

Smooth Trade-Offs between Throughput and Delay in Mobile Ad Hoc Networks

Pan Li, *Member, IEEE*, Yuguang Fang, *Fellow, IEEE*,
Jie Li, *Senior Member, IEEE*, and Xiaoxia Huang, *Member, IEEE*

Abstract—Throughput capacity in mobile ad hoc networks has been studied extensively under many different mobility models. However, most previous research assumes global mobility, and the results show that a constant per-node throughput can be achieved at the cost of very high delay. Thus, we are having a very big gap here, i.e., either low throughput and low delay in static networks or high throughput and high delay in mobile networks. In this paper, employing a practical restricted random mobility model, we try to fill this gap. Specifically, we assume that a network of unit area with n nodes is evenly divided into cells with an area of $n^{-2\alpha}$, each of which is further evenly divided into squares with an area of $n^{-2\beta}$ ($0 \leq \alpha \leq \beta \leq \frac{1}{2}$). All nodes can only move inside the cell which they are initially distributed in, and at the beginning of each time slot, every node moves from its current square to a uniformly chosen point in a uniformly chosen adjacent square. By proposing a new multihop relay scheme, we present smooth trade-offs between throughput and delay by controlling nodes' mobility. We also consider a network of area n^γ ($0 \leq \gamma \leq 1$) and find that network size does not affect the results obtained before.

Index Terms—Mobile ad hoc networks, throughput, delay, restricted mobility.

1 INTRODUCTION

RECENTLY, there has been intensive interest in wireless ad hoc networks. A wireless ad hoc network is an autonomous system in which users can communicate with each other freely. It can be deployed quickly at a very low cost, and can be easily managed and extended. In the future, there is no doubt that more and more objects in our daily lives will have communication capability. In other words, we will have large-scale ad hoc networks in the near future, in which the capacity of the networks is a critical issue.

In the literature, there is a large body of work on the capacity of wireless networks. Gupta and Kumar [12] have shown that the per-node throughput capacity in random ad hoc networks scales as $\Theta(1/\sqrt{n \log n})$, and the per-node transport capacity in arbitrary ad hoc networks scales as $\Theta(1/\sqrt{n})$, respectively, where n is the number of nodes in the network. The results suggest that the capacity of each node diminishes as n goes large. Many other works such as [3], [5], [6], [7] all lead to the same pessimistic result: per-node throughput goes to 0 as the number of nodes in the network goes to infinity.

Realizing that the mobility is an essence of ad hoc networks, Grossglauser and Tse [11] investigate the capacity of mobile ad hoc networks for the first time and show that a per-node throughput of $\Theta(1)$ can be achieved with a two-hop relaying scheme. However, they have not addressed the delay related issues. After that, a big chunk of work explores the capacity and delay in ad hoc networks considering many different mobility models, such as Brownian mobility model [23], random walk model [8], [9], [10], and i.i.d. model [29]. They all show that a constant per-node throughput can be achieved with end-to-end delay bounded by $\Theta(n)$ with possible logarithmic factors. Therefore, we can only either have low throughput and low delay in static networks, or have high throughput and high delay in mobile networks. There exists a very big gap between the two extremes to be explored.

Notice that all the aforementioned research on the capacity of mobile ad hoc networks assumes global mobility, i.e., nodes can move around the whole network. However, this might not always be the case because in many situations nodes only move within a limited region. For example, in a wireless network covering a big city, each network user usually only moves around a small area close to his or her home, including the work place, groceries, restaurants, and so on. As another example, consider a wireless network in a battlefield. Soldiers are not allowed to move around the whole battlefield. Instead, they can only move in their own post areas in the battlefield. In the literature, Mammen and Shah study in [27] the throughput capacity under a restricted mobility model, where each node is allowed to move along a randomly chosen great circle on the unit sphere with a uniform stationary distribution along the great circle. They show that a constant per-node throughput can be achieved with a delay of $\Theta(n \log n)$. In this paper, we investigate the throughput

- P. Li is with the Department of Electrical and Computer Engineering, Mississippi State University, 229 Simrall Engineering Building, Mississippi State, MS 39762-9571. E-mail: li@ece.msstate.edu.
- Y. Fang is with the Department of Electrical and Computer Engineering, University of Florida, NEB 435, Gainesville, FL 32611. E-mail: fang@ece.ufl.edu.
- J. Li is with the Department of Computer Science, University of Tsukuba, Tsukuba Science City, Ibaraki 305-8573, Japan. E-mail: lijie@cs.tsukuba.ac.jp.
- X. Huang is with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Xueyuan Avenue 1068, Shenzhen University Town, Shenzhen, China. Email: xx.huang@sub.siat.ac.cn.

Manuscript received 7 Apr. 2010; revised 26 Dec. 2010; accepted 10 Jan. 2011; published online 17 Mar. 2011.

For information on obtaining reprints of this article, please send e-mail to: tmc@computer.org, and reference IEEECS Log Number TMC-2010-04-0160. Digital Object Identifier no. 10.1109/TMC.2011.54.

capacity under a more practical restricted random mobility model, and attempt to provide a smooth trade-off between throughput and delay to fill the big gap existing in the literature.

Specifically, we assume that the network of unit area with n nodes is evenly divided into $n^{2\alpha}$ cells with an area of $n^{-2\alpha}$, where $0 \leq \alpha \leq \frac{1}{2}$, each of which is further evenly divided into squares with an area of $n^{-2\beta}$, where $0 \leq \alpha \leq \beta \leq \frac{1}{2}$. A node can only move inside the cell which it is initially distributed in, and at the beginning of each time slot, every node moves from its current square to a uniformly chosen point in a uniformly chosen adjacent square. We first show that the per-node throughput capacity under the proposed random mobility model can be upper bounded by $O(n^{\beta-\alpha-\frac{1}{2}})$. Then, by employing a new multihop relay scheme, we show that the per-node throughput capacity can be lower bounded by $\Omega(n^{2\beta-\alpha-1})$. Besides, we also find that the expected average end-to-end packet delay can be bounded by $O(n^{\beta-\alpha+\frac{1}{2}} \log^2 n)$ and $\Omega(n^{2\beta-\alpha} \log n)$ when $0 \leq \alpha < \beta \leq \frac{1}{2}$, and by $O(n^{\beta-\alpha+\frac{1}{2}})$ and $\Omega(n^{2\beta-\alpha})$ when $0 \leq \alpha = \beta \leq \frac{1}{2}$, respectively. Notice that the result in [9], [23], and [27], i.e., achieving $\Theta(1)$ per-node throughput and $\Theta(n \log n)$ average delay, is the same as our result when $\alpha = 0$ and $\beta = \frac{1}{2}$.

Based on the obtained results, we observe that there are trade-offs among capacity, delay, and mobility. If we want to achieve per-node throughput of $\lambda(n)$ bits per second, where $\lambda \in [\frac{1}{n}, 1]^1$ by using a certain mobility pattern (α, β) , the expected average packet delay will be bounded by $\Omega(n\lambda(n) \log n)$ and $O(n\sqrt{\lambda(n)} \log^2 n)$ when $0 \leq \alpha < \beta \leq \frac{1}{2}$, and by $\Omega(n\lambda(n))$ and $O(n\sqrt{\lambda(n)})$ when $0 \leq \alpha = \beta \leq \frac{1}{2}$, respectively. On the other hand, if we want to upper bound the delay by $E[D]$, where $E[D] \in [\sqrt{n}, n]$ when using a certain mobility pattern (α, β) , then the per-node throughput $\lambda(n)$ will be bounded by $\Omega(\frac{E^2[D]}{n^2 \log^4 n})$ and $O(\frac{E[D]}{n \log^2 n})$ when $0 \leq \alpha < \beta \leq \frac{1}{2}$, and by $\Omega(\frac{E^2[D]}{n^2})$ and $O(\frac{E[D]}{n})$ when $0 \leq \alpha = \beta \leq \frac{1}{2}$, respectively.

Moreover, we notice that in the literature the capacity of wireless networks is usually studied either in a dense network or in an extended network. In this study, we also explore the impact of network size on the capacity and delay in the network. In particular, we consider a network of area n^γ , where $0 \leq \gamma \leq 1$, and show that network size does not change the capacity and delay bounds in the network.

The rest of this paper is organized as follows. In Section 2, we give an overview of the related work. In Section 3, we introduce some definitions and models that we will use throughout this paper. Section 4 presents our proposed multihop relay scheme. In Section 5, we derive an upper bound and a lower bound on per-node throughput capacity and delay, respectively, under the proposed mobility model. We also show the trade-offs among capacity, delay, and mobility in this section. Section 6 discusses the impact of network size on the capacity and delay in the network. We finally conclude this paper in Section 7.

1. We only focus on the exponential order while ignoring logarithmic factors here.

2 RELATED WORK

In this section, we introduce the previous work on the capacity of wireless networks.

