On Latency Distribution and Scaling: From Finite to Large Cognitive Radio Networks under General Mobility

Lei Sun Wenye Wang
Department of Electrical and Computer Engineering
North Carolina State University, Raleigh, NC 27695

***Abstract*—Cognitive Radio Networks (CRNs), as a phenomenal technique to improve spectrum efficiency for opportunistic communications, become an integral component in the future communication regime. In this paper, we study the end-to-end latency in CRNs because many CRN applications, such as military networks and emergency networks, are either time-sensitive or dependent on delay performance. In particular, we consider a general mobility framework that captures most characteristics of the existing models and accounts for *spatial heterogeneity* resulting from the scenario that some locations are more likely to be visited by mobile nodes (these can be home in the case of people, or garage in the case of vehicles). By assuming that secondary users are mobile under this general framework, we find that there exists a cutoff point on the *mobility radius* α, which indicates how far a mobile node can reach in the spatial domain, below which the latency has a *heavy-tailed* distribution and above which the tail distribution is bounded by some *Gamma* (*light-tailed*) distribution. A *heavy tail* of the latency implies a significant probability that it takes long time to disseminate a message from the source to the destination and thus a *light-tailed* latency is crucial for time-critical applications. Moreover, as the network grows large, we notice that the latency is asymptotically scalable (linear) with the dissemination *distance* (e.g., the number of hops or Euclidean distance). Another interesting observation is that although the density of primary users adversely impacts the expected latency, it makes no influence on the *dichotomy* of the tail distribution of the latency in finite networks and the linearity of latency in large networks. Our results encourage the CRN deployment for real-time and large applications, when the mobility radius of secondary users is large enough.**

I. Introduction

As an integral component of emerging communication infrastructure and a promising solution to address the challenge of spectrum under-utilization, Cognitive Radio Networks (CRNs) have received considerable attentions [1]. In such networks, there are two types of users: (i) *primary users* who have license from the regulator and thus have priority to utilize spectrum, and (ii) *secondary users* who opportunistically access spectrum without interfering with the coexisting primary users. Many efforts have been made recently to understand the characteristics of CRNs and thus to enable the deployment of such networks for realistic applications, including capacity limits, spectrum sensing, spectrum mobility, and spectrum sharing [1]–[4]. These works have presented a very good understanding of the potential of cognitive communications in optimizing spectrum utilization. However, the key question to the deployment of CRNs is *not* whether the spectrum efficiency is improved, but whether the CRNs are able to support *applications*. For example, spectrum can be overly used, with a very high throughput, but the latency may become extremely long, falling into the traditional problem of the tradeoff between network throughput and latency [5]–[8]. To this end, we aim to study a fundamental problem, i.e., what the stochastic properties of end-to-end latency in CRNs are.

Despite its importance, the latency is an under-explored problem and not well understood in wireless multihop networks. The pioneering work in [9] studied the packet latency for the *fully connected* wireless ad-hoc networks and showed that there exist bounds on the latency which are tight when the number of nodes is large enough. Instead of *full connectivity*, the reference [10] further showed that the latency scales asymptotically at least linearly with the transmission distance in wireless sensor networks when these networks are *percolated*. These results have greatly advanced our understanding of the nature of latency, and also laid a good foundation to approach the problem. Unfortunately, these results may not be applicable to CRNs because (i) asymptotic results were obtained by assuming that wireless nodes are static; and (ii) these results [9], [10] are only derived for *large* networks when the number of nodes approaches to infinity; (iii) these results are derived for *homogeneous networks* in which every node has the same capacity in information propagation.

Particularly, node mobility plays a critical role on the latency, which has been evidenced by earlier results. For instance, the seminal work [5] showed that mobility can improve the capacity in large wireless ad hoc networks at the cost of the delay. This result is obtained by assuming that nodes move according to an ergodic process that are equally likely to visit any portion of the network area. That is, the nodes are *spatially homogeneous.* With the similar assumptions, capacity-delay tradeoffs have been extensively studied under various mobility models, such as under the i.i.d model [6], the Brownian motion [7], the reshuffling model [8] and variants of random walk and random way-point models [11], [12]. Later on, spatial inhomogeneity has been taken into account in [13] where the nodes are restricted to move within the coverage of a home point. These studies motivate an interesting question about the latency under general mobility. Furthermore, it is evident that the asymptotic results, though, provide good insights into network performance, may not explain the latency properties when the number of nodes in real applications is *finite*. As the last point, CRNs feature *heterogeneity* in wireless nodes, since there are two types of nodes, primary nodes and secondary

This work is supported by the NSF Award CNS 0546289, NSF Career Award and Defense Threat Reduction Agency (DTRA) Award HDTRA1-08-1-0024.

nodes [3], which is left open for study on the impact of latency distribution.

Putting all together, in this paper we study the latency distribution in *finite* networks, and the *scaling law* for large networks with *infinite* number of secondary nodes under general mobility. We find that in finite CRNs, the latency of information dissemination depends on the *mobility radius* α, which indicates how far a mobile node can reach in spatial domain. Also, there exists a *cutoff point* on α, below which the latency has a *heavy-tailed* distribution; and above which its tail distribution is bounded by some Gamma distribution. In addition, as the network grows large, the latency asymptotically scales linearly with respect to the *distance* in terms of the number of hops or the Euclidean distance between the source and destination nodes if the network remains fully connected or percolated. It is interesting, though not surprising, that the density of primary nodes presents an adverse influence on the expected latency, but showing no obvious effect on the dichotomy of the latency tail in finite networks and linear scaling law of the asymptotic latency.

The rest of this paper is organized as follows. We describe the mobility and network models, and formulate the latency problem in Section II. In Sections III and IV, we present the results and proofs of our findings on dissemination latency in detail. In Section V, we present simulation results and finally, we conclude in Section VI.

II. System Models and Problem Formulation

In this section, we first describe the network models and then collect basic assumptions, notations and definitions of the metric of interest that will be used throughout the paper.

A. Network Models

We consider a CRN consisting of n mobile secondary users $\{v_1, \dots, v_n\}$ in a torus surface $\Omega_n = [0, \sqrt{\frac{n}{\lambda}}]^2$ (λ is the spatial density of secondary users). Denote $V(t) = (v_1(t), \dots, v_n(t))$ as the positions of secondary users at time t. A set of m channels $\{ch_1, \dots, ch_m\}$ are assumed to be accessible by secondary users. For any $1 \leq k \leq m$, an overlay network of primary users with spatial density λ_{pk} are transmitting with channel ch_k. We assume that $\lambda_{pk} = \lambda_p$ for any k for simplicity. To model the dynamics of the primary traffic, we adopt a synchronized slotted structure, which has been used in [3] to study the connectivity of a large single-channel CRN. Particularly, time is slotted into units and at any time slot, primary users transmitting on any channel ch_k are assumed to be uniformly and independently distributed in Ω_n, and such distribution is i.i.d across slots.

1) Interference Models: In CRNs, there are two types of interference for information dissemination among secondary users: *secondary-secondary* and *primary-secondary* interference. The former interference can be characterized by the well-known *protocol model* [14], which has been widely adopted in homogeneous networks [5], [6], [9], [14]. Particularly, without interference with primary users, a successful transmission from a mobile secondary user v_i to v_j is achievable at time t

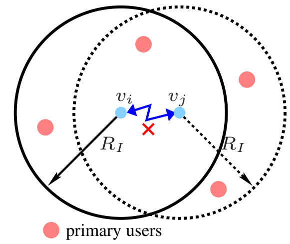


Fig. 1. *Primary-secondary* interference.

if $\|v_i(t) - v_j(t)\| \leq r$ and for any other simultaneously transmitting node on the same channel v_l, $\|v_l(t) - v_j(t)\| \geq (1 + \Delta)r$, where r is the transmission radius of secondary users, and Δ models the guard zone around v_j in which any simultaneously transmission on the same channel causes collision at v_j. In terms of the latter interference, let us denote R_I as the interference range of primary users. And as Fig. 1 shows, two secondary users v_i and v_j are permitted to use the channel ch_k only when there are no primary users on ch_k in the neighborhood, i.e., $\|v_i(t) - u(t)\| > R_I$ for any primary user u transmitting with ch_k, where $u(t)$ is the position of u at time t.

