{\beta}$ secondary nodes for each $T_i \in \mathcal{T}^{(s)}$.

Similar to HSRM in [7], larger $\chi_0$ leads to larger difference between moving area of different users, and larger $\beta$ corresponds to larger $T_i$. Thus, $\chi_0$ and $h$ determine the mobility heterogeneity of our mobility model.

<table>
<thead>
<tr>
<th>$P_{p}^i$</th>
<th>Transmit power of the $i$-th primary pair</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{s}^j$</td>
<td>Transmit power of the $j$-th secondary pair</td>
</tr>
<tr>
<td>$N_0$</td>
<td>Thermal noise power</td>
</tr>
<tr>
<td>$X_{p,tx}^i$</td>
<td>Tx location of the $i$-th primary pair</td>
</tr>
<tr>
<td>$X_{p,rx}^i$</td>
<td>Rx location of the $i$-th primary pair</td>
</tr>
<tr>
<td>$X_{s,tx}^j$</td>
<td>Tx location of the $j$-th secondary pair</td>
</tr>
<tr>
<td>$X_{s,rx}^j$</td>
<td>Rx location of the $j$-th secondary pair</td>
</tr>
<tr>
<td>$I_{p}^i$</td>
<td>Interference power from the primary Txs to the Rx of the $i$-th primary pair</td>
</tr>
<tr>
<td>$I_{sp}^i$</td>
<td>Interference power from the secondary Txs to the Rx of the $i$-th primary pair</td>
</tr>
<tr>
<td>$I_{s}^j$</td>
<td>Interference power from the secondary Txs to the Rx of the $j$-th secondary pair</td>
</tr>
<tr>
<td>$I_{ps}^j$</td>
<td>Interference power from the primary Txs to the Rx of the $j$-th secondary pair</td>
</tr>
<tr>
<td>$R_p^i$</td>
<td>Rate of the $i$-th primary pair</td>
</tr>
<tr>
<td>$R_s^j$</td>
<td>Rate of the $j$-th secondary pair</td>
</tr>
</tbody>
</table>

**Table 1.** Definition of symbols related to achievable rates for each primary and secondary transmit pair
Moreover, we denote the $q$-th primary user of $T_i \in \mathcal{T}^{(p)}$ as $P_i^q$ and its initial position as $X_{p,i}^q$, where $0 \leq i \leq h$ and $1 \leq q \leq n$. The $r$-th secondary user of $T_j \in \mathcal{T}^{(s)}$ as $S_j^r$ and its initial position as $X_{s,j}^r$, where $0 \leq j \leq h$ and $1 \leq r \leq n$. Under our HSRM, $\|P_i^q - X_{p,i}^q\| \leq R_i$, where $R_i = \sqrt{\frac{A_i}{\pi}} = \Theta(n^{\frac{1}{3}})$, and $\|S_j^r - X_{s,j}^r\| \leq R_j$.

### 2.4 Capacity Definition

Throughout the paper, the achievable per-node throughput of the primary and secondary networks is defined the average data rate that each source node can transmit to destination w.h.p. under a particular scheduling and routing scheme.

### 2.5 Fluid Model and Delay

In this paper, we use a fluid model [1] to study the delay performance for the primary and secondary networks. Specifically, we divide each time slot into several packet slots, and the size of the packets will be scaled down to arbitrarily small with respect to the node density $n$ (or $m$) in the networks.

### 3 Network protocols

In this chapter, we first present the routing and scheduling strategy in our mobile cognitive network with GHSM. Then we analyze the throughput and delay performance of the primary and secondary networks. At last we will discuss how mobility heterogeneity of primary and secondary users can affect the throughput and delay scaling of mobile cognitive networks.

#### 3.1 Primary Network Routing Scheme

In our scheme, both the primary and secondary users are moving under the GHSM, so the primary packets would not need to be relayed only by secondary users. Instead, primary packets would be relayed progressively to reach the destination by both primary and secondary users. We still divide the unit area into primary and secondary cells with $a_p = \frac{2 \log n}{n}$ and $a_s = \frac{2 \log m}{m}$, then the primary packets would be routed in these two network grids. Since $\mathcal{T}^{(p)}$ and $\mathcal{T}^{(s)}$ are randomly constructed, it is clear that the chain of relay nodes for a particular primary S-D pair consist of two interlaced primary and secondary relay chains.

We first derive the Critical Relay Type as defined in [7] for the primary nodes, since when the moving pattern is a large type, it is possible that the moving area of a primary node tends to be within a primary cell which will not help exploit the advantage of mobility.