In their seminal work [12], Gupta and Kumar have shown that in *dense networks*, where the network area is fixed and the density of nodes increases as fast as the number of nodes n , the per-node throughput capacity in random ad hoc networks scales as $\Theta(1/\sqrt{n \log n})$, and the per-node transport capacity in arbitrary ad hoc networks scales as $\Theta(1/\sqrt{n})$, respectively, i.e., wireless ad hoc networks cannot scale. Later, many research works are also conducted on *extended networks*, where the density of nodes is fixed and the network area increases linearly with the number of nodes. Franceschetti et al. [7] prove by percolation theory that the same $1/\sqrt{n}$ per-node throughput can also be achieved in random ad hoc networks. Buragohain et al. [3] study the throughput capacity in grid networks and find that the $\Omega(1/d)$ per-node throughput can be achieved, where $1 \leq d \leq \sqrt{n}$ is the average source-destination distance. By allowing an arbitrary small fraction of the nodes to be disconnected, Dousse et al. [5] show that the throughput cannot be improved much. Different from the above studies, Duarte-Melo et al. [6] look into the case of *semiextended* networks, where both node density and network area increase as the number of nodes n increases. Specifically, they assume that the network area is a disk of radius n^γ , $0 < \gamma < \frac{1}{2}$. With a $1/(1+d)^\alpha$ propagation model, they show that the per-node throughput capacity scales as $\Omega(1/n^{1-\gamma})$. We can easily observe that the throughput capacity of wireless ad hoc networks cannot scale linearly with the number of nodes. In addition to the capacity of 2D ad hoc networks studied above, the capacity of 3D ad hoc networks are explored in [13] and [21] and cannot scale either.

Several techniques have been proposed in the literature to improve the capacity of wireless networks. Aeron and Saligrama [1] and Ozgur et al. [30] show that by carrying out cooperative distributed Multi-input-Multi-output transmissions (MIMO), the per-node throughput of ad hoc networks can be increased significantly. There is another body of work, such as [2], [16], [17], [18], [20], [22], [25], [26], [32], [35], [36], places powerful nodes like base stations into ad hoc networks, which can greatly enhance network capacity, too. However, the first approach leads to increased system complexity due to the intelligent collaboration among nodes, while the second one incurs significant deployment cost.

Node mobility has been found as another effective way to improve network capacity. Grossglauser and Tse [11] are the first to show that a per-node throughput of $\Theta(1)$ can be achieved in mobile networks with a two-hop relaying scheme. However, they have not addressed the delay related issues. Following [11], there have been extensive works dealing with the trade-off between the capacity and delay under different mobility models. Lin et al. [23] study the throughput capacity under the Brownian mobility model, and show that the two-hop relaying scheme in [11] can achieve a per-node throughput of $\Theta(1)$ with an expected packet delay of $\Omega(\log n/\sigma^2)$, where σ^2 is the variance parameter of the Brownian mobility model. They

also demonstrate that in order to achieve a delay of $\Theta(n^\alpha/\sigma^2)$ for any $\alpha < 0$, the per-node throughput must drop to $O(1/\sqrt{n})$. Gamal et al. [9], [10] look into the throughput capacity under the random walk model. They propose a similar two-hop relaying scheme to that in [11] and show that it can achieve a per-node throughput of $\Theta(1)$ with a delay of $\Theta(n \log n)$. Besides, [24], [29], [33] study the throughput capacity under i.i.d. mobility model, and find that a constant per-node throughput can be achieved with a very large delay. Sharma et al. [31] examine the problem under hybrid mobility models and show that there are trade-offs between capacity and delay. All these works study the unicast scenario in which each source node only transmits to one destination node. The multicast case is discussed in [14]. Ying et al. [34] also propose joint coding-scheduling algorithms to achieve the optimal throughput and delay trade-off. In this study, we focus on the capacity of mobile ad hoc networks in the unicast case without any coding schemes.

As mentioned before, most previous work on the capacity of mobile ad hoc networks assumes global mobility, which results in a big gap between the performance in static networks and that in mobile networks. In this paper, by considering a more practical restricted random mobility model, we attempt to provide a smooth trade-off between throughput and delay to fill the big gap.

3 NOTATIONS, DEFINITIONS, AND ASSUMPTIONS

In this section, we introduce some notations, definitions, and models that we will use in this paper.

3.1 Notations

We use the following notations [15]:

- $f(n) = O(g(n))$ means $f(n)$ is asymptotically upper bounded by $g(n)$, i.e., $\limsup_{n \rightarrow \infty} |f(n)/g(n)| < \infty$.
- $f(n) = \Omega(g(n))$ means $f(n)$ is asymptotically lower bounded by $g(n)$, i.e., $\liminf_{n \rightarrow \infty} |f(n)/g(n)| > 0$.
- $f(n) = \Theta(g(n))$ means $f(n)$ is asymptotically tight bounded by $g(n)$, i.e., $0 < \liminf_{n \rightarrow \infty} |f(n)/g(n)| \leq \limsup_{n \rightarrow \infty} |f(n)/g(n)| < \infty$.

3.2 Definitions

3.2.1 Throughput

As defined in the usual way, the time average of the number of bits that can be transmitted by each node to its destination is called *the per-node throughput*. The sum of per-node throughput over all the nodes in a network is called *the throughput of the network*.

3.2.2 Feasible Throughput

We say that a per-node throughput, denoted by $\lambda(n)$, is *feasible* if there exists a spatial and temporal scheduling scheme that yields a per-node throughput of $\lambda(n)$ bits per second.

3.2.3 Per-Node Throughput Capacity

We say that the per-node throughput capacity in the network [16] is of order $O(f(n))$ bits per second if there is a deterministic constant $0 < c_1 < +\infty$ such that

$$\liminf_{n \rightarrow +\infty} \text{Prob}(\lambda(n) = c_1 f(n) \text{ is feasible}) < 1,$$

and is of order $\Theta(f(n))$ bits per second if there are deterministic constants $0 < c_2 < c_3 < +\infty$ such that

$$\liminf_{n \rightarrow +\infty} \text{Prob}(\lambda(n) = c_2 f(n) \text{ is feasible}) = 1,$$

$$\liminf_{n \rightarrow +\infty} \text{Prob}(\lambda(n) = c_3 f(n) \text{ is feasible}) < 1.$$

3.2.4 Average Packet Delay

The *delay* of a packet in a network is the time it takes the packet to reach the destination after it leaves the source. As in [23] and [28], we do not consider the queuing delay at the source node since we are more interested in the network delay caused by nodes' mobility. The *average packet delay* of a network is obtained by averaging over all transmitted packets in the network. Besides, we also assume the packet size scales as the per-node throughput.²

3.3 Network Model

We assume that n mobile nodes are initially independently and uniformly distributed on a 2D torus of unit area. We follow the process in [11] to choose random sender-receiver pairs so that each node is a source node for one flow and a destination node for at most $O(1)$ flows. Consider that each source node generates traffic at the same rate, denoted by $\lambda(n)$ bits per second. We also assume that the packet arrival process at each node is independent of the node mobility process.

Besides, a source node sends a packet directly to a destination node if it is within the transmission range of the source node. Otherwise, a source node forwards a packet to one or more relay nodes, which can help forward the packet to other relay nodes, or directly to the destination node if it is within the transmission range.

3.4 Interference Model

We employ the Protocol Model in [12] as the interference model. Suppose a node T_i transmits to a node R_i . Let T_i^t and R_i^t denote the positions of these two nodes, respectively, at time t ($t > 0$). Then, in order for this transmission to be successful, the positions of every other transmitter T_j simultaneously transmitting at time t , denoted by T_j^t , should satisfy

$$|T_j^t - R_i^t| \geq (1 + \Delta)|T_i^t - R_i^t|.$$

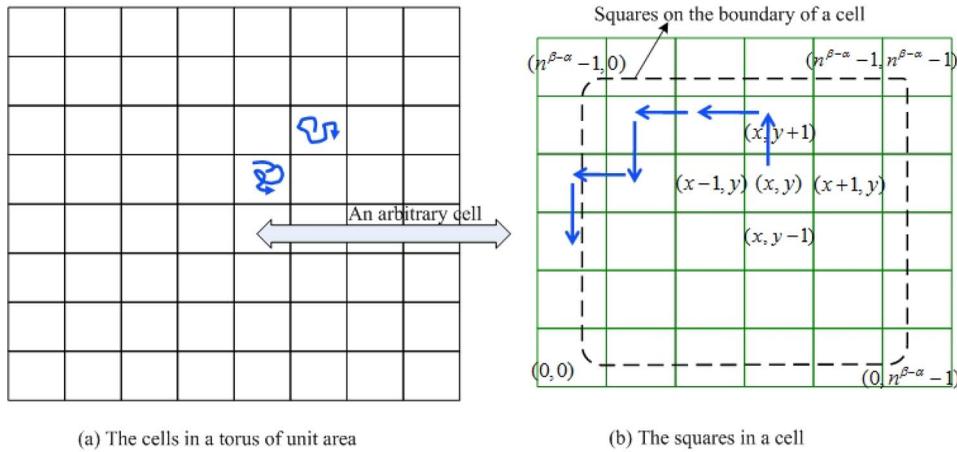
The quantity $\Delta > 0$ models situations, where a guard zone is specified by the protocol to prevent a neighboring node from transmitting at the same time. It also allows for imprecision in the achieved range of transmissions.

Assume that the channel capacity is W bits per second. So, the transmission rate is W bits per second when the protocol model is satisfied, and 0 otherwise.