2) Mobility Models: We consider a general mobility model, $\mathcal{M}(\Phi, \Psi, \alpha)$, which is characterized by three parameters Φ, Ψ, and α over the region Ω_n. First, *spatial heterogeneity*, which accounts for the scenario that mobile nodes are more likely to be found in some area (e.g., the neighborhood of their home in the case of people, or the neighborhood of the garage in the case of vehicles), is taken into account. Particularly, we consider that a node spends most of its time in a small region, and rarely visits the areas far away from it. We model this behavior by assuming that each node v_i has a *home point* [13], located at v_i^h. Nodes move "around" their home points according to independent stationary and ergodic processes. We assume that each home point v_i^h is associated with a fixed point v_i^c, which is called the *center point* of v_i. The center points are regularly placed in Ω_n. For example, $\{v_1^c, \dots, v_n^c\}$ are placed regularly at positions $(\frac{1}{2\sqrt{\lambda}} + \frac{i}{\sqrt{\lambda}}, \frac{1}{2\sqrt{\lambda}} + \frac{j}{\sqrt{\lambda}})$ with $0 \leq i \leq \sqrt{n} - 1$ and $0 \leq j \leq \sqrt{n} - 1$ (we generally assume that n is a square of some integer for simplicity, see Fig. 2). We describe the distribution of the home point v_i^h around v_i^c by a non-increasing probability density function $\Phi_i(x) = \Phi(x - v_i^c)$, which is assumed to be invariant in all directions and used as the first parameter in the mobility model. The second parameter, $\Psi_i(x) = \Psi(x - v_i^h)$ is used to describe the probability density of a node v_i around v_i^h, which is again a non-increasing and direction-invariant function. We assume that Ψ_i is non-zero in and only in a region characterized by a constant α; that is, $\Psi_i(x) = \Psi(x - v_i^h) > 0$ when $\|x - v_i^h\| < \alpha$ and $\Psi_i(x) = \Psi(x - v_i^h) = 0$, otherwise. We refer α as ***mobility radius***.

Remark 1: The idea of "home points" is not new [13] and it has been used to describe the *spatial inhomogeneity* incurred by the mobility of a particular wireless node. We introduce an additional concept, "center points" to model the heterogeneously spatial distribution of the home points, which characterizes the *spatial inhomogeneity* incurred by

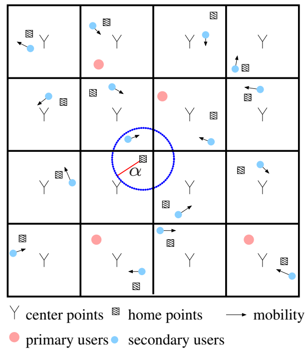


Fig. 2. An illustration of the general mobility $\mathcal{M}(\Phi, \Psi, \alpha)$.

heterogeneous mobility of different wireless nodes. This two-level mobility model accounts for a wide range of mobility patterns. For example, if the probability density function $\Phi(x)$ is a constant function independent of x (i.e., home points are uniformly distributed over Ω_n), $\mathcal{M}(\Phi, \Psi, \alpha)$ reduces to the *Uniform Anisotropic* model in [13]. Furthermore, if the probability density function $\Psi_i(x) = \Psi(x - v_i^h) = \delta(x - v_i^h)$, where $\delta(x)$ is the Dirac impulse function, $\mathcal{M}(\Phi, \Psi, \alpha)$ reduces to the static model in [14], where nodes are assumed to be static and uniformly distributed; if $\Psi(x)$ is also a constant function independent of x and α, $\mathcal{M}(\Phi, \Psi, \alpha)$ reduces to the homogeneous mobility model in [5]; and if $\Psi(x)$ is a threshold function whose value is zero when $x \geq \alpha$ and a nonzero constant when $x < \alpha$, $\mathcal{M}(\Phi, \Psi, \alpha)$ reduces to the *constrained i.i.d* model used in [10].

Mobility of Secondary Users: In this paper, we assume that secondary users are mobile under the *general mobility* $\mathcal{M}(\Phi, \Psi, \alpha)$. To facilitate the study of the dissemination latency of secondary users, we consider three classes according to the *spatial inhomogeneity* of home points:

- Extremely Inhomogeneous Home Points (**EIHP**) mobility $\mathcal{M}(\Phi_E, \Psi, \alpha)$: Home points are fixed and regularly placed over Ω_n. Here $\Phi_E(x) = \delta(x)$.
- Partial Inhomogeneous Home Points (**PIHP**) mobility $\mathcal{M}(\Phi_P, \Psi, \alpha)$: As shown in Fig. 2, center points $\{v_i^c\}_{i=1}^n$ partition the Ω_n into n *subregions* $\{\mathcal{O}_i\}_{i=1}^n$ as Voronoi diagrams. In this class, the home point v_i^h is randomly distributed in $\mathcal{O}_i$.
- Homogeneous Home Points (**HHP**) mobility $\mathcal{M}(\Phi_H, \Psi, \alpha)$: Home points $\{v_i^h\}_{i=1}^n$ are independently and uniformly distributed over Ω_n. Here $\Phi_H(x)$ is a constant density function independent of x.

B. Problem Formulation

We denote $[\mathcal{F}_{m,n}, \mathcal{M}(\Phi, \Psi, \alpha), (\lambda, \lambda_p)]$ as a CRN $\mathcal{F}_{m,n}$, where n secondary users opportunistically access m channels and are mobile under $\mathcal{M}(\Phi, \Psi, \alpha)$, and the spatial densities of secondary users and primary users on any channel are λ and λ_p respectively. We further denote $\mathbb{L}(t)$ as the set of communication links among secondary users at time t and $\mathbb{L}(t)$ is obviously dynamic due to the mobility of the secondary and primary users.

In this paper, we focus on the *dissemination latency*, i.e., *how fast* information can be disseminated from the source to the destination secondary user. Therefore, rebroadcasting and "store-carry-and-forward" communication paradigm (also named mobility-assisted routing) are considered. Specifically, by omitting the propagation delay, when the source v_s broadcasts a message at time 0, all the secondary users connected to v_s in $\mathbb{L}(0)$ receive the message instantly. Denote $l_{i,j}$ as a communication link between secondary users v_i and v_j and $\mathcal{V}(t)$ as the set of secondary users that have received the message at time t.

Definition 1: The *first hitting time* between v_i and v_j is defined as $\mathcal{T}_h(v_i, v_j) \triangleq \inf\{t \geq 0 : l_{i,j} \in \mathbb{L}(t)\}$.

Definition 2: The *dissemination latency* $\mathcal{T}_d$ from the source v_s and the destination v_d is defined as:

$$\mathcal{T}_d \triangleq \inf\{t \geq 0 : v_d \in \mathcal{V}(t)\}.$$

Based on the definitions and system models, we can formulate the problem as

1) In a finite $\mathcal{F}_{m,n}$, what the distribution of the dissemination latency $\mathcal{T}_d$ is;
2) As the network grows large, say to infinity, whether the dissemination latency $\mathcal{T}_d$ is scalable or not.

In $[\mathcal{F}_{m,n}, \mathcal{M}(\Phi, \Psi, \alpha), (\lambda, \lambda_p)]$, three metrics can be used to characterize *how far* two nodes v_i and v_j are apart:

- $d^{(t)}(v_i, v_j)$: the distance between v_i and v_j at time t.
- $d_h(v_i, v_j)$ and $d_c(v_i, v_j)$: the distance between home points and center points of v_i and v_j respectively.

III. The Distribution of $\mathcal{T}_d$ in Finite Networks

In this section, we study how fast information is disseminated among secondary users in *finite* CRNs under general mobility $\mathcal{M}(\Phi, \Psi, \alpha)$. Particularly, we first study the distribution of the dissemination latency $\mathcal{T}_d$ under the three subclasses of mobility models, i.e., EIHP $\mathcal{M}(\Phi_E, \Psi, \alpha)$, PIHP $\mathcal{M}(\Phi_P, \Psi, \alpha)$, and HHP mobility $\mathcal{M}(\Phi_H, \Psi, \alpha)$ defined in Section II, respectively. Then we move on to identify the fundamental properties of $\mathcal{T}_d$ under the general mobility. To proceed, we find the following definitions useful toward the derivation of tail distribution of the latency $\mathcal{T}_d$.