**Definition 1.** The critical relay type of primary nodes $h^*_p$ is denoted by:

$$h^*_p \triangleq \max\{i \mid T_i \in \mathcal{T}^{(p)}, R_i \geq 2\sqrt{2} \sqrt{\pi n}\}. \quad (4)$$

From the above definition, we can calculate the critical relay type as follows:

$$h^*_p = \begin{cases} h, & \text{if } \chi_0 < 1, \\ \left(\frac{1}{\chi_0} - \log \log n + \log 16\pi \right)h, & \text{if } \chi_0 \geq 1. \end{cases} \quad (5)$$

Here $|x| = \max\{n \in Z \mid n \leq x\}$, and $|x|_p = \max\{n \in Z \mid T_n \in \mathcal{T}^{(p)}, n \leq x\}$. It can be seen that $h^*_p = \Theta(h)$.

Now we introduce the main idea of our primary routing scheme. For a particular source node $P_i^{k_i}$, it would first transmit packets to a $T_0$ type relay node since this relay node has access to any primary cells within the unit area and thus can deliver the packets to the destination $P_j^{k_j}$. Then the packets would be forwarded to the next type of relay node whose moving area is smaller but covers the moving area of $P_j^{k_j}$. So on forth, the packets would be delivered to a relay node of the same type of moving pattern as $P_j^{k_j}$, and when they meet the packets would approach the destination. We note that since $\mathcal{T}^{(p)}$ and $\mathcal{T}^{(s)}$ are just part of $\mathcal{T}$, most packets are propagated through both primary and secondary networks and there are four types of delivery case:
With the heterogeneity factor $\chi_0$ and $h = \Theta(\log n)$, the number of eligible relays in each step of Algorithm 1 is larger than 1, w.h.p.

**Proof.** We denote $C_r(x)$ as the circle with radius $r$ centering at $x$. We consider the worse condition as $R^{p-p}$, since the node density of secondary users for each moving pattern is larger than the primary users.
From Algorithm 1, we can see that eligible relays for $T_{k+1}$ should reside in the area of $C_{R_k}(X_{p,k}^{h_k}) \cap C_{R_{k+1}-\sqrt{2\pi p}}(X_{p,j}^{h_j}) > \frac{\pi}{3}(R_{k+1} - \sqrt{2\pi p})^2$. Using properties of Poisson distribution, we have

$$P(\text{the number of eligible relays of } T_{(k+1)} \text{ is } 0) = e^{-\lambda} \left| \lambda \sim \frac{\pi}{3} R_k^2 \right|, R_0 \sim e^{-n \log log n / \log n} \sim \begin{cases} e^{-n^{1-\chi_0}} & \text{if } \chi_0 < 1, \\ e^{-n^{1-\chi_0}} & \text{if } \chi_0 \geq 1, \\ \end{cases}$$

(6)

Then we have $h_s^* = \Theta(h)$. If we denote $h_s^* = \min(j, h_s^*)$, the relay process of $B_s^u$ would take up to $K_{h_s^*} + 2$ steps, for $T_{h_s^*} = T_{s,K_{h_s^*}}$ (the mapping from $T$ to $T^{(s)}$ is shown in Figure 2). We note that a difference in the secondary relay algorithm is that, when $T_0 \notin T^{(s)}$, $T_{s,0}$ nodes are used to relay a secondary packet towards its destination. And we notice that the moving area of $T_{s,0}$ in this case only covers part of the unit torus, so we keep the packet transmitting among $T_{s,0}$ nodes with power of $P_s \alpha_s^{d/2}$ until a specific relay node of $T_{s,0}$ covers the initial position of the destination. We denote each step of this relay process as $S_{s,0} \rightarrow S_{s,0}$. For other steps of the whole relay process, the operation rules are similar to $P_s \rightarrow S_{s,0}$, and we denote them as $S \rightarrow S_{s,0}$. The secondary relay algorithm is shown in Algorithm 2. Similar to Lemma 1, we can verify the feasibility of each step in the secondary relay algorithm.

3.3 Primary and Secondary Scheduling Scheme

After introducing our routing scheme, we will show the scheduling scheme for the primary and secondary networks. Similar to the protocols in Chapter ??, we use the 64-TDMA scheduling scheme for primary and secondary networks, and the duration of a secondary frame equals to the duration of a primary slot. To limit the interference between two networks, we also adopt preservation regions. From the knowledge of previous chapters, we know that this will not cause the decrease of transmission opportunities for secondary users in order sense.
Algorithm 2 Relay Algorithm for Secondary Packet $B_i^k$

**Input:** The secondary source node $S_i^{k_1}$ and destination node $S_j^{k_1}$

**Output:** The $K_{h_1}^* + 1$ intermediate secondary relay nodes

1: $S_i^{k_1}$ moves within its moving area until it meets a node of $T_{s,0}$ in the same secondary cell.
2: while $B_i^k$ is held by a $T_{s,0}$ secondary node whose moving area does not cover $X_i^{h_1}$ do
3: Execute $\mathcal{S}_{p,s}^*$.
4: end while
5: for $k = 0$ to $(K_{h_1}^* - 1)$ do
6: The relay node of $T_{s,k}$ moves within its moving area until it meets a node of $T_{s,k+1}$ in the same secondary cell.
7: Execute $\mathcal{S}_{s,s}^*$ based on the corresponding rules.
8: end for
9: $S_i^{k_1}$ moves within its moving area until it meets $S_j^{k_1}$ in the same secondary cell.
10: Execute $\mathcal{S}_{s,s}^*$ to deliver $B_i^k$ to the destination.