3.5 Restricted Random Mobility Model

As shown in Fig. 1a, the unit torus with n nodes is evenly divided into $n^{2\alpha}$ squares each of which has an area of $n^{-2\alpha}$,

2. As pointed out in [9] and [24], under this assumption, queuing delay at source node can actually be a constant under certain circumstances, which gives us another reason to focus on network delay.



(a) The cells in a torus of unit area

(b) The squares in a cell

Fig. 1. (a) We first divide the unit torus into $n^{2\alpha}$ cells of equal area. (b) We also divide each cell into $n^{2\beta-2\alpha}$ squares of equal area.

where $0 \leq \alpha \leq \frac{1}{2}$. We call these squares “cells” with length of $n^{-\alpha}$. The nodes located inside a cell can only move around within this cell. Specifically, each cell is further evenly divided into squares with an area of $n^{-2\beta}$ for each one, where $0 \leq \alpha \leq \beta \leq \frac{1}{2}$, as shown in Fig. 1b. We call these smaller squares “squares” to be differentiated from the cells. Time is divided into slots of equal duration. Initially, each of the n nodes is assumed to be randomly and independently distributed in each of the $n^{2\beta}$ squares in the network with the same probability. At the beginning of each time slot, a node moves from its current square to a uniformly chosen point in an adjacent square, which is chosen uniformly from its four adjacent squares. The adjacent squares are defined as follows: assume the squares in each cell are numbered by a 2D array, as shown in Fig. 1b. For a square denoted by (x, y) , where $x, y \in (0, n^{\beta-\alpha} - 1)$, the four adjacent squares are $(x - 1, y)$, $(x + 1, y)$, $(x, y - 1)$, $(x, y + 1)$, respectively. Those squares on the boundary of a cell only have three or two adjacent squares in the same cell. Note that we consider the velocity of the nodes in the network to be $\Theta(n^{-\beta})$. The reason is that the average neighborhood range scales as $\Theta(n^{-\beta})$ which is defined by the size of squares. In correspondence to the diminishing neighborhood, we make the velocity of the nodes scale as $\Theta(n^{-\beta})$, too.

Notice that there are four extreme cases of special interest.

- *Case (i):* $\alpha = \beta = 0$. In this case, each node independently moves to a uniformly chosen point in the network in each time slot, which is similar to the i.i.d. mobility model studied in [24], [29], and [33].
- *Case (ii):* $\alpha = \beta \in (0, \frac{1}{2})$. In this case, the network is evenly divided into $n^{2\alpha}$ squares, and each node in each square independently moves to a uniformly chosen point in the same square in each time slot. It is a restricted i.i.d. mobility model.
- *Case (iii):* $\alpha = \beta = \frac{1}{2}$. In this case, the network is evenly divided into n squares each of which has an area of $\frac{1}{n}$. Each node can only move around within these very small squares, which is just like a network with static nodes as studied in [7] and [12].

- *Case (iv):* $\alpha = 0, \beta = \frac{1}{2}$. In this case, nodes can move around the whole network with each square of area $\frac{1}{n}$. In each time slot, a node moves from the current square to a uniformly chosen point in an adjacent square, which is the random walk model in [9] and [10]. Note that the random walk model is the discrete time version of the Brownian motion model considered in [23].

Moreover, we also have the following assumptions:

- A1. Transmissions between two nodes can be carried out only when they are located in the same square or adjacent squares.
- A2. Only the source node of a packet or the relay nodes in squares on the boundary of a cell are allowed to replicate the packet.

Notice that many proposed schemes [9], [10], [31] in the literature satisfy the assumption (A1). Assumption (A2) is made because we limit the mobility of nodes inside the cells they are initially located in.

4 A MULTIHOP RELAY SCHEME

Before we move on with our main results, we first present a new multihop relay scheme for the transmissions between source-destination pairs.

S-I. Each square in the network becomes active in every c ($c \geq 1$) time slots. According to the mobility model, in each time slot every node moves a distance of $\Theta(n^{-\beta})$ at a speed of $\Theta(n^{-\beta})$. So the length of a time slot is set to be $\Theta(1)$. The value of c will be determined later.

S-II. A node in an active square can transmit its packet to another node only if they are in the same square or in two adjacent squares. Specifically, each time slot is further divided into three subslots A, B, and C, of the same length. For a square with N nodes, each subslot is further divided into N minislots of the same length.

- In subslot A, if $N \geq 2$, every node acting as source transmits one-by-one in a minislot its packet to another randomly chosen node in the same square, which acts as a relay. This relay node can also be a destination node.

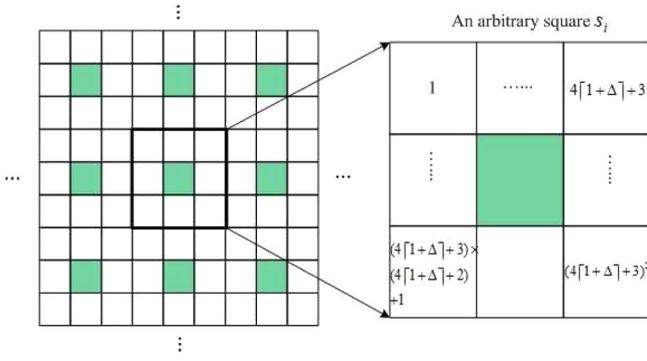


Fig. 2. An example for dividing the network into groups of $(4\lceil 1 + \Delta \rceil + 3)^2$ squares.

- In subslot B, if this active square is an interior square, nothing happens. Otherwise, if this active square is on the boundary of a cell, every node in this square acting as a relay transmits one-by-one in a minislot its packet to another randomly chosen node in the adjacent square in the adjacent cell, which also acts as a relay. This relay node can also be a destination node.
- In subslot C, if $N \geq 2$, every node acting as a relay transmits one-by-one in a minislot its packet to the destination node if it is in the same square. Otherwise, nothing happens.

In other words, after receiving a packet from a source node in the same cell, a relay node carries this packet and moves around until it goes to the boundary of the cell, where this relay node forwards the packet to another node in an adjacent cell. Thus, a packet is transmitted at least once and at most twice in each of the cells it goes through.

Since we allow a transmission between two nodes only if they are located in the same square or in two adjacent squares, we choose the transmission range to be $\sqrt{5}$ times the side length of the square, i.e., $\sqrt{5}n^{-\beta}$. Thus, we have the following result.

Lemma 1. *If there is always at least one node within the transmission range of a transmitter, a lower bound on the transmission rate of each square is $W/(4\lceil 1 + \Delta \rceil + 3)^2$. Note that $\lceil x \rceil$ is the ceiling function of x , which returns the smallest integer not smaller than x , or formally,*

$$\lceil x \rceil = \min\{n \in \mathbb{Z} \mid n \geq x\},$$

where \mathbb{Z} is the set of real numbers.

Proof. We show a lower bound on the transmission rate of each square by proposing a scheduling scheme as follows: we group all the squares in the network into sets, each of which contains $(4\lceil 1 + \Delta \rceil + 3)^2$ squares. As shown in Fig. 2, the $(4\lceil 1 + \Delta \rceil + 3)^2$ squares in each set are numbered from 1 to $(4\lceil 1 + \Delta \rceil + 3)^2$ in the same way. We further divide time into sequences of successive slots, denoted by t ($t = 0, 1, 2, 3, \dots$). During a slot t , all squares that are numbered $\lceil t \bmod (4\lceil 1 + \Delta \rceil + 3)^2 \rceil + 1$ are allowed to transmit packets.

Consider a time slot when a square s_i is active and a node m in s_i is chosen to transmit its packet to another

node m' during the whole time slot. Recall that the side length of a square is $n^{-\beta}$ as we assume in Section 3.5. Since we only allow transmissions within the same square or between two adjacent squares, the distance between the transmitter m and its receiver, denoted by $d_{(m,m')}$, satisfies

$$d_{(m,m')} \leq \sqrt{5}n^{-\beta}.$$

Notice that those squares where the transmitters may interfere with the transmission from node m to node m' are located along the perimeters of concentric squares centered at s_i . Denote by q an interfering transmitter to the transmission of node m , and q' its intended receiver. Then, the minimal distance between node q and node m' , denoted by $d_{(q,m')}$, satisfies

$$\begin{aligned} d_{(q,m')} &\geq \left\{ 2 \cdot \frac{[4\lceil 1 + \Delta \rceil + 3] - 1}{2} - 1 \right\} \cdot n^{-\beta} \\ &= (4\lceil 1 + \Delta \rceil + 1)n^{-\beta}. \end{aligned}$$

Noting that $4 > \sqrt{5}$ and $\lceil 1 + \Delta \rceil \geq 1 + \Delta$, we can obtain that

$$\begin{aligned} d_{(q,m')} &\geq \sqrt{5}(1 + \Delta)n^{-\beta} \\ &\geq (1 + \Delta)d_{(m,m')}, \end{aligned}$$

i.e., the protocol model can be satisfied for the transmission between node m and node m' . Thus, according to the interference model, we can obtain that in $(4\lceil 1 + \Delta \rceil + 3)^2$ time slots, each square in the network has a chance to transmit at a constant transmission rate W , which results in the average transmission rate of each square being $W/(4\lceil 1 + \Delta \rceil + 3)^2$. \square

Thus, from Lemma 1, we can have that $c = (4\lceil 1 + \Delta \rceil + 3)^2$ in S-I. Note that the transmission rate of each square is obtained for a square assuming one node is transmitting in the whole time slot. Since in our relay scheme, a time slot is further divided into minislots, one for each node in the square, all the nodes within the transmission range share this transmission rate. We will give some more explanations when we derive a lower bound on the per-node throughput later.