Definition 3: If Z and Z' are random variables such that $\mathbb{P}(Z > z) \leq \mathbb{P}(Z' > z)$ for all z, we say that Z is *stochastically dominated* by Z' and write $Z \overset{\mathcal{D}}{<} Z'$; and if $Z \overset{\mathcal{D}}{<} Z'$, there exists a random variable $\hat{Z}'$, which has the same distribution of Z' such that $Z \leq \hat{Z}'$ ($\hat{Z}'$ is called a *coupling* of Z' [15].)

Definition 4: If Z and Z' are random variables such that $\mathbb{P}(Z > z) \leq \mathbb{P}(Z' > z)$ for large z, we say that the Z's **tail** is *stochastically dominated* by Z'.

Remark 2: *Coupling* is a very important tool in probability theory which is used throughout the paper. To use the coupling method, *stochastic domination* is required (as shown in Definition 3). However, in finite CRNs, we are interested in the tail distribution of the dissemination latency $\mathcal{T}_d$, which implies that only *stochastic tail domination* needs to be considered.

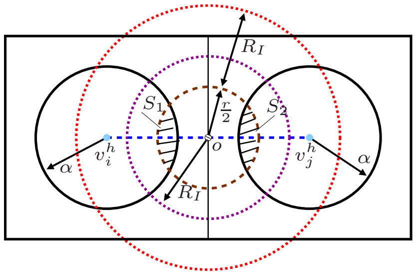


Fig. 3. Calculation of the first hitting time under EIHP Mobility.

Therefore, in order to use *coupling*, we need the following lemma, which bridges the gap between *stochastic domination* and *stochastic tail domination*.

Lemma 1: Given non-negative i.i.d random variables $\{X_i\}_{i=1}^{\infty}$ and $\{Y_i\}_{i=1}^{\infty}$ where $\mathbb{P}(X_i > t) \le \mathbb{P}(Y_i > t)$ for large t, i.e., the *tails* of the former are *stochastically dominated* by the latter, there exist i.i.d random variables $\{\bar{X}_i\}_{i=1}^{\infty}$, which have the same tail distribution with $\{X_i\}_{i=1}^{\infty}$ and are *stochastically dominated* by $\{Y_i\}_{i=1}^{\infty}$. Furthermore, for any finite k, $\sum_{i=1}^{k} X_i$ has the same tail distribution with $\sum_{i=1}^{k} \bar{X}_i$.

Proof: (Sketch.) Assume $\mathbb{P}(X_i > t) \le \mathbb{P}(Y_i > t)$ when $t > t_c$ for some finite constant t_c. We construct $\{\bar{X}_i\}_{i=1}^{\infty}$ as $\bar{X}_i = 0$ when $X_i \le t_c$ and $\bar{X}_i = X_i$ otherwise. This proves the first part. For the second part, we only need to show $\mathbb{P}(\bar{X}_1 + \bar{X}_2 > t) = \mathbb{P}(X_1 + X_2 > t)$ for large t:

$$\begin{aligned}\mathbb{P}(\bar{X}_1 + \bar{X}_2 > t) &= \mathbb{P}(\bar{X}_1 < t_c)\mathbb{P}(\bar{X}_1 + \bar{X}_2 > t | \bar{X}_1 < t_c) + \\ &\mathbb{P}(\bar{X}_1 > t - t_c)\mathbb{P}(\bar{X}_1 + \bar{X}_2 > t | \bar{X}_1 > t - t_c) \\ &+\mathbb{P}(t_c < \bar{X}_1 < t - t_c)\mathbb{P}(\bar{X}_1 + \bar{X}_2 > t | t_c < \bar{X}_1 < t - t_c).\end{aligned}$$

Note that the third item on the right hand side is equal to its counterpart of $\mathbb{P}(X_1 + X_2 > t)$ and the first two items are on the higher order of the third item as $t \to \infty$. Thus $\mathbb{P}(\bar{X}_1 + \bar{X}_2 > t) \to \mathbb{P}(X_1 + X_2 > t)$ for large t and this completes the proof. ∎

A. Distribution of $\mathcal{T}_d$ under EIHP Mobility $\mathcal{M}(\Phi_E, \Psi, \alpha)$

Prior study [10] has shown that *propagation delay* in networks whose topologies change frequently (e.g., due to mobility) is negligible, in comparison with the latency incurred by the topology dynamics. Therefore, $\mathcal{T}_d$ can be coupled as the sum of a sequence of the first hitting time $\mathcal{T}_h$ between secondary users along a communication path from the source to the destination node. Hence we study $\mathcal{T}_h$ first. In EIHP mobility, secondary users move around home points, which are overlaid with center points, with the Euclidean distance between neighboring home points being $\sqrt{\frac{1}{\lambda}}$ (see Fig. 2). The following lemma studies the property of the first hitting time $\mathcal{T}_h(v_i, v_j)$ between v_i and v_j with neighboring home points.

Theorem 2: Given secondary users v_i and v_j in a finite CRN $[\mathcal{F}_{m,n}, \mathcal{M}(\Phi_E, \Psi, \alpha), (\lambda, \lambda_p)]$ with $d_c(v_i, v_j) = \sqrt{\frac{1}{\lambda}}$, we have i) $\mathbb{P}(\mathcal{T}_h(v_i, v_j) = \infty) = 1$ if $\alpha < \frac{\sqrt{\frac{1}{\lambda}} - r}{2}$; ii) otherwise, $E(T_h(v_i, v_j)) < \infty$ and $\mathbb{P}(T_h(v_i, v_j) > t) \le e^{-c_1 t}$ for sufficiently large t and some positive constant c_1.

Proof: At time t if and only if $d^{(t)}(v_i, v_j) < r$, nodes v_i and v_j may communicate directly. Also, $d^{(t)}(v_i, v_j) > d_h(v_i, v_j) - 2\alpha$ for all t. If $\alpha < \frac{\sqrt{\frac{1}{\lambda}} - r}{2}$, $d^{(t)}(v_i, v_j) > r$ for all t, which implies that v_i and v_j cannot communicate with each other. This completes the proof of part i).

For $\alpha > \frac{\sqrt{\frac{1}{\lambda}} - r}{2}$, let $\mathcal{E}_t$ denote the event that there exists no communication link between v_i and v_j at time t and $\bar{\mathcal{E}}_t$ as its complement. As shown in Fig. 3, a necessary condition for $\bar{\mathcal{E}}_t$ is that there exist no primary users on channel ch_k, for some $1 \le k \le m$, within the circle centered at o with radius R_I, and a sufficient condition for $\bar{\mathcal{E}}_t$ is that v_i lies in the shaded region S_1, v_j in S_2 and no primary users on ch_k in the bigger circle centered at o for some k. Note that for any $1 \le k \le m$, the number of primary users on ch_k is $\lambda_p \frac{n}{\lambda}$ and the probability that no primary users on ch_k is located within a circle with radius e is given by $\mathbb{P}_a(e) = (1 - \frac{\lambda \pi e^2}{n})^{\frac{n\lambda_p}{\lambda}}$. Therefore,

$$\begin{aligned}0 &< \Big(1 - (1 - \mathbb{P}_a(R_I + r/2))^m\Big)\bar{\Psi}(S_1)\bar{\Psi}(S_2) \\ &< 1 - \mathbb{P}(\mathcal{E}_t) = \mathbb{P}(\bar{\mathcal{E}}_t) < 1 - (1 - \mathbb{P}_a(R_I))^m < 1 \quad (1)\end{aligned}$$

where $\Big(1 - (1 - \mathbb{P}_a(R_I + r/2))^m\Big)\bar{\Psi}(S_1)\bar{\Psi}(S_2)$ characterizes the probability of the sufficient condition and $1 - (1 - \mathbb{P}_a(R_I))^m$ characterizes the probability of the necessary condition for $\bar{\mathcal{E}}_t$, respectively, and $\bar{\Psi}(S) = \int_S \Psi dS$. To proceed, we next find an index set $\mathcal{I}$ such that $\{\mathcal{E}_t\}_{t \in \mathcal{I}}$ are independent and let $\varepsilon = \mathbb{P}(\mathcal{E}_t)$ for convenience.