**Primary Scheduling Scheme** For the primary network, we also organize $T_{s,k} \in \mathcal{T}^{(p)}$ and $T_{p,k}$ denotes the $k$-th type of moving pattern in $\mathcal{T}^{(p)}$. Then the scheduling scheme includes $K_{h_p}^* + 2$ phases, for $T_{h_p} = T_{p,k_1}$, and each phase costs one primary frame.

**Phase 0** During the active slot of each primary cell, randomly choose a source node $P_i^{k_1}$ and relay packet $B_i^k$ to a random node of $T_0$ through $\mathcal{P}^{s,p}$ or $\mathcal{P}^{s,s}$.

For $k = 1, 2, \ldots, K_{h_p}^*$,

**Phase $k$** During the active slot of each cell, two types of transmission within this cell could happen in this phase: (1) transmission between primary node of $P_i^{k_k-1}$ and feasible relay node of $T_{k_{k-1}+1}$ ($T_{k_{k-1}} = T_{p,k-1}$) for primary packet $B_i^k$ ($K_k \leq j \leq h_p^*$); (2) transmission between primary node of $T_{p,k-1}$ and $P_j^{k_{k-1}}$ for a packet $B_p^{k_{k-1}}$ to reach the destination. One of such transmissions would be selected randomly to perform. Otherwise the cell stays idle.

**Phase $K_{h_p}^* + 1$** During the active slot of each cell, all pairs of nodes ($P_i^{h_p^*}$, $P_j^{k_1}$) ($h_p^* \leq j \leq h$) with in the cell are eligible for transmission to relay $B_i^k$ to the destination. One of such pairs would be selected randomly to transmit. Otherwise the cell stays idle.

**Secondary Scheduling Scheme** According to the routing scheme, the secondary network are required to relay both primary and secondary packets. Thus we propose a scheduling scheme consist of more phases to guarantee transmission opportunity for both kinds of packets. Specifically, we present a $K_{[h_p^*]} + K_{h_p^*} + 2$ phases scheme. For the first $K_{[h_p^*]}$ phases, secondary nodes serve as relay for primary packets; for the other phases, secondary nodes relay their own packets to the destinations.

For $k = 1, 2, \ldots, K_{[h_p^*]}$,

**Phase $k$** During the active slot of each secondary cell, randomly choose a node $S_i^{u_{k,k-1}}$ to transmit to a feasible relay node of $T_{k_{k+1}}$ for primary packet $B_i^k$ ($K_k \leq j \leq h_p^*$). If such pair does not exist, stay idle.
Phase $K_{h_p^*} + 1$ During the active slot of each secondary cell, randomly choose a secondary source node $S^h_{k}$ and relay packet $B^i_j$ to a randomly selected $S^h_{K_s}$ through $S^h_{K_s}$.

For $k = 1, 2, ..., K_{h_p^*}$,

Phase $K_{h_p^*} + 1 + k$ During the active slot of each secondary cell, two types of transmission within this cell could happen: (1) transmission between $S^h_{K_k-1}$ and a feasible relay node $S^h_{K_k}$ for secondary packet $B^i_j (K_k \leq j \leq h_p^*)$; (2) transmission between $S^h_{K_k-1}$ and $S^h_{K_k-1}$ for a packet $B^i_j$ to reach the destination. One of such transmissions would be selected randomly to perform. Otherwise the cell stays idle.

Phase $K_{h_p^*} + K_{h_p^*} + 2$ During the active slot of each secondary cell, all pairs of nodes $(S^h_{h_p^*}, S^h_{j})$ ($h_p^* \leq j \leq h$) with in the cell are eligible for transmission to relay $B^i_j$ to the destination. One of such pairs would be selected randomly to transmit. Otherwise the cell stays idle.

4 Throughput and Delay Scalings for the Primary Network

In this section, we will study the throughput and delay performance for the primary network based on our predefined protocols.

4.1 Throughput Performance

Before we proceed to the derivation of throughput performance of the primary network, we provide the following lemma.

Lemma 2. [7] Lemma 6 At any moment, the number of $P^k_i$ (for $T_i \in T(p), 0 \leq i \leq h$) nodes in a primary cell is $\Theta(\log n)$, w.h.p.; at any moment, the number of $S^h_i$ (for $T_i \in T(s), 0 \leq i \leq h$) nodes in a secondary cell is $\Theta(\log n)$, w.h.p..

We first consider the data rate of each transmission in primary and secondary networks using the following lemma.

Lemma 3. During the routing process of primary packets, each delivery type $P^{s \rightarrow p}$, $P^{p \rightarrow s}, P^{s \rightarrow s}, P^{s \rightarrow p}$ can achieve a constant data rate in each cell.

