Throughput, Delay, and Mobility in Wireless Ad Hoc Networks

Pan Li
Mississippi State University.
Mississippi State, MS 39762-9571
Email: li@ece.msstate.edu

Yuguang Fang
University of Florida, Gainesville, FL 32611
Xidian University, Xi’an, China
Email:fang@ece.ufl.edu

Jie Li
University of Tsukuba,
Tsukuba Science City, Japan
Email: lijie@cs.tsukuba.ac.jp

Abstract—Throughput capacity in wireless ad hoc networks has been studied extensively under many different mobility models such as i.i.d. mobility model, Brownian mobility model, random walk model, etc. Most of these research works assume global mobility, i.e., each node moves around in the whole network, and the results show that a constant per-node throughput can be achieved at the cost of very high expected average end-to-end delay. Thus, we are having a very big gap here, either low throughput and low delay in static networks or high throughput and high delay in mobile networks. In this paper, employing a more practical restricted random mobility model, we try to fill in this gap. Specifically, we assume a network of unit area with \( n \) nodes is evenly divided into \( n^{-\alpha} \) 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 \( 0 \leq \alpha \leq \beta \leq \frac{1}{2} \). 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. Proposing a new multi-hop relay scheme, we present an upper bound and a lower bound on per-node throughput capacity and expected average end-to-end delay, respectively. We finally explicitly show smooth trade-offs between throughput and delay by controlling nodes' mobility.

I. INTRODUCTION

There has been intensive interest in the study of the capacity of wireless networks since the seminal work by Gupta and Kumar [10]. This work has 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. [5] prove by percolation theory that the same \( 1/\sqrt{n} \) per-node throughput can also be achieved in random ad hoc networks. By allowing an arbitrary small fraction of the nodes to be disconnected, Dousse et al. [3] show that the throughput cannot be improved much. Different from the above studies, Duarte-Melo et al. [4] look into the case of semi-extended networks, where both node density and network area increase as the number of nodes \( n \) increases. Specifically, they assume the network area is a disk of radius \( n^{\gamma} \), \( 0 < \gamma < \frac{1}{2} \). With a \( \frac{1}{1+d^n} \) 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.

Aeron et al. [1] and Ozgur et al. [23] then study the scaling laws for wireless networks employing MIMO (Multi-Input-Multi-Output) techniques, respectively, and show that the per-node throughput can be increased significantly. However, we have to bear with the increased complexity due to the intelligent collaboration among nodes. Furthermore, many researchers also explore the capacity of hybrid wireless networks, i.e., ad hoc networks with partial infrastructure, like [12]–[14], [18], [19], [25], [27]. The results show that throughput can only be enhanced with a large number of base stations which means very high deployment cost. Recently, Li et al. [15] explore the capacity of heterogeneous wireless networks and find that the throughput can be much higher than that in pure ad hoc networks.

Different from all the above works which study static networks, in this paper, we investigate how to achieve higher throughput by leveraging nodes’ mobility. Grossglauser and Tse [9] 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 [9], there have been extensive works dealing with the trade-off between capacity and delay under different mobility models. Lin et al. [16] study the throughput capacity under Brownian mobility model, and show that the two-hop relaying scheme in [9] can achieve a per-node throughput of \( \Theta(1) \) with an expected packet delay of \( \Omega(\log n/\sigma^2) \) where \( \sigma^2 \) 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. [7], [8] have looked into the throughput capacity under random walk model, and have proposed a similar two-hop relaying scheme to that in [9] and shown that it can achieve a per-node throughput of \( \Theta(1) \)

This work was partially supported by the U.S. National Science Foundation under grants CNS-0916391, CNS-0716450 and CNS-0721744, and China 111 Project under grant B08038. The work of J. Li was also partially supported by Grand-in-Aid for Scientific Research from Japan Society for Promotion of Science and Technology (JSPS), the Okawa Foundation for Information and Telecommunications.