Denote η as a *renewal* interval for secondary users, i.e., for any $t > 0$, $\{v_i(t') : t' \le t\}$ and $\{v_i(t'' + \eta) : t'' \ge t\}$ are independent. And by the system model defined in the Section II, the *renewal* for primary users is 1 second. Define $\rho = \max\{\eta, 1\}$ and ρ is the *renewal* for the CRN. That is, after ρ seconds, primary and secondary users completely forget where they were ρ seconds ago. Denote $\{\rho_i\}_{i=1}^{\infty}$ as a sequence of i.i.d random variables with the same distribution as ρ. Now we consider the index set $\mathcal{I}_t = \{0, t_1, \ldots, t_{N(t)}\} \subset (0, t]$, where $t_k = \sum_{i=1}^{k} \rho_i$ and $N(t) = |\mathcal{I}_t| = \max\{k : t_k \le t\}$. Observe that

$$\mathbb{P}(T_h(v_i, v_j) > t) \le \mathbb{P}(\cap_{s \in \mathcal{I}_t} \mathcal{E}_s) = \prod_{s \in \mathcal{I}_t} \mathbb{P}(\mathcal{E}_s) \quad (2)$$

where the last equality is by the independency of $\{\mathcal{E}_s\}_{s \in \mathcal{I}_t}$ and by conditioning on $N(t)$, we have

$$\mathbb{P}(T_h(v_i, v_j) > t) \le E(\varepsilon^{N(t)}) = E(e^{-\beta N(t)}),$$

where $\beta = -\log \varepsilon > 0$. In addition, for any $\tau > 0$,

$$\begin{aligned}&E(e^{-\beta N(t)}) = E(e^{-\beta N(t)} I_{\{N(t) \le \tau t\}}) + \\ &E(e^{-\beta N(t)} I_{\{N(t) \le \tau t\}}) \le \mathbb{P}(N(t) \le \tau t) + e^{-\beta \tau N(t)}.\end{aligned}$$

Note that the finite sum of exponentially bounded random variables is still exponentially bounded [15], [16]. Thus, if we can show that $\mathbb{P}(N(t) \le \tau t)$ is exponentially bounded, we will finish the proof. In order to proceed, we assume that the tails of renewals $\{\rho_i\}_{i=1}^{\infty}$ are exponentially bounded. This assumption is reasonable considering the network is finite, which has been well-explained in many mobility models [5], [10], [11], [13], [16]. We next show that $\mathbb{P}(N(t) \le \tau t)$ is exponentially bounded. By letting $k = \tau t$, $\mathbb{P}(N(t) \le \tau t) = \mathbb{P}(\sum_{i=1}^{\tau t} \rho_i > t) = \mathbb{P}(\sum_{i=1}^{k} \rho_i > \frac{k}{\tau})$. The last item is obviously bounded by some exponential variable considering $\{\rho_i\}_{i=1}^{\infty}$ are exponentially bounded. This completes the proof. ■

We next present our main result on the tail distribution of the dissemination latency $\mathcal{T}_d$ under EIHP mobility.

Proposition 1: Given $[\mathcal{F}_{m,n}, \mathcal{M}(\Phi_E, \Psi, \alpha), (\lambda, \lambda_p)]$ with finite users, if $\alpha > \frac{\sqrt{\frac{1}{\lambda}} - r}{2}$, the tail distribution of the dissemination latency $\mathcal{T}_d$ is stochastically dominated by a Gamma distribution, $\Gamma(2\sqrt{n}, c_2)$; otherwise, $\mathcal{T}_d$ has a *heavy-tailed* distribution and $\mathbb{P}(\mathcal{T}_d = \infty) > 0$.

Proof: (Sketch.) As the end to end latency, $\mathcal{T}_d$ is clearly bounded by the transmission delay along any path from the source v_s to destination v_d. Theorem 2 shows that, if $\alpha > \frac{\sqrt{\frac{1}{\lambda}} - r}{2}$, a link exists between two neighboring secondary users with positive probability. Therefore, we can identify a *Manhattan path* through which v_s first transmits the message vertically until the message reaches the secondary user whose center point has the same horizontal coordinate with v_d^c, and then transmits the message horizontally to v_d as shown in Fig. 4. Denote $\{X_k\}_{k=1}^{\infty}$ as a sequence of random variables with identical distributions as the first hitting time between neighboring secondary users. Note that a *Manhattan path* consists of at most $2\sqrt{n}$ communication links and thus $\mathcal{T}_d \le \sum_{k=1}^{2\sqrt{n}} X_k$.

The next challenge is that the first hitting time of neighboring links, i.e., X_i and X_{i+1} are not independent. To tackle this challenge, we assume that after receiving the message, each secondary user will hold this message for a *renewal* time ρ before it tries to relay the message. Let $\{\rho_i\}_{i=1}^{\infty}$ be a sequence of *renewals* and $Y_k = X_k + \rho_k$. It is clear that $\mathcal{T}_d \le \sum_{k=1}^{2\sqrt{n}} Y_k$. Note that Y_k is bounded by $exponential(c_2)$ (since both X_k and ρ_k are both exponentially bounded) and $\{Y_k\}_{k=1}^{\infty}$ are clearly independent. Let $\{\hat{Y}_k\}_{k=1}^{\infty}$ be a sequence of independent random variables distributed as $exponential(c_2)$, we have

$$\mathbb{P}(\mathcal{T}_d > t) \le \mathbb{P}(\sum_{k=1}^{2\sqrt{n}} Y_k > t) \le \mathbb{P}(\sum_{k=1}^{2\sqrt{n}} \hat{Y}_k > t), \quad (3)$$

where the last inequality is from Lemma 1 and *coupling* (Definition 3). By the *moment generating function* technique, we know that Y follows a Gamma distribution, $\Gamma(2\sqrt{n}, c_2)$. This completes the proof for $\mathcal{T}_d$.

When $\alpha < \frac{\sqrt{\frac{1}{\lambda}} - r}{2}$, Theorem 2 says that the first hitting time between any two secondary users $\mathcal{T}_h(v_i, v_j) = \infty$. Therefore, $\mathcal{T}_d = \infty$, which completes the proof. ■

B. Distribution of $\mathcal{T}_d$ under PIHP Mobility $\mathcal{M}(\Phi_P, \Psi, \alpha)$

Note that the main difference between PIHP and EIHP mobility is that home points in the former are randomly located, and thus for neighboring secondary users v_i and v_j with $d_c(v_i, v_j) = \sqrt{\frac{1}{\lambda}}$, $d_h(v_i, v_j) \ne \sqrt{\frac{1}{\lambda}}$ in PIHP mobility. But under PIHP mobility, $d_h(v_i, v_j)$ is still bounded and we

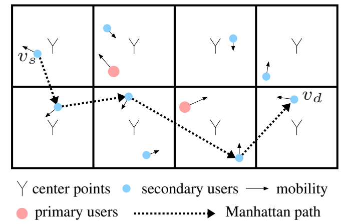


Fig. 4. A Manhattan path between v_s and v_d under EIHP mobility.

have $\mathbb{P}(d_h(v_i, v_j) \le \sqrt{\frac{5}{\lambda}}) = 1$. Thus, by similar proof to Theorem 2, we are able to see that for any v_i and v_j with $d_c(v_i, v_j) = \sqrt{\frac{1}{\lambda}}$, if $\alpha > \frac{\sqrt{\frac{5}{\lambda}} - r}{2}$, the first hitting time $\mathcal{T}_h(v_i, v_j)$ is *exponentially* bounded; and if $\alpha < \frac{\sqrt{\frac{5}{\lambda}} - r}{2}$, $\mathbb{P}(\mathcal{T}_h(v_i, v_j) = \infty) > 0$. Therefore, through the similar proof as that of Proposition 1, we have the following result:

Proposition 2: Given $[\mathcal{F}_{m,n}, \mathcal{M}(\Phi_P, \Psi, \alpha), (\lambda, \lambda_p)]$ with finite users, if $\alpha > \frac{\sqrt{\frac{5}{\lambda}} - r}{2}$, the tail distribution of $\mathcal{T}_d$ is stochastically dominated by a Gamma distribution $\Gamma(2\sqrt{n}, c_3)$ for some positive constant c_3; otherwise, $\mathcal{T}_d$ is *heavy tail* distributed and $\mathbb{P}(\mathcal{T}_d = \infty) > 0$.

C. Distribution of $\mathcal{T}_d$ under HHP Mobility $\mathcal{M}(\Phi_H, \Psi, \alpha)$

When HHP mobility is considered, home points are homogeneously distributed in the whole network Ω_n. Therefore, the distance between home points of secondary users v_i and v_j is homogeneous and may be any value in the interval $(0, \sqrt{\frac{n}{2\lambda}})$ (see Fig. 5; note that Ω_n is a torus surface without border effect). We next show that, to overcome the *randomness* of $d_h(v_i, v_j)$, secondary users need to move over the whole network (large mobility capability α) to eliminate the *heavy tail* of the first hitting time.