Proof. Based on analysis in previous sections, we know that for our TDMA scheduling scheme with frame structure in Figure 2 and the defined preservation regions, each primary and secondary cell can achieve a constant data rate w.h.p. Under the same scheduling scheme, we can also support each step of the routing process of primary packets with a constant data rate.

Then we can derive the per-node throughput of the primary network with the following theorem.

Theorem 1. With the proposed primary scheduling and routing scheme, the primary network can achieve the per-node throughput w.h.p.:

$$\lambda(N) = \Theta\left(\frac{1}{h^2 \log n}\right).$$

Proof. Our proof follows a similar logic as [7]. We divide the routing process into three parts: input, relay and output.

In the input process, the primary packets are initiated from all active primary cells to a $T_0$ relay node through $P^{p \rightarrow p}$ or $P^{p \rightarrow s}$, thus the aggregated throughput over the primary network in this process is

$$A_{input}(N) = \Theta\left(\frac{1}{h_p}\right) = \Theta\left(\frac{n}{\log n}\right).$$

Since each input process consumes $\frac{1}{K_{h_p^*}^2 + 2}$ fraction of the scheduling cycle, we have

$$\lambda_{input}(N) = A_{input}(N) \cdot \frac{1}{K_{h_p^*}^2 + 2} \geq A_{input}(N) \cdot \frac{1}{h_p} \cdot \frac{1}{6} = \Theta\left(\frac{1}{p^2 \log n}\right).$$
In the relay process, we have four different delivery situations. For $\mathcal{P}^{\text{in} \rightarrow \text{p}}$ and $\mathcal{P}^{\text{p} \rightarrow \text{out}}$, the aggregated throughput is the same as $\lambda_{\text{input}}(N)$. However these two situations would last for a constant fraction of the primary scheduling cycle $\frac{K_i}{\lambda_i^{\prime \prime} + 2}$, so we have

$$\lambda'_{\text{relay}}(N) = \lambda_{\text{input}}(N) \cdot \frac{1}{(h_p n)} \cdot \frac{K_i}{\lambda_i^{\prime \prime} + 2} \sim \lambda_{\text{input}}(N) \cdot \frac{1}{h_n} = \Theta\left(\frac{1}{h_n \log n}\right).$$

For $\mathcal{P}^{\text{out} \rightarrow \text{p}}$ and $\mathcal{P}^{\text{p} \rightarrow \text{out}}$, the delivery is executed in the secondary network. So we have

$$\lambda''_{\text{relay}} = \Theta\left(\frac{1}{a_s}\right) = \Theta\left(\frac{n^{\beta}}{\log n}\right),$$

and the process consumes $\frac{K_i}{\lambda_i^{\prime \prime} + 2}$ fraction of secondary scheduling cycle, we have

$$\lambda''_{\text{relay}}(N) = \lambda''_{\text{relay}}(N) \cdot \frac{1}{(h_p n)} \cdot \frac{K_i}{\lambda_i^{\prime \prime} + 2} \sim \lambda''_{\text{relay}}(N) \cdot \frac{1}{h_n} = \Theta\left(\frac{n^{\beta - 1}}{h_n \log n}\right).$$

The throughput of the output process is similar to the situation $\mathcal{P}^{\text{p} \rightarrow \text{out}}$ of the relay process, except that it consumes $\frac{1}{\lambda_i^{\prime \prime} + 2}$ fraction of the primary scheduling cycle. Therefore we have $\lambda_{\text{output}}(N) = \Theta\left(\frac{1}{h_n \log n}\right)$.

Consequently, the primary network can achieve a per-node throughput of $\lambda(N) = \min\left(\Theta\left(\frac{1}{h_n \log n}\right), \Theta\left(\frac{1}{h_n \log n}\right), \Theta\left(\frac{n^{\beta - 1}}{h_n \log n}\right)\right) = \Theta\left(\frac{1}{h_n \log n}\right)$.

### 4.2 Delay Performance

We have the following theorem to count the delay for the primary network.

**Theorem 2.** With the proposed primary relay algorithm, the primary network could achieve the following delay performance w.h.p.:

$$D_j(N) = \Theta(h^2 \log n + h^2 n^{\frac{\lambda}{\lambda_0}} \log^2 n + h^2 \log^2 n + h^2 n^{(1 - h_p^{\prime \prime} n^\chi)} \log n),$$

where $j$ denotes the type of primary destination node, and $h_p' = \min(h_p^*, j)$ as we defined previously.

**Proof.** We divide the whole routing process into input, relay and output process. To derive the delay performance of the primary network, we would evaluate the average number of frames during each of the three processes for a primary packet $B_i^p$ to reach its destination.