Y. Fang is also a Changjiang Scholar Chair Professor with National Key Laboratory of Integrated Services Networks, Xidian University, China.
with a delay of $\Theta(n \log n)$. Besides, [17], [22], [26] 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. [24] examine the problem under hybrid mobility models and show that there are trade-offs between capacity and delay.

We notice that all the aforementioned research works on the throughput capacity in wireless networks assume 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 in 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, Mannen et al. study in [20] the throughput capacity under 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 capacity under a more practical restricted random mobility model. We assume the network of unit area with $n$ nodes is evenly divided into $2^{2\alpha}$ cells with an area of $n^{-2\alpha}$, where $0 \leq \alpha \leq 1/2$, each of which is further evenly divided into squares with an area of $n^{-2\beta}$, where $0 \leq \alpha \leq \beta \leq 1/2$. 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 $O(n^{\beta - \alpha - 2})$. Then, by employing a new multi-hop relay scheme, we show that the per-node throughput capacity can be lower bounded by $\Omega(n^{\beta - \alpha - 1})$. Besides, we also find that the expected average end-to-end packet delay can be bounded by $O(n^{\beta - \alpha + \frac{1}{2}} \log n)$ and $\Omega(n^{\beta - \alpha} \log n)$ when $0 \leq \alpha < \beta \leq \frac{1}{2}$, and by $O(n^{\beta - \alpha + \frac{1}{2}})$ and $\Omega(n^{\beta - \alpha})$ when $0 \leq \alpha = \beta \leq \frac{1}{2}$, respectively. Notice that the result in [7], [16], [20], 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 result, 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 $\lambda \in \left[\frac{1}{\sqrt{n}}, 1\right]$ by using a certain mobility pattern ($\alpha, \beta$), the expected average packet delay will be bounded by $\Omega(n \log n)$ and $O(n^{\sqrt{\lambda(n)} \log n})$ when $0 \leq \alpha < \beta \leq \frac{1}{2}$, and by $\Omega(n \log n)$ and $O(n^{\sqrt{\lambda(n)}})$ when $0 \leq \alpha = \beta \leq \frac{1}{2}$. On the other hand, if we want to upper bound the delay by $E[D]$ where $E[D] \in \left[\sqrt{n}, n\right]$ when using a certain mobility pattern ($\alpha, \beta$), then the per-node throughput $\lambda(n)$ will be bounded by $\Omega\left(\frac{E[D]}{n^{\beta - 2\alpha}}\right)$ and $O\left(\frac{E[D]}{n^{\beta - \alpha}}\right)$ when $0 \leq \alpha < \beta \leq \frac{1}{2}$, and by $\Omega\left(\frac{E[D]}{n^{\beta - 2\alpha}}\right)$ and $O\left(\frac{E[D]}{n^{\beta - \alpha}}\right)$ when $0 \leq \alpha = \beta \leq \frac{1}{2}$.

Moreover, notice that most of the previous research on mobile ad hoc networks just shows that a constant per-node throughput can be achieved with very large average delay. Thus, we 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. While in this paper, by restricting nodes’ mobility, we can provide a smooth trade-off between throughput and delay in the network.

The rest of this paper is organized as follows. In Section II we introduce some definitions and models we will use in this study. Section III presents our proposed multi-hop relay scheme. In Section IV, 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. We finally conclude this paper in Section V.

II. Definitions and Assumptions

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

A. Definitions

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.

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/sec.

Per-node Throughput Capacity: We say that the per-node throughput capacity in the network (12) is of order $O(f(n))$ bits per second if there is a deterministic constant $0 < c_1 < +\infty$ such that

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

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

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

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

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 [16], [21], 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.$^2$


$^2$As pointed out in [7], [17], under this assumption, queuing delay at source node can actually be a constant, which gives us another reason to focus on network delay.
where 

\[ s \text{atisfy: } n \]

D. Restricted Random Mobility Model

is satisfied, and 0 otherwise.

\[ W \]

B. Network Model

As shown in Fig. 1(a), the unit torus of \( n \) nodes is evenly divided into \( n^{2\alpha} \) squares each of which has an area of \( n^{-\alpha} \), 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 \( 0 \leq \beta \leq \frac{1}{2} \), as shown in Fig. 1(b). 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 an 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 two-dimensional array, as shown in Fig. 1(b). For a square denoted by \((x, y)\), 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.