Lemma 3: Given a CRN $[\mathcal{F}_{m,n}, \mathcal{M}(\Phi_H, \Psi, \alpha), (\lambda, \lambda_p)]$ with finite users, if $\alpha > \frac{\sqrt{\frac{n}{2\lambda}} - r}{2}$, the first hitting time $T_h(v_i, v_j)$ is exponentially bounded; otherwise, $T_h(v_i, v_j)$ has a heavy tail and $\mathbb{P}(T_h(v_i, v_j) = \infty) > 0$.

Proof: As shown in Fig. 5, there may exist a communication link between v_i and v_j (i.e., $\mathbb{P}(\mathcal{T}_h(v_i, v_j) < \infty) = 1$), if and only if v_j^h is located in the solid circle $\mathcal{C}$ centered at v_i^h. Thus the probability that v_j^h is distributed outside $\mathcal{C}$ (i.e., $\mathbb{P}(\mathcal{T}_h(v_i, v_j) = \infty) > 0$), will incur a heavy tail of $\mathcal{T}_h(v_i, v_j)$ (that is, $E(\mathcal{T}_h(v_i, v_j)) = \infty$). Therefore, to eliminate the *heavy tail*, $\mathcal{C}$ must cover the whole network Ω_n, which requires $2\alpha + r > \sqrt{\frac{n}{2\lambda}} \Rightarrow \alpha > \frac{\sqrt{\frac{n}{2\lambda}} - r}{2}$. When $\alpha > \frac{\sqrt{\frac{n}{2\lambda}} - r}{2}$, v_i and v_j may communicate with each other with some positive probability at any time. Hence, with the similar proof of Theorem 2, we can show that $T_h(v_i, v_j)$ is exponentially bounded. This completes the proof. ■

Under HHP mobility, any secondary user v_i may receive the message directly from the source v_s, and any v_i that carries the message may in turn copy this message to all secondary users it encounters along its trajectory. Hence, we cannot apply the coupling method in calculating $\mathcal{T}_d$ hop by hop along the end

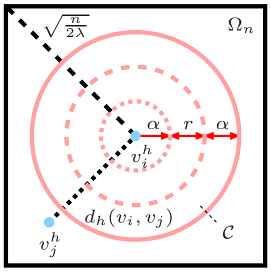


Fig. 5. An illustration of the first hitting time in HHP mobility.

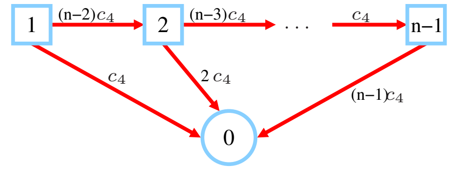


Fig. 6. Illustration of the Markov chain. Each state represents the number of secondary users carrying the message.

to end path for HHP mobility. Instead, we use a stochastic model to analyze $\mathcal{T}_d$ and obtain the following result:

Proposition 3: Given $[\mathcal{F}_{m,n}, \mathcal{M}(\Phi_H, \Psi, \alpha), \mathcal{M}_H]$ with finite users, if $\alpha < \frac{\sqrt{\frac{n}{2\lambda}}-r}{2}$, $\mathcal{T}_d$ has a *heavy-tailed* distribution and $\mathbb{P}(\mathcal{T}_d = \infty) > 0$; and if $\alpha > \frac{\sqrt{\frac{n}{2\lambda}}-r}{2}$, the tail of $\mathcal{T}_d$ is stochastically dominated by a *Gamma distribution*.

Proof: When $\alpha < \frac{\sqrt{\frac{n}{2\lambda}}-r}{2}$, there exists some positive probability that all the home points $\{v_j^h, j \neq s\}$ are located outside the circle centered at v_s^h with radius $2\alpha + r$, which implies $\mathbb{P}(\mathcal{T}_d = \infty) > 0$ and thus a heavy tail.

When $\alpha > \frac{\sqrt{\frac{n}{2\lambda}}-r}{2}$, Lemma 3 shows that the tail of the first hitting time $\mathcal{T}_h(v_i, v_j)$ is stochastically dominated by $exponential(c_4)$ for some constant c_4. If we can show that when $\mathcal{T}_h(v_i, v_j)$ is distributed as $exponential(c_4)$, the tail of the resulting dissemination latency $\mathcal{T}_d'$ is stochastically dominated by a *Gamma distribution*, then by Lemma 1 and *coupling*, which shows that $\mathbb{P}(\mathcal{T}_d > t) < \mathbb{P}(\mathcal{T}_d' > t)$ for large t, we complete the proof.

Assume $\mathcal{T}_h(v_i, v_j) \sim exponential(c_4)$ for any v_i and v_j. Denote by ζ the number of secondary users, which carry the message sent by source v_s before this message is successfully delivered to the destination v_d. The proof is based on modeling ζ as an absorbing finite-state Markov chain. The Markov chain consists of states $k = 0, 1, 2, \ldots, n-1$. The state $k > 0$ denotes $\zeta = k$ and the state 0 denotes the absorbing state that v_d successfully receives the message.

When secondary users hit each other, messages will be forwarded to the ones without a copy of the message. Therefore, when there are k secondary users carrying the message, the message is sent to another secondary user at rate $c_4k(n-1-k)$ with the transition from k to $k+1$, and to destination v_d at rate c_4k with transition from k to 0, as shown in Fig. 6. The chain jumps from state k to $k+1$ with probability $\frac{n-1-k}{n-k}$ and transits from k to 0 with probability $\frac{1}{n-k}$. The sojourn time S_k in state k is exponentially distributed with intensity $c_4k(n-k)$. Since $S_1, S_2, \ldots, S_{n-1}$ are mutually independent. Thus $\mathbb{P}(\zeta = k) = \frac{1}{n-k}\prod_{j=1}^{k-1}\frac{n-1-j}{n-j} = \frac{1}{n-1}$. Conditioning $\mathcal{T}_d'$ on ζ, we have $\mathcal{T}_d'|(\zeta = k) = \sum_{j=1}^{k} S_j$, which is clearly exponentially bounded and therefore, $\mathbb{P}(\mathcal{T}_d' > t) = \sum_{k=1}^{n-1}\mathbb{P}(\zeta = k)\mathbb{P}(\mathcal{T}_d'|(\zeta = k) > t) = \frac{1}{n}\sum_{k=1}^{n-1}\mathbb{P}(\mathcal{T}_d'|(\zeta = k) > t)$ is exponentially bounded (note that the sum of exponential variables is still exponentially distributed). That is, the tail of $\mathcal{T}_d'$ is bounded by $exponential(c_5) = Gamma(1, c_5)$ for some positive constant c_5. This completes the proof. ■

Based on Propositions 1, 2 and 3, we summerize the fundamental property of the latency $\mathcal{T}_d$ as follows:

Theorem 4: In a CRN $[\mathcal{F}_{m,n}, \mathcal{M}(\Phi, \Psi, \alpha), (\lambda, \lambda_p)]$ with finite users, there exists a **cutoff point** on the *mobility range* α, above which the tail distribution of dissemination latency $\mathcal{T}_d$ is bounded by some *Gamma* distribution; below which $\mathcal{T}_d$ has a *heavy-tailed* distribution and $\mathbb{P}(\mathcal{T}_d = \infty) > 0$.

Remark 3: A *heavy tail* of the dissemination latency $\mathcal{T}_d$ implies a significant probability that it takes long time to disseminate a message from the source to the destination. Thus a *light-tailed* dissemination latency $\mathcal{T}_d$ is crucial for time-critical applications in CRNs. Theorem 4 tells that to achieve a *light-tailed* dissemination latency (note that *Gamma distribution* is a type of light-tailed distribution), the *mobility radius* of secondary users α need to be larger than some cutoff point, which is specifically identified in Proposition 1 for EIHP, Proposition 2 for PIHP and Proposition 3 for HHP, respectively. This result encourages the existing endeavor of deploying CRN for practical applications, including time-critical applications, such as emergency networks and military networks. In addition, note that the proofs of Propositions 1, 2 and 3 are independent of the spatial density λ_p of primary users. We need to clarify here that this "independency" is related to the *stochastic dichotomy*. However, λ_p may have a negative impact on the dissemination latency $\mathcal{T}_d$(see simulation results in Fig. 8). Particularly, as shown in the proof of Theorem 2, the expected first hitting time $E(\mathcal{T}_h(v_i, v_j))$ and thus the expected dissemination latency $E(\mathcal{T}_d)$ are obviously increasing functions of the probability $\mathbb{P}(\mathcal{E}_t)$ that no communication link between v_i and v_j at time t, which increases as λ_p increases (see Eq. (1)).