**Input Process** Now we consider the delay performance of input process, where primary source node $P_i^h$ will transmit to a relay node of $T^h_0$. Since for each type $T_i^h \in T(h)$ there are $\Theta(\log n)$ primary nodes within a primary cell, the probability that $P_i^h$ is chosen to transmit during the active time slot of the cell is $\mathbb{P}(P_i^h \text{ is chosen | the primary cell is active}) = \Theta(\frac{1}{h_n \log n})$. It is obvious that $\mathbb{P}(\text{the primary cell is active}) = \frac{1}{\lambda_i} = \Theta\left(\frac{1}{h_n \log n}\right)$, then we can obtain the probability for a successful transmission for $P_i^h$. In the input process is $\mathbb{P}_{\text{input}} = \Theta\left(\frac{1}{h_n \log n}\right)$. Delay performance in this process is $D_{\text{input}}(N) = \Theta(h^2 \log n)$.

**Relay Process** Now we consider the delay performance of relay process. During the relay process, the delivery of packet $B_i^p$ is propagated through $\mathcal{P}^{\text{p} \rightarrow \text{in}}, \mathcal{P}^{\text{p} \rightarrow \text{out}}, \mathcal{P}^{\text{out} \rightarrow \text{p}}, \mathcal{P}^{\text{p} \rightarrow \text{out}}$. We first consider the delay of $\mathcal{P}^{\text{p} \rightarrow \text{in}}$. In this process, $\mathcal{P}^{\text{p} \rightarrow \text{in}}$ happens between $P_i^{h_{i-1}}$ and $P_i^{h_i}$, for $1 \leq i \leq h_p^*$. It has been proved in [7] that the probability of a successful $\mathcal{P}^{\text{p} \rightarrow \text{in}}$ is $\Theta(n^{-\frac{\chi}{\lambda_0}})$. Considering that the primary cell is active for $\frac{1}{\lambda_i^{\prime \prime} + 2} = \Theta\left(\frac{1}{h_n \log n}\right)$ fraction of the primary scheduling cycle, and there are $\Theta(\log n)$ nodes within the cell for $T_{i-1}, T_{i}$, we obtain $\mathbb{P}_{\text{in} \rightarrow \text{p}} = \Theta\left(\frac{h}{h_n \log n}\right)$, and $D_{\text{in} \rightarrow \text{p}} = h n^{\frac{\lambda}{\lambda_0}} \log^2 n$.

Next we consider the delay of $\mathcal{P}^{\text{p} \rightarrow \text{in}}$. It happens between $P_i^{h_{i-1}}$ and $S_i^h$, for $1 \leq i < h_p^*$. In this case, the probability of a successful $\mathcal{P}^{\text{p} \rightarrow \text{in}}$ is also $\Theta(n^{-\frac{\chi}{\lambda_0}})$ because the node density of secondary users is larger than the primary users which will not lead to the decrease of this probability. Using the same method we can obtain $D_{\text{p} \rightarrow \text{in}} = h n^{\frac{\lambda}{\lambda_0}} \log^2 n$. 


For $\mathfrak{P}^{s\rightarrow s}$, the situation is similar to $\mathfrak{P}^{p\rightarrow p}$ except that the transmission happens $\frac{1}{\lambda} + \frac{\log n}{n^{3-\epsilon}}$ fraction of the secondary scheduling cycle, and this will not influence the delay in order sense.

At last, we consider the situation of $\mathfrak{P}^{s\rightarrow p}$. Recall that for $\mathfrak{P}^{p\rightarrow s}$ during the relay process, each primary user broadcast its packet $B^p_i$ with power $P_{p,h_p}^{1/2}$ to all secondary users residing in the secondary cell. It is clear that there are at least $O(n^{3-1})$ secondary users in each primary cell, i.e., at least $O(n^{3-1})$ copies of packet $B^p_i$ are propagated in the secondary network until they are transmitted back to the primary network through $\mathfrak{P}^{p\rightarrow p}$.

Based on the proof in [7], we can obtain that the probability that an eligible primary relay node $P^u_i$ resides in the same secondary cell as $S_{n-1}^{n-1}$ in a active time slot is $\Theta(n^{-\alpha/h \log n})$, then at least one copy of $B^p_i$ could be forwarded successfully, because

$$P(\text{forwarded}) = 1 - (1 - \Theta(n^{-\alpha/h \log n}))^n \sim \Theta(1),$$

for $n^{-\alpha/h} = \Theta(1)$. Thus we obtain $D_{\mathfrak{P}^{p\rightarrow p}} = h \log^2 n$.

To eliminate the effects of copying $B^p_i$, once a copy successfully execute $\mathfrak{P}^{s\rightarrow p}$ and forward $B^p_i$ back to the primary network, other copies are outdated and would be rejected by any $P^u_i$. Such assumption is to guarantee that no bottleneck effects would happen regarding the throughput of the primary network.