We also note that every time a square is active, in subslot A a transmission initiated by a source can take place only when the square contains at least two nodes. We have the following result.

Lemma 2. *Any square has at least two nodes with positive constant probability, denoted by p_c , for any $0 \leq \beta \leq \frac{1}{2}$.*

Proof. For an arbitrary square, the probability that there are at least two nodes in it, denoted by p_c , is

$$\begin{aligned} p_c &= 1 - (1 - n^{-2\beta})^n - n \cdot n^{-2\beta}(1 - n^{-2\beta})^{n-1} \\ &= 1 - e^{n \log(1 - n^{-2\beta})} - n^{1-2\beta} e^{(n-1) \log(1 - n^{-2\beta})} \\ &\geq 1 - \frac{1}{e^{n^{1-2\beta}}} - n^{1-2\beta} \frac{1}{e^{(n-1)n^{-2\beta}}}. \end{aligned}$$

So, as $n \rightarrow \infty$, $p_c \rightarrow 1$ when $0 \leq \beta < \frac{1}{2}$, and $p_c \geq 1 - \frac{2}{e} > 0$ when $\beta = \frac{1}{2}$, which is also a constant. \square

Similarly, we note that in subslot B a transmission initiated by a relay node in a square on the boundary of a cell can be carried out only when the adjacent square in the adjacent cell has at least one node. Thus, we have the following result.

Lemma 3. Any square has at least one node with positive constant probability, denoted by p'_c , for any $0 \leq \beta \leq \frac{1}{2}$, and $p_c \leq p'_c \leq 1$.

5 CAPACITY, DELAY, AND MOBILITY TRADE-OFFS

Recall that in our restricted random mobility model, the nodes' mobility pattern are controlled by two parameters: α and β . Since obviously capacity and delay in the network greatly depend on the mobility of the nodes, there is a trade-off among capacity, delay, and mobility when we tune these two parameters. In this section, we first investigate per-node throughput capacity and average packet delay with respect to α and β in Sections 5.1 and 5.2, respectively, when $0 < \alpha < \beta < \frac{1}{2}$. We then study the capacity and the delay in four extreme cases in Section 5.3. Finally, we discuss the trade-offs among capacity, delay, and mobility in Section 5.4.

5.1 Throughput Capacity and Mobility

We first derive an upper bound on the throughput capacity for all the schemes satisfying assumptions (A1) and (A2). Recall that a transmission is considered to be successful if the Protocol Model is satisfied. Suppose that T_i^t and T_j^t transmit to R_i^t and R_j^t , respectively, at the same time t . Then, we have

$$\begin{aligned} |R_i^t - R_j^t| &\geq |T_i^t - R_j^t| - |T_i^t - R_i^t| \\ &\geq (1 + \Delta)|T_j^t - R_j^t| - |T_i^t - R_i^t|. \end{aligned}$$

Similarly,

$$\begin{aligned} |R_j^t - R_i^t| &\geq |T_j^t - R_i^t| - |T_j^t - R_j^t| \\ &\geq (1 + \Delta)|T_i^t - R_i^t| - |T_j^t - R_j^t|. \end{aligned}$$

Adding these two inequalities together, we can obtain

$$|R_i^t - R_j^t| \geq \frac{\Delta}{2} (|T_i^t - R_i^t| + |T_j^t - R_j^t|).$$

This means that the disks centered at the receivers, with radius $\frac{\Delta}{2}$ times the lengths of hops, are essentially disjoint, which we define as "Disjoint Disks."

Since the whole area of the network is 1, then we have

$$\sum_{(T_i^t, R_i^t) \in P(t)} S_E \leq 1, \quad (1)$$

where $P(t)$ is the set of all transmitter-receiver pairs carrying out transmissions at time t , and S_E is the area occupied by a single hop's transmission, which we call "Exclusion Area." Obviously, an exclusion area is equal to the area of disjoint disks, i.e.,

$$S_E = \frac{\pi \Delta^2}{4} d_{(T_i^t, R_i^t)}^2,$$

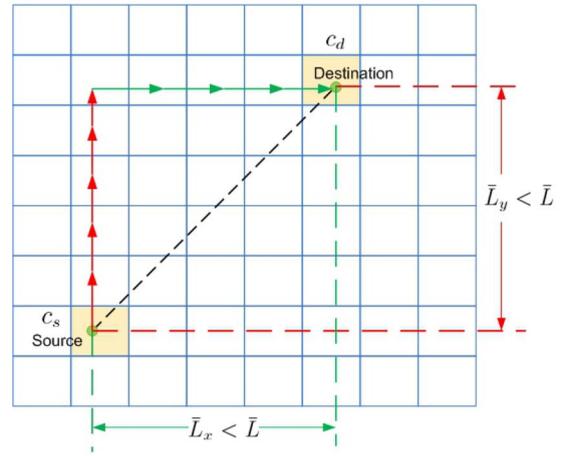


Fig. 3. The cells a packet may go through.

where $d_{(T_i^t, R_i^t)}$ is the distance between a transmitter-receiver pair (T_i^t, R_i^t) .

Thus, combining with (1), we obtain

$$\sum_{(T_i^t, R_i^t) \in P(t)} d_{(T_i^t, R_i^t)}^2 \leq \frac{4}{\pi \Delta^2}.$$

Besides, since at any time there are at most $\frac{n}{2}$ concurrent transmitter-receiver pairs, according to Cauchy-Schwartz inequality, we have

$$\begin{aligned} \sum_{(T_i^t, R_i^t) \in P(t)} d_{(T_i^t, R_i^t)} &\leq \sqrt{\sum_{(T_i^t, R_i^t) \in P(t)} d_{(T_i^t, R_i^t)}^2 \cdot \sum_{(T_i^t, R_i^t) \in P(t)} 1^2} \\ &\leq \sqrt{\frac{2n}{\pi \Delta^2}}. \end{aligned}$$

Assume that the average distance between the source and destination of a bit is \bar{L} . Recall that we assume the packet arrival process at each node is independent of the node mobility process. Thus, the source node and the destination node are located in each cell with the same probability when a packet arrives. So, we have $\bar{L} = \Theta(1)$.³ Since packets are relayed at least once in those cells they go through, and the average transmission distance between a source node and a relay node is at least $\bar{L}n^{-\beta}$ due to assumption (A1), the average cumulative transmitted distance of a bit, denoted by \bar{L}_{trans} , can be calculated as⁴

$$\bar{L}_{trans} \geq \frac{\bar{L}}{\sqrt{2}n^{-\alpha}} \cdot \bar{L}n^{-\beta} = \frac{\bar{L}^2}{\sqrt{2}} n^{\alpha-\beta}. \quad (2)$$

Denote the duration of a time slot by T . Thus, we can obtain an upper bound on $\lambda(n)$ by having the following inequality:

$$\begin{aligned} \lambda(n)T \cdot n \cdot \bar{L}_{trans} &\leq WT \cdot \sum_{(T_i^t, R_i^t) \in P(t)} d_{(T_i^t, R_i^t)} \\ &\leq WT \sqrt{\frac{2n}{\pi \Delta^2}}, \end{aligned} \quad (3)$$

3. It has been proved in Mathematics that the mean distance between two random points in a unit square is around 0.52. So, we use \bar{L} instead in the following since it is just a constant.

4. Transmission distance is the distance a bit is transmitted in a single-hop, while cumulative transmitted distance is the total distance a bit is transmitted from its source to its destination.

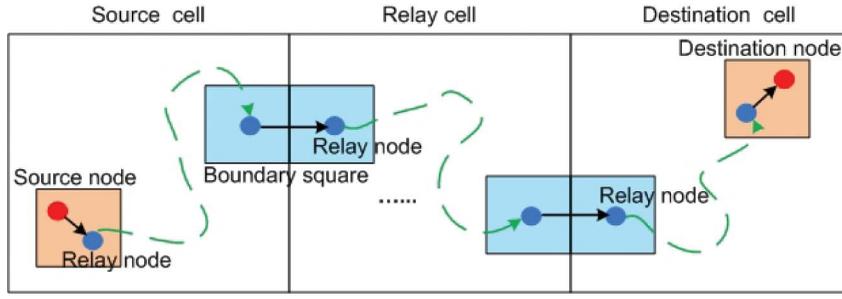


Fig. 4. An example for relaying a packet for a source-destination pair. The dotted lines mean that the relay nodes will carry a packet moving around until it hits a boundary square, or the square where the destination node lies.