Notice that there are four extreme cases of special interests.

- 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 [17], [22], [26].
- 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 divided into \( n \) squares each of which has an area of \( \frac{1}{2} \). Each node can only move around within these very small squares, which is just like a network with static nodes as studied in [5], [10].
- Case iv), \( \alpha = 0, \beta = \frac{1}{2} \). In this case, nodes can move around the whole network with each square of area \( \frac{1}{2} \). 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 [7], [8]. Note that random walk model is the discrete time version of the Brownian motion model considered in [16].

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.

\[ \Delta \]

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/sec when the protocol model is satisfied, and 0 otherwise.

C. Interference Model

We employ the Protocol Model in [10] 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/sec when the protocol model is satisfied, and 0 otherwise.

D. Restricted Random Mobility Model

As shown in Fig. 1(a), the unit torus of \( n \) nodes is evenly divided into \( n^{2\alpha} \) squares each of which has an area of \( n^{-\alpha} \), 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 \( 0 \leq \alpha \leq \frac{1}{2} \). 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 an 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 two-dimensional array, as shown in Fig. 1(b). For a square denoted by \((x, y)\), 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.

Notice that there are four extreme cases of special interests.

- 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 [17], [22], [26].
- 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 divided into \( n \) squares each of which has an area of \( \frac{1}{4} \). Each node can only move around within these very small squares, which is just like a network with static nodes as studied in [5], [10].
- Case iv), \( \alpha = 0, \beta = \frac{1}{2} \). In this case, nodes can move around the whole network with each square of area \( \frac{1}{4} \). 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 [7], [8]. Note that random walk model is the discrete time version of the Brownian motion model considered in [16].

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 ([7], [8], [24]) 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.

III. A Multi-hop Relay Scheme

Before we move on with our main results, we first present our multi-hop relay scheme for the transmissions between source-destination pairs:

S-I: Each square in the network becomes active in every $c \geq 1$ time slots. 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 sub-slots A, B, and C, of the same length. For a square with $N$ nodes, each sub-slot is further divided into $N$ mini-slots of the same length.

- In sub-slot A, if $N \geq 2$, every node acting as source transmits one-by-one in a mini-slot 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.
- In sub-slot B, if this active square is an interior square, nothing happens. Otherwise, i.e., 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 mini-slot 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 sub-slot C, if $N \geq 2$, every node acting as a relay transmits one-by-one in a mini-slot 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 $W/(4[1 + \Delta] + 3)^2$. Note that $[x]$ is the ceiling function of $x$, which returns the smallest integer not smaller than $x$, or formally,

$$[x] = \min\{n \in \mathbb{Z} \mid n \geq x\},$$

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 $(4[1 + \Delta] + 3)^2$ squares. As shown in Fig. 2, the $(4[1 + \Delta] + 3)^2$ squares in each set are numbered from 1 to $(4[1 + \Delta] + 3)^2$ in the same way. We further divide time into sequences of successive slots, denoted by $t$ ($t = 0, 1, 2, 3, ...$). During a slot $t$, all squares that are numbered $[t \mod (4[1 + \Delta] + 3)^2] + 1$ 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 II-D. 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

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

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

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

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 $(4[1 + \Delta] + 3)^2$ 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 $W/(4[1 + \Delta] + 3)^2$.

Thus, from Lemma 1, we can have that $c = (4[1 + \Delta] + 3)^2$ in S-I. Note that the transmission rate of each square is obtained for a square assuming one node is transmitting in
Similarly, a transmission is considered to be successful if the nodes’ mobility pattern are controlled by two parameters: $\alpha$ and $\beta$. 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 $\alpha$ and $\beta$ in Section IV-A and Section IV-B, respectively, when $0 < \alpha < \beta < \frac{1}{2}$. We then study the capacity and the delay in four extreme cases in Section IV-C. Finally, we discuss the trade-offs among capacity, delay, and mobility in Section IV-D.