**Output Process** During the output process, $\mathfrak{P}^{p\rightarrow p}$ happens between $P^u_i$ and $P^j_k$, for $h_p = \min(j, h^*_p)$. The probability of $P^u_i$'s meeting $P^j_k$ in the same cell is

$$P_{\mathfrak{P}^{p\rightarrow p}} = \sum_{Q} P(X_{P^u_i}^u \in Q)P(X_{P^j_k}^k \in Q) \geq \sum_{Q} \left( \frac{\alpha}{\pi R_{h_p}^2} \right)^2 \sim R_{h_p}^2 \left( \frac{\alpha}{\pi R_{h_p}^2} \right)^2 = \Theta \left( \frac{1}{h \log n} \right),$$

where $Q$ denotes the set of cells inside $C_{h_p} \cap C_{h_p}$. For the worst case where $j \geq h_p$, the probability that pair $(P^u_i, P^j_k)$ is chosen to transmit is $\Theta \left( \frac{1}{h \log n} \right)$, thus we can obtain $D_{output}(N) = h^2 n^{1-h_p \alpha/h} \log n$.

Combining all the delay for each step of the whole relay process, we can obtain the delay performance of primary packets:

$$D_j(N) = \Theta(h^2 \log n + h^2 n^{-\alpha/h} \log^2 n + h^2 \log^2 n + h^2 n^{1-h^*_p \alpha/h} \log n),$$

where we use the fact that $\mathfrak{P}^{p\rightarrow p}, \mathfrak{P}^{p\rightarrow s}, \mathfrak{P}^{s\rightarrow s}, \mathfrak{P}^{s\rightarrow p}$ each consumes $\Theta(h)$ steps to relay the primary packets.

## 5 Throughput and Delay Scalings for the Secondary Network

In this section we study the throughput and delay performance for the secondary network. The secondary network is different from the primary network for they can only access the spectrum opportunistically, and we use preservation regions to guarantee such assumption. According to analysis in previous sections, we know that the adoption of preservation regions will not cause the decrease of throughput and delay performance in order sense. Therefore we proceed without discussing some specific issues regarding preservation regions.

### 5.1 Throughput Performance

Similar to Lemma 3, we have the following lemma.

**Lemma 4.** During the routing process of secondary packets, each delivery type $\mathfrak{S}^{s\rightarrow s}$ and $\mathfrak{S}^{s\rightarrow s}$ can achieve a constant data rate in each cell.

Then we can derive the per-node throughput of the secondary network with the following theorem.

**Theorem 3.** With the proposed secondary scheduling and routing scheme, the secondary network can achieve the per-node throughput w.h.p.:

$$\lambda(M) = \Theta \left( \frac{1}{h^2 \log n} \right).$$
Proof. We still divide the routing process into three parts: input, relay and output.

In the input process the secondary packets are initiated from all active secondary cells to a $T_{s,0}$ type relay node through $\mathcal{G}_0^{s \rightarrow s}$, thus

$$\lambda_{input}(M) = \Lambda_{input}(M) \cdot \frac{1}{(h, n^\tau)} \cdot \frac{1}{K_{\frac{1}{2}^s} + \frac{1}{K_{\frac{1}{2}^s} + 2}} = \Theta\left(\frac{n^\tau}{\log n} \cdot \frac{1}{(h, n^\tau)} \cdot \frac{1}{n} \right) = \Theta\left(\frac{1}{\log n}\right).$$

In the relay process, secondary packets are delivered through $\mathcal{G}^{s \rightarrow s}$, it can support an average per-node throughput of

$$\lambda_{relay}(M) = \Lambda_{\mathcal{G}^{s \rightarrow s}}(M) \cdot \frac{1}{(h, n^\tau)} \cdot \frac{K_{\frac{1}{2}^s} + 1}{K_{\frac{1}{2}^s} + K_{\frac{1}{2}^s} + 2} = \Theta\left(\frac{n^\tau}{\log n} \cdot \frac{1}{(h, n^\tau)} \right) = \Theta\left(\frac{1}{h \log n}\right).$$

In the output process, the situation is similar to that of the input process, we have

$$\lambda_{output}(M) = \Theta\left(\frac{1}{h^2 \log n}\right).$$

Consequently, the secondary network can achieve a per-node throughput of $\lambda(M) = \min\left(\Theta\left(\frac{1}{\tau \log n}\right), \Theta\left(\frac{1}{h \log n}\right)\right) = \Theta\left(\frac{1}{h^2 \log n}\right)$.

5.2 Delay Performance

We have the following theorem to count the delay for the primary network.

**Theorem 4.** With the proposed secondary relay algorithm, the secondary network could achieve the following delay performance w.h.p.:

$$D_j(M) = \Theta\left(h^2 \log n + h \log n + h^2 n^3 \frac{\beta}{h^2} \log^2 n + h^2 n^{(\beta - h^2 \frac{\beta}{h^2})} \log n\right),$$

(12)

where $j$ denotes the type of primary destination node, and $h_p' = \min\left(h_p, j\right)$ as we defined previously.

Proof. We still divide the whole routing process into input, relay and output process. We evaluate the average number of secondary frames during each of the three processes for a secondary packet $B_j^s$ to reach its destination.