from which we can have

$$\lambda(n) \leq \frac{2W}{\sqrt{\pi}\bar{L}^2\Delta} n^{\beta-\alpha-\frac{1}{2}},$$

i.e.,

$$\lambda(n) = O(n^{\beta-\alpha-\frac{1}{2}}). \quad (4)$$

Next, we derive a lower bound on the throughput capacity based on the proposed relay scheme. We choose a routing strategy as shown in Fig. 3. In particular, packets from a source cell c_s are relayed first along the cells with the same x-coordinate as c_s , then along those with the same y-coordinate as the destination cell c_d . Since a packet is transmitted at most twice in each cell, as shown in Fig. 3 a packet is at most transmitted $2 \cdot (2\bar{L}/n^{-\alpha})$, i.e., $4\bar{L}n^\alpha$ times in the network. Thus, with the throughput of each node being $\lambda(n)$, every node has to support a traffic rate of at most $4\bar{L}n^\alpha\lambda(n)$. Besides, since the transmission range of each node is chosen to be $\sqrt{5}n^{-\beta}$, there are $n \cdot \pi(\sqrt{5}n^{-\beta})^2$, i.e., $5\pi n^{1-2\beta}$ nodes which will share the same channel. From Lemmas 2 and 3, on average, an active square can have transmissions for at least $T/3$ during a time slot T with a probability no smaller than p_c . Thus, by Lemma 1, a lower bound on available transmission rate is $p_c W/3c$. Therefore, a per-node throughput $\lambda(n)$ is feasible if

$$4\bar{L}n^\alpha\lambda(n) \leq \frac{p_c W/3c}{5\pi n^{1-2\beta}},$$

holds.⁵ Thus, we can obtain that

$$\lambda(n) = \frac{p_c W}{60\pi\bar{L}c} n^{2\beta-\alpha-1},$$

bits per second is feasible, and hence,

$$\lambda(n) = \Omega(n^{2\beta-\alpha-1}). \quad (5)$$

Notice that in (4) and (5), we have

$$\left(\beta - \alpha - \frac{1}{2}\right) - (2\beta - \alpha - 1) = \frac{1}{2} - \beta > 0,$$

for any $0 < \alpha < \beta < \frac{1}{2}$, which is consistent with the fact that the order of the upper bound is always no lower than that of the lower bound.

5. A similar approach is employed in [12] and [31].

5.2 Average Packet Delay and Mobility

We first introduce a concept called “the First Hitting Time.” Let $X(t)$ be a discrete-time Markov chain on state space S , with initial stationary distribution Π . We have the following definition [31].

Definition 1. The first hitting time for the set of states A , denoted by T^A , is the first time instant at which $X(t)$ enters A , i.e.,

$$T^A = \inf\{t \geq 0 : X(t) \in A\},$$

with $X(0)$ being distributed according to Π .

Besides, we have the following result on the first hitting time for a single state in case of a 2D torus containing n nodes uniformly distributed in $\sqrt{n} \times \sqrt{n}$ squares [31]. In this case, $X(t)$ denotes a node’s position at time t and the state space S is the set of n squares.

Lemma 4. Let T^A denote the first hitting time for a single state A on a 2D torus of size $\sqrt{n} \times \sqrt{n}$. Then, the expectation of T^A , denoted by $E[T^A]$, is

$$E[T^A] = \Theta(n \log n).$$

We note that we limit our discussions to the case when $0 < \alpha < \beta < \frac{1}{2}$ in this part. The extreme cases will be discussed later. Thus, a packet needs to go through many cells in order to arrive at the destination node. Here, we first find a lower bound on the average packet delay by employing the same routing strategy as that used for deriving a lower bound on the throughput capacity. One such an example is shown in Fig. 4.

We ignore the transmission time of the packets, since the transmissions are carried out at a much smaller time scale than the nodes’ mobility. We also ignore the queuing delay as in [31]. In this study, we mainly focus on the delay caused by nodes’ random mobility.

We define three kinds of cells. A cell in which the source node of a packet is located is called a *Source Cell*. A cell which the destination node of a packet lies in is called a *Destination Cell*. The other cells which a packet goes through are called *Relay Cells*. In Lemma 2, we have proved that when $0 < \beta < \frac{1}{2}$, a square has at least two nodes with probability 1. Thus, a source node can always find a relay node.

Denote the time needed for the relay node in source cell to first hit a boundary square by T_S . Note that a cell is a square instead of a torus with size $n^{\beta-\alpha} \times n^{\beta-\alpha}$. Thus, due to

the edge effects, by Lemma 4, the expectation of T_S , denoted by $E[T_S]$, is

$$E[T_S] = \Omega(n^{2\beta-2\alpha} \log n).$$

Besides, the probability that the adjacent square in the adjacent cell is not empty, denoted by p_0 , is

$$p_0 = 1 - (1 - n^{-2\beta})^n = 1 - e^{n \log(1-n^{-2\beta})} = 1 - \frac{O(1)}{e^{n^{1-2\beta}}},$$

which approaches to 1 as $n \rightarrow +\infty$. So, a packet can be relayed between cells successfully with high probability.

Let T_R denote the time spent in each relay cell, and $E[T_R]$ the expectation. Notice that Lemma 4 gives the expected first hitting time when n nodes are uniformly distributed in $\sqrt{n} \times \sqrt{n}$ squares, while in a relay cell, a node needs to hit a boundary from the opposite boundary. Since [23] shows that the travel time increases as the distance, the expected boundary hitting time in a relay cell is lower bounded by the expected first hitting time defined before, i.e.,

$$E[T_R] = \Omega(n^{2\beta-2\alpha} \log n).$$

Finally, let T_D denote the time spent in destination cell, and $E[T_D]$ the expectation. Since we assume that the packet arrival process is independent of the node mobility, the destination node is uniformly distributed in the destination cell. Thus, similar to the above, we have

$$E[T_D] = \Omega(n^{2\beta-2\alpha} \log n).$$

Recall that the average distance between a source and a destination is $\bar{L} = \Theta(1)$. Denote by H the average number of cells a packet has to go through before it reaches the destination. Then, the expectation of H , denoted by $E[H]$, is

$$E[H] = \Theta\left(\frac{\bar{L}_x + \bar{L}_y}{n^{-\alpha}}\right) = \Omega\left(\frac{\bar{L}}{n^{-\alpha}}\right) = \Omega(n^\alpha),$$

where \bar{L}_x and \bar{L}_y are the distances between the source cell and the destination cell along the x -axis and along the y -axis, respectively, as shown in Fig. 3.

Denote by D and $E[D]$ the average packet delay and the expectation of the average packet delay, respectively. We can obtain

$$\begin{aligned} E[D] &= E[T_S] + E[H]E[T_R] + E[T_D] \\ &= \Omega(n^{2\beta-2\alpha} \log n), \end{aligned} \quad (6)$$

which is a lower bound on the expected average packet delay.

Next, we attempt to find an upper bound on $E[D]$ which is irrelevant to routing schemes. Chen and Gunes [4] show that the cover time of a 2D grid with n nodes is $O(n \log^2 n)$. So the first hitting time T_S in a cell satisfies $E[T_S] = O(n^{2\beta-2\alpha} \log^2 n)$. Recall that according to assumption (A2), a packet is relayed from one cell to another when it goes to the boundary squares of the current cell. Thus, we can have

$$E[D] = O(E[H]n^{2\beta-2\alpha} \log^2 n). \quad (7)$$

Similar to (3), we can also have

$$\lambda(n)T \cdot n \cdot E[H] \cdot \bar{L}n^{-\beta} \leq WT \sqrt{\frac{2n}{\pi\Delta^2}},$$

i.e.,

$$\lambda(n)E[H] = O(n^{\beta-\frac{1}{2}}). \quad (8)$$

Thus, combining (7) and (8), we obtain

$$\lambda(n)E[D] = O(n^{3\beta-2\alpha-\frac{1}{2}} \log^2 n).$$

According to (5), we can finally arrive at an upper bound on the expected average packet delay, i.e.,

$$E[D] = O(n^{\beta-\alpha+\frac{1}{2}} \log^2 n). \quad (9)$$

Note that comparing (6) and (9), we find that

$$\left(\beta - \alpha + \frac{1}{2}\right) - (2\beta - \alpha) = \frac{1}{2} - \beta > 0,$$

for any $0 < \alpha < \beta < \frac{1}{2}$, i.e., the order of the upper bound is always higher than that of the lower bound.

5.3 Extreme Cases of Special Interest

Here, we discuss the capacity and delay in extreme cases of special interest.

Case I: $\alpha = \beta = 0$.

In this case, each node independently moves to a uniformly chosen point in the network in each time slot. The average cumulative transmitted distance of a bit, i.e., \bar{L}_{trans} , changes from (2) into $\bar{L}_{trans} = \Omega(1)$. Besides, similar to (3), we have $\lambda(n) \cdot n \cdot \bar{L}_{trans} = O(\sqrt{n})$, and hence

$$\lambda(n) = O(n^{-\frac{1}{2}}).$$

Moreover, according to our relay scheme, a source node can always directly transmit to a destination node since they are in the same square. Thus, we can easily obtain

$$\lambda(n) = \Omega\left(\frac{1}{n}\right), \quad E[D] = \Omega(1).$$

Furthermore, since source nodes and destination nodes are in the same square, (7) changes into $E[D] = \Theta(E[H] \cdot 1)$. Similar to (8), we also have $\lambda(n)E[H] = O(n^{-\frac{1}{2}})$. Thus, we obtain that $\lambda(n)E[D] = O(n^{-\frac{1}{2}})$, which yields

$$E[D] = O(n^{\frac{1}{2}}).$$

Note that Neely and Modiano [29] present a scheme that can achieve $\Theta(1)$ per-node throughput with expected average packet delay of $\Theta(n)$ under i.i.d. mobility model. In their model, they evenly divide the network into n squares and the communication range can be as short as $n^{-\frac{1}{2}}$, which is very different from our case here.