IV. The Scalability of $\mathcal{T}_d$ in Large CRNs

We next study the dissemination latency $\mathcal{T}_d$ in large mobile CRNs. Our study on the distribution of $\mathcal{T}_d$ in Section III implies that $E(\mathcal{T}_d) \to \infty$ as the network size grows large, which indicates that the distribution of $\mathcal{T}_d$ cannot be used to measure *how fast* information is disseminated in large CRNs. Therefore, in large CRNs, we investigate the *sclability* of $\mathcal{T}_d$ with respect to the *distance* $\mathcal{D}$ between the source v_s and destination v_d. First, we present the following theorem, which is helpful to our analysis:

Theorem 5 (Liggett's subadditive ergodic theorem, [17]): Let $\{\mathcal{Z}_{h,q}\}$ be a collection of random variables indexed by integers satisfying $0 \leq h < q$. Suppose $\{\mathcal{Z}_{h,q}\}$ has the following properties: (i) $\mathcal{Z}_{0,q} \leq \mathcal{Z}_{0,h} + \mathcal{Z}_{h,q}$; (ii) For each q,

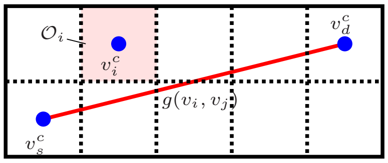


Fig. 7. Illustration of the dissemination direction.

$\mathbb{E}(|\mathcal{Z}_{0,q}|) < \infty$ and $\mathbb{E}(\mathcal{Z}_{0,q}) \geq cq$ for some constant $c > -\infty$; (iii) The distribution of $\{\mathcal{Z}_{h,h+k:k\geq 1}\}$ does not depend on h; (iv) For each $k \geq 1$, $\{\mathcal{Z}_{qk,(q+1)k} : q \geq 0\}$ is a stationary sequence; (v) If $k \geq 1$, $\{\mathcal{Z}_{qk,(q+1)k} : q \geq 0\}$ are ergodic. Then we have (a) $\zeta = \lim_{q\to\infty} \mathbb{E}(\mathcal{Z}_{0,q})/q = inf_{q\geq 1} E(\mathcal{Z}_{0,q})/q$; (b) $\mathcal{Z} = \lim_{q\to\infty} \mathcal{Z}_{0,q}/q$ exists almost surely; (c) $\mathbb{E}(\mathcal{Z}) = \zeta$; and (d) $\mathcal{Z} = \zeta$ almost surely.

A. Scalability of $\mathcal{T}_d$ under EIHP and PIHP Mobility

Since the study of the first hitting time $\mathcal{T}_h(v_i, v_j)$ between neighboring secondary users v_i and v_j under EIHP or PIHP models in Section III is independent of the network size, these results still hold for large CRNs. That is, when the *mobility radius* $\alpha > \frac{\sqrt{\frac{1}{\lambda}}-r}{2}$ for EIHP or $\alpha > \frac{\sqrt{\frac{5}{\lambda}}-r}{2}$ for PIHP mobility, the first hitting time $\mathcal{T}_h(v_i, v_j)$ is exponentially bounded. Otherwise, $\mathcal{T}_h(v_i, v_j)$ and thus $\mathcal{T}_d$ have heavy tails independent of the transmission distance $\mathcal{D}$(i.e., $\mathcal{T}_d$ is unscalable with $\mathcal{D}$). Therefore, we only need to study the scalability of $\mathcal{T}_d$ with *exponentially bounded* $\mathcal{T}_h(v_i, v_j)$. In addition, when it comes to the *distance* $\mathcal{D}$ between the source v_s and destination v_d, it can be any *p-norm* metric function and we consider two of the most popular metrics *transmission hops* and *Euclidean distance*. As analyzed in Section III-A and III-B, hop by hop communication is necessary for **EIHP** and **PIHP** mobility, which indicates that *transmission hops* can describe "how far" more accurately than the Euclidean distance in these two models. Therefore, $\mathcal{D}$ here denotes the *Manhattan distance* between v_s^c and v_d^c by which the maximum number of transmission hops between v_s and v_d can be expressed as $\sqrt{\lambda}\mathcal{D}$. We next present our main result.

Proposition 4: Given $\alpha > \frac{\sqrt{\frac{1}{\lambda}}-r}{2}$ for a large network $[\mathcal{F}_{m,n}, \mathcal{M}(\Phi_E, \Psi, \alpha), (\lambda, \lambda_p)]$ (or $\alpha > \frac{\sqrt{\frac{5}{\lambda}}-r}{2}$ for $[\mathcal{F}_{m,n}, \mathcal{M}(\Phi_P, \Psi, \alpha), (\lambda, \lambda_p)]$), there exists some finite constant κ such that $\mathbb{P}(\lim_{\mathcal{D}\to\infty} \frac{\mathcal{T}_d}{\mathcal{D}} = \kappa) = 1$.

To initiate the proof of Proposition 4, we first define the following notations. Denote $d_c^{(1)}(v_i, v_j)$ as the *Manhattan distance* between center points v_i^c and v_j^c for any v_i and v_j. Let N_h be the set of secondary users defined as,

$$N_h \triangleq \{v_i : d_c^{(1)}(v_s, v_i) = h\sqrt{\frac{1}{\lambda}}\},$$

where $\sqrt{\frac{1}{\lambda}}$ is the *Manhattan distance* between center points of neighboring secondary users (see Fig. 2). The *information dissemination direction* from the source v_s to destination v_d is denote by $g(v_s, v_d)$, which is the straight line joining v_s^c and v_d^c as shown in Fig. 7. We denote $v(h)$ as the secondary user whose *distance* to v_s is $h\sqrt{\frac{1}{\lambda}}$ and which is in the information dissemination direction, i.e.,

$$v(h) \triangleq \{v_i : v_i \in N_h \text{ and } g(v_s, v_d) \cap \mathcal{O}_i \neq \emptyset\},$$

where $\mathcal{O}_i$ is the cell associated with v_i^c as shown in Fig. 7. We next define the collection of indexed variables by $\mathcal{T}_{h,q}$ as the *dissemination latency* from nodes $v(h)$ to $v(q)$ (thus $\mathcal{T}_d = \mathcal{T}_{0,\mathcal{D}\sqrt{\lambda}}$). Therefore, Proposition 4 is equivalent to showing $\mathbb{P}(\lim_{q\to\infty} \frac{\mathcal{T}_{0,q}}{q} = \kappa\sqrt{\lambda}) = 1$, which can be proved by Liggett's theorem.

Particularly, if we can show that the sequence $\{\mathcal{T}_{h,q}, h \leq q\}$ satisfies the conditions $(i) - (v)$ of Liggett's theorem, we can finish our proof. By definition, $\mathcal{T}_{0,q}$ is the shortest time that $v(q)$ will receive the message from $v(0)$, which is clearly at most $\mathcal{T}_{0,h} + \mathcal{T}_{h,q}$. Condition (i) is thus verified. As the latency cannot be negative, we have $E(\mathcal{T}_{0,q}) > 0$. To compute an upper bound of $E(\mathcal{T}_{0,q})$, we consider a Manhattan path between nodes $v(0)$ and $v(q)$ (see Fig.4) and thus have $E(\mathcal{T}_{0,q}) \leq qE(\mathcal{T}_h(v_i, v_j))$. Theorem 2 tells that the first hitting time $\mathcal{T}_h(v_i, v_j)$ is exponentially bounded and thus $E(\mathcal{T}_{0,q}) < \infty$. This attests condition (ii). Conditions (iii) and (iv) are clearly satisfied, as $\mathcal{T}_{h,q}$ is defined in a stationary way. The following lemma is to prove that the sequence $\mathcal{T}_{h,q}$ is ergodic, i.e., $\{\mathcal{T}_{h,q}, h \leq q\}$ satisfies condition (v). Particularly, it shows that $\mathcal{T}_{h,q}$ is mixing (i.e., roughly speaking, asymptotically independent), which is a stronger property than ergodicity.