**Input Process** First, we consider the delay performance of input process, where secondary source node $S^k_i$ will transmit to a relay node of $T_{s,0}$ through $\mathcal{G}_0^{s \rightarrow s}$. Since for each type $T_i \in \mathcal{T}^{(s)}$ there are $\Theta(n)$ secondary nodes within a secondary cell, the probability that $S^k_i$ is chosen to transmit during the active time slot of the cell is

$$\mathbb{P}(S^k_i \text{ is chosen | the secondary cell is active}) = \Theta\left(\frac{1}{n \log n}\right).$$

It is obvious that

$$\mathbb{P}(\text{the secondary cell is active}) = \frac{1}{K_{\frac{1}{2}^s} + \frac{1}{K_{\frac{1}{2}^s} + 2}} = \Theta\left(\frac{1}{h}\right),$$

then we can obtain the probability for a successful transmission for $S^k_i$ in the input process is $\mathbb{P}'_{input} = \Theta\left(\frac{1}{h \log n}\right)$. Delay performance in this process is $D_{input}(M) = \Theta\left(h^2 \log n\right)$.

**Relay Process** Now we move to the delay performance of relay process. During the relay process, the delivery of packet $B_j^s$ consists of two parts: relay process through $\mathcal{G}_0^{s \rightarrow s}$ and relay process through $\mathcal{G}^{s \rightarrow s}$. We first study the case with $\mathcal{G}_0^{s \rightarrow s}$.

Recall that in the secondary routing scheme, we keep packet $B_j^s$ propagated among relay nodes of $T_{s,0}$ until we find one such relay node whose moving area covers the initial position $X^k_{i,j}$, i.e., $X^{\hat{k}_j}_{i,j} - X^{k}_i < R_{\hat{k}_j} - \sqrt{2\alpha_{s,j}}$. We now show that such an eligible relay node of $T_{s,0}$ could be found through a constant number $\tau$ of $\mathcal{G}_0^{s \rightarrow s}$.

Since each elements in $\mathcal{T}^{(s)}$ is randomly selected from $\mathcal{T}$ with probability $\frac{1}{2}$, if $T_{s,0}$ is not 0-th type, then $T_{s,0}$ is $\Theta(1)$-th type, w.h.p., because $(\frac{1}{2})^{\Theta(1)} \to 0$. We denote $T_{s,0}$ is $\tau_0$-th type. It is obvious that the probability that $X^{\hat{k}_j}_{i,j}$ locates within the moving area of $S_i^{k_0}$ is $\pi R_{\hat{k}_j}^2 = \pi n^{-\tau_0 \chi_0 / h}$. Then the probability that it costs $\omega(1)$ number of $\mathcal{G}_0^{s \rightarrow s}$ to find a relay node of $T_{s,0}$ is $(1 - \pi n^{-\tau_0 \chi_0 / h} \omega(1)) \to 0$, for $n^{-\tau_0 \chi_0 / h} \sim \Theta(1)$. 


We have proved that the primary network can achieve \( \lambda = 6.1 \) Optimal Throughput and Delay Performance type of moving pattern. At last, we will show some possible extensions to our work. Specifically, we compare our primary and secondary networks under our routing and scheduling scheme.

### 6 Discussion on the Delay and Throughput Performance

In this section, we will discuss how mobility heterogeneity of primary and secondary users can affect the delay and throughput performance of cognitive network. We first evaluate the optimal performance of primary and secondary networks under our routing and scheduling scheme. Specifically, we compare our results with [7] to show the advantages of our scheme. Then we discuss the performance in different type of moving pattern. At last, we will show some possible extensions to our work.

#### 6.1 Optimal Throughput and Delay Performance

We have proved that the primary network can achieve \( \lambda(N) = \Theta\left(\frac{1}{n^3 \log n}\right) \), w.h.p.. Since \( h = \Theta(\log n) \), the achievable per-node throughput \( \lambda(N) = \Theta\left(\frac{1}{\log^4 n}\right) \).

For the optimal delay performance for primary network, we have

\[
D_{\text{optimal}}(N) = \begin{cases} 
\Theta(n^{1-\chi_0} \log^3 n), & 0 \leq \chi_0 < 1 - \frac{\log \log n}{\log n}, \\
\Theta(\log^4 n), & \chi_0 \geq 1 - \frac{\log \log n}{\log n}.
\end{cases}
\]

We can see that for the primary network, the near-optimal throughput and delay, as well as the delay-throughput tradeoff \( \Theta(poly(\log n)) \) can be achieved when \( h = \Theta(\log n) \) and \( \chi_0 > 1 \).