Case II: $\alpha = \beta \in (0, \frac{1}{2})$.

Following the steps in Sections 5.1 and 5.2, we can obtain

$$\lambda(n) = O(n^{-\frac{1}{2}}), \quad \lambda(n) = \Omega(n^{\alpha-1}),$$

and

$$E[D] = O(n^{\frac{1}{2}}), \quad E[D] = \Omega(n^\alpha).$$

Case III: $\alpha = \beta = \frac{1}{2}$.

Along the line in Sections 5.1 and 5.2, we can obtain that

$$\lambda(n) = \Theta(n^{-\frac{1}{2}}), \quad E[D] = \Theta(n^{\frac{1}{2}}).$$

Note that in this case, our mobile network is just like a static network as studied in [7] and [12]. Gupta and Kumar [12] have shown that a static random network can achieve per-node throughput of $\sqrt{n \log n}$, while Franceschetti et al. [7] later show that the per-node throughput of $1/\sqrt{n}$ is also achievable, which is consistent with our result here. We also notice that neither of these two works consider the delay.

Case IV: $\alpha = 0, \beta = \frac{1}{2}$.

In this case, the network turns into a torus of size $\sqrt{n} \times \sqrt{n}$ and we can have

$$\lambda(n) = \Theta(1), \quad E[D] = \Theta(n \log n),$$

which are the same as the results derived in [9].

Combining the above results with those obtained in Sections 5.1 and 5.2, we arrive at the following two theorems.

Theorem 1. *Under the restricted random mobility model, the per-node throughput, denoted by $\lambda(n)$, scales as*

$$\lambda(n) = \begin{cases} \Omega(n^{2\beta-\alpha-1}), \\ O(n^{\beta-\alpha-\frac{1}{2}}). \end{cases}$$

Theorem 2. *Under the restricted random mobility model, the expected average packet delay, denoted by $E[D]$, scales as*

$$E[D] = \begin{cases} \Omega(n^{2\beta-\alpha} \log n), \\ O(n^{\beta-\alpha+\frac{1}{2}} \log^2 n), \end{cases}$$

when $0 \leq \alpha < \beta \leq \frac{1}{2}$, and scales as

$$E[D] = \begin{cases} \Omega(n^{2\beta-\alpha}), \\ O(n^{\beta-\alpha+\frac{1}{2}}), \end{cases}$$

when $0 \leq \alpha = \beta \leq \frac{1}{2}$.

5.4 Trade-Offs among Capacity, Delay, and Mobility

Recall that two parameters α and β together define a mobility pattern, and β also defines a lower bound on the transmission range. In Theorem 1, we observe that in order to obtain higher throughput, we need to increase β while decreasing α , which means that we need to allow nodes to move in a larger area while have a smaller transmission range. However, by doing this, the delay will also increase as the throughput increases, as we can find in Theorem 2. As a result, obviously, there are trade-offs among capacity, delay, and mobility.

Notice that the lower bound on throughput capacity and on average packet delay, shown in (5) and (6), respectively, can be achieved with the same routing strategy. So, from Theorems 1 and 2, we can obtain that when the per-node throughput of $\lambda(n) = n^{2\beta-\alpha-1}$ bits per second is achievable, the expected average packet delay is lower bounded by $n^{2\beta-\alpha} \log n$ and upper bounded by $n^{\beta-\alpha+\frac{1}{2}} \log^2 n$ when $0 \leq \alpha < \beta \leq \frac{1}{2}$, and lower bounded by $n^{2\beta-\alpha}$ and upper bounded by $n^{\beta-\alpha+\frac{1}{2}}$ when $0 \leq \alpha = \beta \leq \frac{1}{2}$.

Let $x = 2\beta - \alpha - 1$, i.e., the exponential order of the lower bound on the per-node throughput. Then, $-1 \leq x \leq 0$. Since $0 \leq \alpha \leq \beta \leq \frac{1}{2}$, we can have

$$\beta \geq \frac{x+1}{2}.$$

Then, the exponential order of the upper bound on the expected average packet delay is

$$\beta - \alpha + \frac{1}{2} = x - \beta + \frac{3}{2} \leq \frac{x}{2} + 1,$$

and the exponential order of the lower bound on the expected average packet delay is

$$2\beta - \alpha = x + 1.$$

Thus, with the achievable per-node throughput being $\lambda(n) = n^x$, where $-1 \leq x \leq 0$, an upper bound and a lower bound on the expected average packet delay are

$$E[D] = O(n\sqrt{\lambda(n)} \log^2 n), \quad E[D] = \Omega(n\lambda(n) \log n),$$

when $0 \leq \alpha < \beta \leq \frac{1}{2}$, and are

$$E[D] = O(n\sqrt{\lambda(n)}), \quad E[D] = \Omega(n\lambda(n)),$$

when $0 \leq \alpha = \beta \leq \frac{1}{2}$.

Corollary 1. *Under the restricted random mobility model, we can achieve per-node throughput of $\lambda(n)$ bits per second, where $\lambda \in [\frac{1}{n}, 1]$ with the expected average packet delay bounded as*

$$E[D] = \begin{cases} \Omega(n\lambda(n) \log n), \\ O(n\sqrt{\lambda(n)} \log^2 n), \end{cases}$$

when $0 \leq \alpha < \beta \leq \frac{1}{2}$, and bounded as

$$E[D] = \begin{cases} \Omega(n\lambda(n)), \\ O(n\sqrt{\lambda(n)}), \end{cases}$$

when $0 \leq \alpha = \beta \leq \frac{1}{2}$.

Moreover, let $y = \beta - \alpha + \frac{1}{2}$, i.e., the exponential order of the upper bound on the expected average packet delay. Then, $\frac{1}{2} \leq y \leq 1$. Since $\alpha \geq 0$, we have

$$\beta \geq y - \frac{1}{2}.$$

Then, the exponential order of the lower bound on the per-node throughput is

$$2\beta - \alpha - 1 = y + \beta - \frac{3}{2} \geq 2y - 2,$$

and the exponential order of the upper bound on the per-node throughput is

$$\beta - \alpha - \frac{1}{2} = y - 1.$$

So, when the expected average packet delay is upper bounded by $E[D] = O(n^{\beta-\alpha+\frac{1}{2}} \log^2 n)$, the per-node throughput capacity can be lower bounded and upper bounded as

$$\lambda(n) = \Omega\left(\frac{E^2[D]}{n^2 \log^4 n}\right), \quad \lambda(n) = O\left(\frac{E[D]}{n \log^2 n}\right),$$

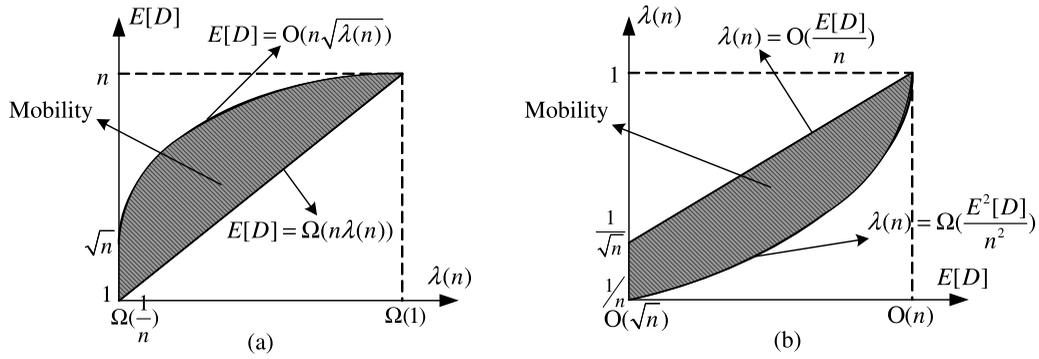


Fig. 5. Trade-offs among capacity, delay, and mobility. (a) An upper bound and a lower bound on the expected average packet delay when an achievable per-node throughput ranges from $\frac{1}{n}$ to 1. (b) An upper bound and a lower bound on the per-node throughput when an upper bound on the expected average packet delay ranges from \sqrt{n} to n . We ignore the logarithmic terms in both cases as in [23].

respectively, when $0 \leq \alpha < \beta \leq \frac{1}{2}$, and be lower bounded and upper bounded as

$$\lambda(n) = \Omega\left(\frac{E^2[D]}{n^2}\right), \quad \lambda(n) = O\left(\frac{E[D]}{n}\right),$$

respectively, when $0 \leq \alpha = \beta \leq \frac{1}{2}$.

Corollary 2. Under the restricted random mobility model, if the expected average packet delay is upper bounded by $E[D]$ where $E[D] \in [\sqrt{n}, n]$, then the per-node throughput $\lambda(n)$ is bounded as

$$\lambda(n) = \begin{cases} \Omega\left(\frac{E^2[D]}{n^2 \log^4 n}\right), \\ O\left(\frac{E[D]}{n \log^2 n}\right), \end{cases}$$

when $0 \leq \alpha < \beta \leq \frac{1}{2}$, and as

$$\lambda(n) = \begin{cases} \Omega\left(\frac{E^2[D]}{n^2}\right), \\ O\left(\frac{E[D]}{n}\right), \end{cases}$$

when $0 \leq \alpha = \beta \leq \frac{1}{2}$.