Lemma 6: The sequence $\{\mathcal{T}_{q,q+1}, q \geq 0\}$ is mixing.

Proof: We compute $\mathcal{T}_{q,q+1}$ by the following construction: Denote $\mathcal{N}_{q,k}$ as the set of secondary users whose distance to $v(q)$ is less than $k\sqrt{\frac{1}{\lambda}}$, that is, $\mathcal{N}_{q,k} \triangleq \{v_i : d_c^{(1)}(v(q), v_i) < k\sqrt{\frac{1}{\lambda}}\}$ and $\mathcal{T}_{q,q+1}^{(k)}$ as the transmission delay from $v(q)$ to $v(q+1)$ where nodes $v \in \mathcal{N}_{q,k}$ are used as relays. Observe that

$$\lim_{k\to\infty} \mathbb{P}\Big(\mathcal{T}_{q,q+1}^{(k)} < t\Big) = \mathbb{P}\Big(\mathcal{T}_{q,q+1} < t\Big)$$

for all t. Thus $\{\mathcal{T}_{q,q+1}\}$ is mixing by

$$\begin{aligned} &\lim_{k\to\infty} \mathbb{P}\Big((\mathcal{T}_{q,q+1} < t) \cap (\mathcal{T}_{q+2k,q+2k+1} < t')\Big) \\ &= \lim_{k\to\infty} \mathbb{P}\Big((\mathcal{T}_{q,q+1}^{(k)} < t) \cap (\mathcal{T}_{q+2k,q+2k+1}^{(k)} < t')\Big) \\ &= \lim_{k\to\infty} \mathbb{P}\Big(\mathcal{T}_{q,q+1} < t\Big)\mathbb{P}\Big(\mathcal{T}_{q+2k,q+2k+1}^{k} < t'\Big) \quad \forall t, t'. \end{aligned}$$

The second equality follows that $\mathcal{T}_{q,q+1}^{(k)}$ and $\mathcal{T}_{q+2k,q+2k+1}^{k}$ are independent, as they depend on non-intersected node sets $\mathcal{N}_{q,k}$ and $\mathcal{N}_{q+2k,k}$. ■

Putting all together, We conclude that $\{\mathcal{T}_{h,q}, h \leq q\}$ satisfies all the conditions of Liggett's theorem and thus prove Proposition 4. Since the proof for the scalability of $\mathcal{T}_d$ under PIHP mobility is similar, we omit the details. Next we study the scalability of $\mathcal{T}_d$ under HHP mobility.

B. Scalability of $\mathcal{T}_d$ under HHP Mobility

Connectivity is a prerequisite for network applications to ensure that information can be disseminated to the entire network. Note that in the previous analysis of latency in large CRNs under **EIHP** and **PIHP** mobility, a fundamental requirement is that the *mobility radius* α is larger than some critical values, which is actually used to ensure that the networks are *fully connected*. That is, there exists a communication path (may be dynamic over time) between any two secondary users. By techniques used to derive the *full connectivity* in its counterpart in homogeneous networks [14], we can identity that $2\alpha + r = \Theta(\sqrt{\log n})$ [1] is required for a fully connected CRN under HHP mobility, which is impractical to be satisfied when the number of nodes n is large. Therefore, we consider a *percolated* network [3], [10], in which there exists a *giant component* consisting of $\Theta(n)$ nodes well scattered over the whole network. In a percolated network, information can still be be disseminated to the entire network through the giant component. The results in *continuum percolation* [3], [10] show that given the spatial densities (λ, λ_p), finite $2\alpha + r$ is enough for a percolated CRN under **HHP** mobility.

In terms of the latency scalability in a percolated CRN under HHP mobility, a similar problem has been studies in its counterpart in homogeneous networks [10]. In particular, Kong and Yeh [10] show that in a percolated mobile homogeneous network, where wireless nodes are mobile under *constrained i.i.d mobility*, i.e, home points are independently and uniformly distributed and any node v_i is mobile within the circular region $\mathcal{A}(v_i^h, \alpha)$ centered at its home point v_i^h with radius α according to an uniform stationary distribution, the latency asymptotically scales linearly with the transmission distance (Euclidean). To make the best use of the existing findings, we compare our CRN model $[\mathcal{F}_{m,n}, \mathcal{M}(\Phi_H, \Psi, \alpha), (\lambda, \lambda_p)]$ with the homogeneous network studied in [10] and identify two main differences. First, in our model, a node v_i is independently mobile within $\mathcal{A}(v_i^h, \alpha)$ according to an arbitrary stationary distribution Ψ (not necessarily the *uniform* distribution). Second, primary users are present in our CRNs, which constrains the communications among secondary users. However, we find that the proofs in [10] require neither *uniform* distribution of v_i around v_i^h, nor non-interference from other nodes (e.g., primary users). Indeed, the fundamental requirement for proofs in [10] is that, given any two nodes v_i and v_j with $d_h(v_i, v_j) = \|v_i^h - v_j^h\| < 2\alpha + r$, the expected first hitting time $E(\mathcal{T}_h(v_i, v_j)) < \infty$. Fortunately, our earlier study shown in Theorem 2 can satisfy this condition. Therefore, we are able to extend the results [10] to our network model $[\mathcal{F}_{m,n}, \mathcal{M}(\Phi_H, \Psi, \alpha), \mathcal{M}_H]$. This leads to the linear scaling law of information dissemination latency with respect to transmission distance (Euclidean). Particularly, we have

Proposition 5: Given any two nodes v_s and v_d in the *giant component* of a percolated CRN under HHP mobility $[\mathcal{F}_{m,n}, \mathcal{M}(\Phi_H, \Psi, \alpha), (\lambda, \lambda_p)]$, there exists a finite and positive constant κ such that $\mathbb{P}(\lim_{\mathcal{D}\to\infty} \frac{\mathcal{T}_d}{\mathcal{D}} = \kappa) = 1$.

Based on Proposition 4 and 5, we summarize the latency scalability in large mobile CRNs as follows:

Theorem 7: Given a large *connected* (fully connected or percolated) mobile CRN $[\mathcal{F}_{m,n}, \mathcal{M}(\Phi, \Psi, \alpha), (\lambda, \lambda_p)]$, there exists a finite constant κ such that $\mathbb{P}(\lim_{\mathcal{D}\to\infty} \frac{\mathcal{T}_d}{\mathcal{D}} = \kappa) = 1$.

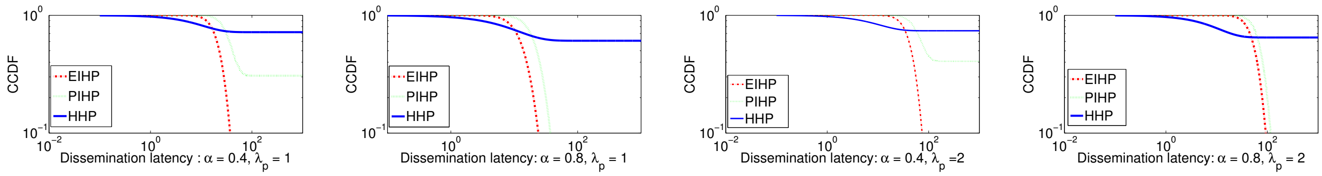
Remark 4: Scalability has been one of the most fundamental problems that has discouraged the deployment of large wireless networks [6], [14]. Theorem 7 reveals that in large connected CRNs, the dissemination latency $\mathcal{T}_d$ asymptotically scales linearly with the initial *distance* between the source and destination, i.e., the message sent by a source reaches its destination at a fixed asymptotic speed. This result enables the feasible deployment of CRNs for large applications. Moreover, note that the proofs for the *linear scalability* of the latency in Propositions 4 and 5 are invariant to the spatial density of primary users λ_p, which indicates that the *linearity* of the latency, but not the latency itself, is independent of primary users. Particularly, as shown in the proof of Proposition 4, the asymptotic value $\lim_{\mathcal{D}\to\infty} \frac{\mathcal{T}_d}{\mathcal{D}}$ is proportional to the expected first hitting time $E(T_h(v_i, v_j))$, which implies that λ_p (see Theorem 2) has an adverse impact on the dissemination latency. This has been validated by simulations in Fig. 9. Furthermore, a prerequisite for Proposition 5 is that the network is percolated and the reference [3] shows that when the density of primary users is larger than some critical value, the CRN cannot percolate. Thus a subtle assumption for Proposition 5 is that the density λ_p is smaller than such critical value.