For the secondary network, the achievable per-node throughput is \( \lambda(M) = \Theta\left(\frac{1}{\log^4 n}\right) \). The optimal delay performance can be characterized as follows,

\[
D_{\text{optimal}}(M) = \begin{cases} 
\Theta(n^{\beta-\chi_0} \log^3 n), & 0 \leq \chi_0 < \beta - \frac{\log \log n}{\log n}, \\
\Theta(\log^4 n), & \chi_0 \geq \beta - \frac{\log \log n}{\log n}.
\end{cases}
\]

We can see that for the secondary network, the near-optimal throughput and delay, as well as the delay-throughput tradeoff \( \Theta(poly(\log n)) \) can be achieved when \( h = \Theta(\log n) \) and \( \chi_0 > \beta \).

Now we compare our results with [7] to show the advantages of our routing and scheduling scheme. In [7], the author proposed a transmission scheme with an achievable throughput \( \lambda(n) = \Theta\left(\frac{1}{\log n}\right) \) and delay

\[
D_{\text{optimal}}(n) = \begin{cases} 
\Theta(n^{\beta-\chi_0} \log n), & 0 \leq \chi_0 < \beta - \frac{3 \log \log n}{\log n}, \\
\Theta(\log^4 n), & \chi_0 \geq \beta - \frac{3 \log \log n}{\log n},
\end{cases}
\]

for \( n \) primary users. Thus both our scheme and the scheme in [7] can achieve near-optimal throughput of \( \Theta(poly(\log n)) \). However our scheme can guarantee more primary nodes \( \Theta(hn) \) with better delay performance as shown in Figure 3. For the secondary network, our scheme can achieve similar performance with respect to throughput and delay to their scheme.
6.2 Performance of Different Moving Type

Under our proposed routing and scheduling scheme, the primary and secondary users can achieve near-optimal throughput regardless of their different moving types. However the delay performance is highly related to the moving pattern each node possesses, for

\[ D_j(N) = \Theta(h^2 \log n + h^2 n^\frac{\chi_0}{2} \log^2 n + h^2 \log^2 n + h^2 n^{(1-h^2)\frac{\chi_0}{2}} \log n), \]

\[ D_j(M) = \Theta(h^2 \log n + h \log n + h^2 n^{\frac{\chi_0}{2}} \log^2 n + h^2 n^{(1-h_0)\frac{\chi_0}{2}} \log n), \]

where \( h_p = \min(j, h^*_p) \) \( T_j \in T^{(p)} \), \( h'_p = \min(j, h^*_p) \) \( T_j \in T^{(s)} \).

Obviously, \( D_j(N) \) will decrease when moving type of primary nodes increases. For 0-th type primary nodes, \( D_0(N) = \Theta(n \log^4 n) \) with \( \lambda_0(N) = \Theta(\frac{1}{\log^4 n}) \), and our results indicate a similar phenomenon as described in [4], i.e., \( D(n) = n \lambda(n) \), for \( \lambda(n) = 1 \). When \( j \) becomes larger, the delay performance \( D_j(N) \) is improved with the help of relay nodes of lower moving type, which indicates that heterogeneous mobility introduces transmission diversity into the network to reduce transmission delay and guarantee near-optimal throughput performance.

We also notice that the optimal performance occurs to the primary nodes with minimal moving area. When \( 0 < \chi_0 < 1 \), the primary nodes of \( T_{p,h_p} \) achieves optimal delay \( D_{K_{h_p}}(N) = \Theta(n^{1-\chi_0} \log^3 n) \); when \( \chi_0 \geq 1 \), the primary nodes whose moving patterns are at least \( (h^*_p - \log \log n - \Theta(1)) \)-th type can achieve optimal delay of \( D_{K_{h_p}}(N) = \Theta(\log^4 n) \).

For the secondary network, the delay performance is similar to that of the primary network. It is also presented that better performance can be achieved for secondary nodes of larger moving type. However, when users tend to have strong moving ability, the overall performance of secondary network is worse than the primary network by an order of \( \Theta(n^{\beta-1}) \) due to larger node density. Otherwise both networks can achieve the same near-optimal performance.

7 Conclusion

This paper studies the throughput and delay scaling laws of GHSM cognitive radio networks. In the GHSM model, we study the scaling laws of two tiers of mobile nodes with heterogeneous moving pattern. For primary transmissions, we utilize both primary and secondary nodes to forward primary packets progressively to the destination through a chain of relay nodes with continuous moving pattern. By employing such a cooperative routing strategy, we guarantee the primary users a better optimal performance, and the delay-throughput tradeoff achieves a near-optimal order of \( \Theta(poly \log n) \) when heterogeneity is increased sufficiently. For the secondary transmissions, only secondary nodes are used to relay packets due to the priority of primary users. We prove that the secondary network can still achieves the same performance as shown in [7] even though secondary relay process may suffer from moving ability gaps among nodes with different moving patterns. The impact of different moving ability on throughput and delay performance is also discussed and possible extensions are provided for future work.

\[^4\text{Even if } T_0 \notin T^{(p)}, \text{ we have } D_{K_0}(N) = \Theta(n \log^3 n) \text{ for } T_{p,0} \text{ is } \Theta(1)-\text{th type w.h.p. in this case.}\]
References