We further illustrate the trade-offs among capacity, delay, and mobility more explicitly in Fig. 5, ignoring the logarithmic terms as in [23]. Our results show that by controlling nodes' mobility pattern (α, β) , we can have smooth trade-offs between throughput and delay.

6 IMPACT OF NETWORK SIZE ON CAPACITY AND DELAY

Notice that in the literature the capacity of wireless networks is usually studied either in a dense network (with area 1) or in an extended network (with area n). In this section, we explore the impact of network size on capacity and delay in the network.

In particular, we consider a network with area n^γ , where $0 \leq \gamma \leq 1$. We still divide the network into $n^{2\alpha}$ cells and $n^{2\beta}$ squares, and the mobility model remains the same as that in Section 3.5. But, in this case, each cell has an area of $n^{\gamma-2\alpha}$ and each square has an area of $n^{\gamma-2\beta}$, respectively.

Upper bound on capacity. Following the steps in Section 5.1, (1) changes into

$$\sum_{(T_i^t, R_i^t) \in P(t)} S_E \leq n^\gamma.$$

Since $S_E = \frac{\pi \Delta^2}{4} d_{(T_i^t, R_i^t)}^2$, we can have that

$$\sum_{(T_i^t, R_i^t) \in P(t)} d_{(T_i^t, R_i^t)}^2 \leq \frac{4n^\gamma}{\pi \Delta^2},$$

and hence,

$$\begin{aligned} \sum_{(T_i^t, R_i^t) \in P(t)} d_{(T_i^t, R_i^t)} &\leq \sqrt{\sum_{(T_i^t, R_i^t) \in P(t)} d_{(T_i^t, R_i^t)}^2} \cdot \sum_{(T_i^t, R_i^t) \in P(t)} 1^2 \\ &\leq \sqrt{\frac{2n^{\gamma+1}}{\pi \Delta^2}}. \end{aligned}$$

Besides, \bar{L}_{trans} , the average cumulative transmitted distance of a bit in (2), is now

$$\begin{aligned} \bar{L}_{trans} &\geq \frac{\bar{L}}{\sqrt{2n^{\frac{\gamma}{2}-\alpha}}} \cdot \bar{L} n^{-\beta} \\ &= \frac{\bar{L}^2}{\sqrt{2}} n^{\alpha-\beta-\frac{\gamma}{2}}, \end{aligned}$$

where $\bar{L} \approx 0.52n^{\frac{\gamma}{2}}$. Therefore, similar to (3), we get

$$\lambda(n)T \cdot n \cdot \bar{L}_{trans} \leq WT \sqrt{\frac{2n^{\gamma+1}}{\pi \Delta^2}},$$

which leads to

$$\begin{aligned} \lambda(n) &\leq \frac{2W}{\sqrt{\pi} \bar{L}^2 \Delta} n^{\beta-\alpha+\gamma-\frac{1}{2}} \\ &\approx \frac{7.4W}{\sqrt{\pi} \Delta} n^{\beta-\alpha-\frac{1}{2}}, \end{aligned}$$

i.e.,

$$\lambda(n) = O(n^{\beta-\alpha-\frac{1}{2}}). \quad (10)$$

which is the same upper bound as that in (4).

Lower bound on capacity. Since a packet is transmitted at most twice in each cell, it is at most transmitted $2 \cdot (2\bar{L}/n^{\frac{\gamma}{2}-\alpha})$, i.e., $4\bar{L}n^{\alpha-\frac{\gamma}{2}}$ times in the network. Thus, every

node has to support a traffic rate of at most $4\bar{L}n^{\alpha-\frac{\gamma}{2}}\lambda(n)$ if per-node throughput is $\lambda(n)$. Besides, the transmission range of each node in this case is $\sqrt{5}n^{\frac{\gamma}{2}-\beta}$, and hence, there are $n^{1-\gamma} \cdot \pi(\sqrt{5}n^{\frac{\gamma}{2}-\beta})^2$, i.e., $5\pi n^{1-2\beta}$ nodes which share the same channel. Thus, similar to (5), a per-node throughput $\lambda(n)$ is feasible if

$$4\bar{L}n^{\alpha-\frac{\gamma}{2}}\lambda(n) \leq \frac{p_c W/3c}{5\pi n^{1-2\beta}},$$

holds. Therefore, we can have that

$$\lambda(n) = \frac{p_c W}{60\pi\bar{L}c} n^{2\beta-\alpha+\frac{\gamma}{2}-1},$$

i.e.,

$$\lambda(n) = \Omega(n^{2\beta-\alpha-1}), \quad (11)$$

which is the same result as derived in (11).

From (10) and (11), we can find that network size does not affect the capacity bounds in the network. Similarly, we can also show that it does not change the delay bounds, either, since we still have the same number of cells and squares, and hence, the same number of states in the Markov chain state space S .

7 CONCLUSION

Mobile ad hoc networks have been proved to be able to provide nondiminishing per-node throughput even when the number of nodes in the network goes to infinity. However, the price to pay is that the end-to-end delay in the network is very high, on the order of $\Theta(n)$ with possible logarithmic factors. Thus, we can only have either very low throughput, $1/\Theta(\sqrt{n})$, and short delay, $\Theta(\sqrt{n})$, in static ad hoc networks, or much higher throughput, $\Theta(1)$, and much longer delay, $\Theta(n)$, in mobile ad hoc networks. There is a big gap between the two extremes, which gives rise to an interesting question: how can we fill the gap?

This question is surely a valid question. In some scenarios, we do not need high throughput, but have a strict requirement on delay. For example, in a tactical network, we may only exchange limited and encrypted information once in a while to reduce the chance of being eavesdropped. But the information we deliver is expected to be received in time. In other cases, we expect high throughput, but can tolerate a reasonable length of delay. A delay tolerant network (DTN) or a social network can be such an example.

Aiming at addressing the above question, in this paper, we propose a new multihop relaying scheme, and investigate the throughput, delay, and mobility in wireless ad hoc networks. Instead of global mobility, we consider a more practical restricted random mobility model, and find that we can provide smooth trade-offs between throughput and delay in mobile ad hoc networks by controlling nodes' mobility pattern (α, β) . In addition, currently we only consider network delay when analyzing the trade-offs between throughput and delay. We will take queuing delay into consideration in our future work.

ACKNOWLEDGMENTS

This work was partially supported by the US National Science Foundation under grants CNS-0916391, CNS-0716450, CNS-0721744, and CNS-1147813/1147851, and the China 111 Project under grant B08038. The work of J. Li was also partially supported by Grand-in-Aid for Scientific Research from the Japan Society for Promotion of Science and Technology (JSPS), the Okawa Foundation for Information and Telecommunications. The work of X. Huang was also partially supported by the National Natural Science Foundation of China under grant 60903192 and the National High Technology Research and Development Program of China (863 Program) under grant 2011AA010503. An earlier version of this paper was presented at INFOCOM 2010 [19].

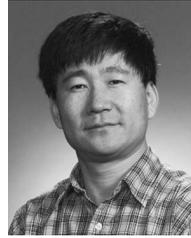
REFERENCES

- [1] S. Aeron and V. Saligrama, "Wireless Ad Hoc Networks: Strategies and Scaling Laws for the Fixed SNR Regime," *IEEE Trans. Information Theory*, vol. 53, no. 6, pp. 2044-2059, June 2007.
- [2] A. Agarwal and P. Kumar, "Capacity Bounds for Ad Hoc and Hybrid Wireless Networks," *ACM SIGCOMM Computer Comm. Rev.*, vol. 34, no. 3, pp. 71-81, July 2004.
- [3] C. Buraagohain, S. Suri, C. Toth, and Y. Zhou, "Improved Throughput Bounds for Interference-Aware Routing in Wireless Networks," *Proc. Computing and Combinatorics (COCOON '07)*, July 2007.
- [4] A. Chen and E. Gunes, "On the Cover Time of Random Geometric Graphs," *Proc. Int'l Colloquium on Automata, Languages and Programming*, July 2005.
- [5] O. Dousse, M. Franceschetti, and P. Thiran, "On the Throughput Scaling of Wireless Relay Networks," *IEEE Trans. Information Theory*, vol. 52, no. 6, pp. 2756-2761, June 2006.
- [6] E. Duarte-Melo, A. Josan, M. Liu, D. Neuhoff, and S. Pradhan, "The Effect of Node Density and Propagation Model on Throughput Scaling of Wireless Networks," *Proc. IEEE Int'l Symp. Information Theory (ISIT '06)*, July 2006.
- [7] M. Franceschetti, O. Dousse, D.N. Tse, and P. Thiran, "Closing the Gap in the Capacity of Wireless Networks via Percolation Theory," *IEEE Trans. Information Theory*, vol. 53, no. 3, pp. 1009-1018, Mar. 2007.
- [8] A. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-Delay Trade-Off in Wireless Networks," *Proc. IEEE INFOCOM*, Mar. 2004.
- [9] A. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-Delay Trade-Off in Wireless Networks—Part I: The Fluid Model," *IEEE Trans. Information Theory*, vol. 52, no. 6, pp. 2568-2592, June 2006.
- [10] A. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-Delay Trade-Off in Wireless Networks—Part II: Constant-Size Packets," *IEEE Trans. Information Theory*, vol. 52, no. 11, pp. 5111-5116, Nov. 2006.
- [11] M. Grossglauser and D. Tse, "Mobility Increases the Capacity of Ad Hoc Wireless Networks," *IEEE/ACM Trans. Networking*, vol. 10, no. 4, pp. 477-486, Aug. 2002.
- [12] P. Gupta and P. Kumar, "The Capacity of Wireless Networks," *IEEE Trans. Information Theory*, vol. 46, no. 2, pp. 388-404, Mar. 2000.
- [13] P. Gupta and P. Kumar, "Internets in the Sky: The Capacity of Three Dimensional Wireless Networks," *Comm. in Information and Systems*, vol. 1, pp. 33-50, 2001.
- [14] C. Hu, X. Wang, and F. Wu, "Motioncast: On the Capacity and Delay Tradeoffs," *Proc. ACM MobiHoc*, May 2009.
- [15] D. Knuth, *The Art of Computer Programming*. Addison-Wesley, 1998.
- [16] U. Kozat and L. Tassiulas, "Throughput Capacity of Random Ad Hoc Networks with Infrastructure Support," *Proc. ACM MobiCom*, June 2003.
- [17] P. Li and Y. Fang, "Impacts of Topology and Traffic Pattern on Capacity of Hybrid Wireless Networks," *IEEE Trans. Mobile Computing*, vol. 8, no. 12, pp. 1585-1595, Dec. 2009.