V. Simulation

In this section we provide simulation results to support our theoretical analysis on *distribution* and *scalability* of latency in finite and infinite CRNs, respectively. In these simulations, time is partitioned into unit slots and in each time slot, primary users are uniformly distributed at random and secondary users are uniformly distributed around their home points (i.e., Ψ is uniform). Furthermore, home points are uniformly distributed around the center points under PIHP mobility (i.e., Φ_P is uniform). The transmission range r of secondary users and the interference range R_I of primary users are set as $r = 0.1$ kilometer (km) and $R_I = 0.3$ (km), respectively. Secondary users opportunistically access $m = 2$ channels.

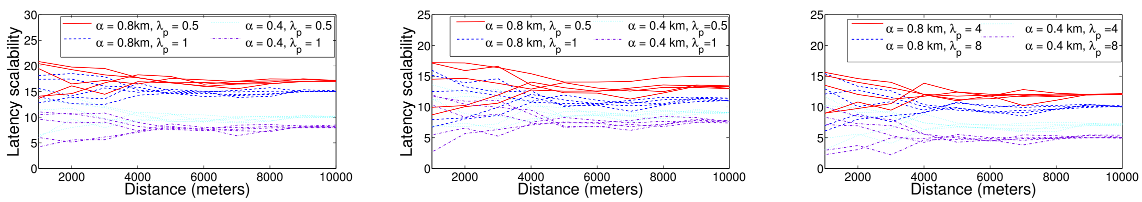
We first study a finite CRN where $n = 16$ secondary users are mobile within an 2×2 (km^2) area (i.e., $\lambda = 4$ per km^2). Fig. 8 illustrates the complementary distribution (CCDF) of the latency $\mathbb{P}(\mathcal{T}_d > t)$ on a log-log scale for EIHP, PIHP and HHP models with different values of the *mobility radius* α and the spatial density λ_p of primary users. The probability is calculated based on the average of 1000 independent simulations. It is observed in Fig. 8 that as λ_p increases, the curves move right-ward, which indicates the increasing expected dissemination latency. However, regardless of the value of λ_p, when $\alpha = 0.4$ (km), which is larger than the *cutoff point* under *EIHP* but smaller than those under PIHP and HHP, the dissemination latency $\mathcal{T}_d$ has a light tail under EIHP but *heavy tails* under PIHP and HHP. As α increases to 0.6 (km), which is larger than the *cutoff point* in PIHP, but still less than that in HHP, the *heavy tail* of $\mathcal{T}_d$ in PIHP disappears, but $\mathcal{T}_d$ in HHP presents a *heavy* tail. These results are in good agreement with our theoretical analysis in Propositions 1, 2 and 3, and arguments in Remark 3.

[1]Given any function $f(n)$ and $g(n)$, $f(n) = O(g(n)) \leftrightarrow \limsup_{n\to\infty} \frac{f(n)}{g(n)} < \infty$; $f(n) = \Theta(g(n)) \leftrightarrow f(n) = O(g(n))$ and $g(n) = O(f(n))$; n is the number of nodes.



(a) Log-Log scale (b) Log-Log scale (c) Log-Log scale (d) Log-Log scale

Fig. 8. CCDF of dissemination latency $\mathcal{T}_d$ under general mobility.



(a) EIHP (b) PIHP (c) HHP

Fig. 9. The latency scalability $\frac{\mathcal{T}_d}{\mathcal{D}}$ (s/km) based on 5 independent simulations.

We further perform a series of simulations to validate our asymptotic results in large networks. Fig. 9(a) and 9(b) show the latency *scalability* in large CRNs under EIHP and PIHP models, respectively, where the spatial density of secondary users is $\lambda = 4$ (per km^2). As shown in 9(a) and 9(b), no matter how large the *mobility radius* α is, the dissemination latency $\mathcal{T}_d$ scales linearly with the dissemination distance $\mathcal{D}$ (Manhattan distance). Moreover, The *latency scalability* in a large percolated CRN under HHP mobility, where the spatial density of secondary users is set as $\lambda = 200$ (per km^2) to ensure percolation, is shown in Fig. 9(c), which shows that in percolated CRNs, the dissemination latency $\mathcal{T}_d$ scales linearly with the dissemination distance $\mathcal{D}$ (Euclidean distance) as $\mathcal{D}$ increases. In addition, as shown in Fig. 9, the *scalability* decreases as the spatial density λ_p increases. These observations provide a straightforward illustration of Propositions 4 and 5 and arguments in Remark 4.

VI. Conclusions

We have studied in this paper the distribution of the information dissemination latency $\mathcal{T}_d$ in *finite* CRNs and the scalability of $\mathcal{T}_d$ in *large* CRNs under general mobility. We found that in finite networks, there exists a cutoff point on the *mobility radius* α of secondary users, above which the tail distribution of $\mathcal{T}_d$ is bounded by some Gamma distribution and below which $\mathcal{T}_d$ has a *heavy-tailed* distribution. When networks become large, the dissemination latency $\mathcal{T}_d$ is (linearly) scalable with respect to the dissemination distance. Our results demonstrate that when secondary users can move in a large region, a *Gamma* distributed (light-tailed) latency in finite networks, or a scalable latency in large networks, is achievable, which encourages the deployment of CRNs for real-time and large applications.

References

[1] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt Generation/Dynamic Spectrum Access/Cognitive Radio Wireless Networks: A Survey," *Computer Networks*, May 2006.

[2] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Distributed Coordination in Dynamic Spectrum Allocation," in *IEEE Journal on Selected Areas in Communication*, Jan 2006.

[3] W. Ren, Q. Zhao, and A. Swami, "On the Connectivity and Multihop Delay of Ad Hoc Cognitive Radio Networks," in *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, April 2011.

[4] L. Sun and W. Wang, "On Study of Achievable Capacity with Hybrid Relay in Cognitive Radio Networks," in *Proceedings of IEEE GLOBECOM*, November 2009.

[5] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad hoc wireless networks," in *IEEE Transactions on Networking*, vol. 10, no. 2, August 2002, pp. 477–486.

[6] M. J. Neely and E. Modiano, "Capacity and Delay Tradeoffs for Ad-Hoc Mobile Networks," in *IEEE Transactions on Information Theory*, vol. 51, no. 6, June 2005.

[7] A. E. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-Delay Trade-off in Wireless Networks," in *Proceedings of IEEE INFOCOM*, Apr 2004.

[8] S. Toumpis and A. Goldsmith, "Large Wireless Networks under Fading, Mobility, and Delay Constraints," in *Proceedings of IEEE INFOCOM*, Apr 2004.

[9] Y. Xu and W. Wang, "The Speed of Information Propagation in Large Wireless Networks," in *Proceedings of IEEE INFOCOM*, April 2008.

[10] Z. Kong and E. M. Yeh, "On the Latency for Information Dissemination in Mobile Wireless Networks," in *Proceedings of ACM MOBIHOC*, May 2008.

[11] S. Toumpis and A. Goldsmith, "Delay and Capacity Trade-offs in Mobile Ad Hoc Networks: A Global Perspective," in *Proceedings of IEEE INFOCOM*, Apr 2006.

[12] A. A. Hanbali, A.A.Kherani, R. Groenevelt, P. Nain, and E. Altman, "Impact of Mobility on the Performance of Relaying in Ad Hoc Networks," in *Proceedings of IEEE INFOCOM*, Apr 2006.

[13] M. Garetto, P. Giaccone, and E. Leonardi, "Capacity Scaling in Delay Tolerant Networks with Heterogeneous Mobile Nodes," in *Proceedings of ACM MOBIHOC*, May 2007.

[14] P. Gupta and P. R. Kumar, "The Capacity of Wireless Networks," in *IEEE Transactions on Information Theory*, vol. 46, no. 2, March 2000, pp. 388–404.

[15] T. Lindavll, *Lectures on the Coupling Method*. John Wiley & Sons, Inc.

[16] H. Cai and D. Y. Eun, "Crossing Over the Bounded Domain: From Exponential to Power-law Inter-meeting Time in MANET," in *Proceedings of ACM MOBICOM*, August 2007.

[17] T. Liggett, "An Improved Subadditive Ergodic Theorem," in *The Annals of Probability*, vol. 13, no. 4, 1985, pp. 1279–1285.