- [18] P. Li and Y. Fang, "The Capacity of Heterogeneous Wireless Networks," *Proc. IEEE INFOCOM*, Mar. 2010.
- [19] P. Li, Y. Fang, and J. Li, "Throughput, Delay, and Mobility in Wireless Ad Hoc Networks," *Proc. IEEE INFOCOM*, Mar. 2010.
- [20] P. Li, X. Huang, and Y. Fang, "Capacity Scaling of Multihop Cellular Networks," *Proc. IEEE INFOCOM*, Apr. 2011.
- [21] P. Li, M. Pan, and Y. Fang, "The Capacity of Three-Dimensional Wireless Ad Hoc Networks," *Proc. IEEE INFOCOM*, Apr. 2011.
- [22] P. Li, C. Zhang, and Y. Fang, "Capacity and Delay of Hybrid Wireless Broadband Access Networks," *IEEE J. Selected Areas in Comm.*, vol. 27, no. 2, pp. 117-125, Feb. 2009.
- [23] X. Lin, G. Sharma, R. Mazumdar, and N. Shroff, "Degenerate Delay-Capacity Tradeoffs in Ad-Hoc Networks with Brownian Mobility," *IEEE/ACM Trans. Information Theory*, vol. 52, no. 6, pp. 2777-2784, June 2006.
- [24] X. Lin and N.B. Shroff, "The Fundamental Capacity-Delay Tradeoff in Large Mobile Ad Hoc Networks," *Proc. Third Ann. Mediterranean Ad Hoc Networking Workshop (MedHoc '04)*, June 2004.
- [25] B. Liu, Z. Liu, and D. Towsley, "On the Capacity of Hybrid Wireless Networks," *Proc. IEEE INFOCOM*, Mar. 2003.
- [26] B. Liu, P. Thiran, and D. Towsley, "Capacity of a Wireless Ad Hoc Network with Infrastructure," *Proc. ACM MobiHoc*, Sept. 2007.
- [27] J. Mammen and D. Shah, "Throughput and Delay in Random Wireless Networks with Restricted Mobility," *IEEE Trans. Information Theory*, vol. 53, no. 3, pp. 1108-1116, Mar. 2007.
- [28] R. Moraes, H. Sadjadpour, and J. Garcia-Luna-Aceves, "On Mobility-Capacity-Delay Trade-Off in Wireless Ad Hoc Networks," *Proc. IEEE/ACM 12th Ann. Int'l Symp. Modeling, Analysis, and Simulation of Computer and Telecomm. Systems (MASCOTS '04)*, Oct. 2004.
- [29] M. Neely and E. Modiano, "Capacity and Delay Tradeoffs for Ad-Hoc Mobile Networks," *IEEE Trans. Information Theory*, vol. 51, no. 6, pp. 1917-1937, June 2005.
- [30] A. Ozgur, O. Leveque, and D. Tse, "Hierarchical Cooperation Achieves Optimal Capacity Scaling in Ad Hoc Networks," *IEEE Trans. Information Theory*, vol. 53, no. 10, pp. 3549-3572, Oct. 2007.
- [31] G. Sharma, R. Mazumdar, and N. Shroff, "Delay and Capacity Trade-Offs in Mobile Ad Hoc Networks: A Global Perspective," *IEEE/ACM Trans. Networking*, vol. 15, no. 5, pp. 981-992, Oct. 2007.
- [32] S. Toumpis, "Capacity Bounds for Three Classes of Wireless Networks," *Proc. ACM MobiHoc*, May 2004.
- [33] S. Toumpis and A. Goldsmith, "Large Wireless Networks Under Fading, Mobility, and Delay Constraints," *Proc. IEEE INFOCOM*, Mar. 2004.
- [34] L. Ying, S. Yang, and R. Srikant, "Optimal Delay-Throughput Trade-Offs in Mobile Ad Hoc Networks," *IEEE Trans. Information Theory*, vol. 54, no. 9, pp. 4119-4143, Sept. 2008.
- [35] A. Zemplianov and G. Veciana, "Capacity of Ad Hoc Wireless Networks with Infrastructure Support," *IEEE J. Selected Areas in Comm.*, vol. 23, no. 3, pp. 657-667, Mar. 2005.
- [36] X. Zhu, P. Li, Y. Fang, and Y. Wang, "Throughput and Delay in Cooperative Wireless Networks with Partial Infrastructure," *IEEE Trans. Vehicular Technology*, vol. 58, no. 8, pp. 4620-4627, Oct. 2009.



Pan Li received the BE degree in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2005, and the PhD degree in electrical and computer engineering from the University of Florida, Gainesville, in 2009, respectively. Currently, he is an assistant professor in the Department of Electrical and Computer Engineering, Mississippi State University. His research interests include capacity and connectivity analysis, medium access control, routing algorithms, and cross-layer optimization and design in wireless networks. He is a member of the IEEE and the ACM.



Yuguang Fang received the PhD degree in systems engineering from Case Western Reserve University in January 1994 and the PhD degree in electrical engineering from Boston University in May 1997. He was an assistant professor in the Department of Electrical and Computer Engineering at the New Jersey Institute of Technology from July 1998 to May 2000. He then joined the Department of Electrical and Computer Engineering at the University of Florida in May 2000 as an assistant professor, got an early promotion to an associate professor with tenure in August 2003, and to a full professor in August 2005. He held the University of Florida Research Foundation (URFF) professorship from 2006 to 2009. He is also active in professional activities. He has served on several editorial boards of technical journals, including *IEEE Transactions on Communications*, *IEEE Transactions on Wireless Communications*, *IEEE Transactions on Mobile Computing*, and *ACM Wireless Networks*. He has been actively participating in professional conference organizations such as serving as the steering committee cochair for QShine, the technical program vice-chair for IEEE INFOCOM 2005, technical program symposium cochair for IEEE GlobeCom 2004, and a member of the technical program committee for IEEE INFOCOM (1998, 2000, 2003-2010). He has published more than 200 papers in refereed professional journals and conferences. He received the US National Science Foundation Faculty Early Career Award in 2001 and the US Office of Naval Research Young Investigator Award in 2002. He was the recipient of the Best Paper Award from the IEEE International Conference on Network Protocols (ICNP) in 2006 and the recipient of the IEEE TCGN Best Paper Award from the IEEE High-Speed Networks Symposium at IEEE GlobeCom in 2002. He is a fellow of the IEEE and a member of the ACM.



Jie Li received the BE degree in computer science from Zhejiang University, Hangzhou, China, the ME degree in electronic engineering and communication systems from the China Academy of Posts and Telecommunications, Beijing, China, and the Dr. Eng degree from the University of Electro-Communications, Tokyo, Japan. He has been with the University of Tsukuba, Japan, where he is a full professor. His research interests include mobile distributed multimedia computing and networking, OS, network security, modeling, and performance evaluation of information systems. He has served as a secretary for the Study Group on System Evaluation of IPSJ (Information Processing Society of Japan), on several editorial boards including the *IPSJ Journal*, and on the steering committees of the SIG of System Evaluation (EVA) of IPSJ, the SIG of DataBase System (DBS) of IPSJ, and the SIG of MoBiLe computing and ubiquitous communications of IPSJ. He has been a cochair of several international symposia and workshops. He has also served on the program committees for several international conferences such as IEEE ICDCS, IEEE INFOCOM, IEEE GlobeCom, and IEEE MASS. He received the best paper award from IEEE NAECON 1997. He is a senior member of the IEEE and the ACM, and a member of IPSJ.



Xiaoxia Huang received the BE and ME degrees in electrical engineering from the Huazhong University of Science and Technology, China, in 2000 and 2002, respectively, and the PhD degree in electrical and computer engineering from the University of Florida in 2007. Currently, she is an associate researcher at the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences. Her research interests include wireless sensor networks, wireless communications, and mobile computing. She is a member of the IEEE.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.