### B. 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

$$|R^i_t - R^j_t| \geq |T^i_t - R^j_t| - |T^j_t - R^i_t| \geq (1 + \Delta) |T^i_t - R^j_t| - |T^j_t - R^i_t|,$$

Similarly,

$$|R^j_t - R^i_t| \geq |T^j_t - R^i_t| - |T^i_t - R^j_t| \geq (1 + \Delta) |T^j_t - R^i_t| - |T^i_t - R^j_t|.$$

Adding these two inequalities together, we can obtain

$$|R^i_t - R^j_t| \geq \frac{\Delta}{2}(|T^i_t - R^j_t| + |T^j_t - R^i_t|).$$

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

$$\sum_{(T^i_t, R^i_t) \in P(t)} S_E \leq 1$$

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, exclusion area is equal to the area of disjoint disks, i.e.,

$$S_E = \frac{\pi \Delta^2}{4} \left( \frac{d^2(T^i_t, R^i_t)}{2} \right)$$

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

$$\sum_{(T^i_t, R^i_t) \in P(t)} d^2(T^i_t, R^i_t) \leq \frac{4}{\pi \Delta^2}.$$

Besides, since at any time, there are at most $\frac{\pi}{4}$ concurrent transmitter-receiver pairs, according to Cauchy-Schwartz inequality, we have

$$\sum_{(T^i_t, R^i_t) \in P(t)} d^2(T^i_t, R^i_t) \leq \frac{\pi}{4} \sum_{(T^i_t, R^i_t) \in P(t)} d^2(T^i_t, R^i_t) \sum_{(T^j_t, R^j_t) \in P(t)} d^2(T^j_t, R^j_t) \leq \frac{2n}{\pi \Delta^2}.$$

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, and hence 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 destination node is at least $\frac{1}{2} n^{-\beta}$ due to assumption (A1), the average cumulative transmitted distance of a bit, denoted by $\bar{L}_{trans}$, can be calculated as

$$\bar{L}_{trans} \geq \frac{\bar{L}}{\sqrt{2n^{-\alpha}}} \cdot \frac{1}{4} n^{-\beta} = \frac{\bar{L}}{4\sqrt{2}} n^{-\beta} \cdot n^{-\beta} = \frac{\bar{L}}{4\sqrt{2}} n^{-\beta}.$$

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

$$\lambda(n) T \cdot n \cdot \bar{L}_{trans} \leq WT \sum_{(T^i_t, R^i_t) \in P(t)} d(T^i_t, R^i_t) \leq WT \frac{2n}{\pi \Delta^2}.$$

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.
\[ \lambda(n) \leq \frac{8W}{\sqrt{n}L\Delta} n^{\beta - \alpha - \frac{1}{4}}, \]

i.e.,
\[ \lambda(n) = \Theta(n^\beta - \alpha - \frac{5}{2}). \] (4)

We next derive a lower bound on the throughput capacity based on the proposed relay scheme. We also choose a routing protocol so that the cells a packet goes through approximately lie on the same straight line connecting the two cells where the source node and the destination node are located, respectively. Since a packet is transmitted at most twice in each cell, a packet is at most transmitted \(2 \cdot (2L/n - \alpha)\), i.e., \(4Ln^\alpha\) 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 \(4Ln^\alpha \lambda(n)\). Besides, since the transmission range of each node is chosen to be \(\sqrt{5n^{-\beta}}\), there are \(n \cdot (\sqrt{5n^{-\beta}})^2\), i.e., \(5n^{1-2\beta}\) nodes which will share the same channel. From Lemma 2 and Lemma 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 \(p_cW/3c\). Thus, a per-node throughput \(\lambda(n)\) is feasible if

\[ 4L\ln \lambda(n) \leq \frac{p_cW/3c}{5n^{1-2\beta}}, \]

holds\(^4\). Thus, we can obtain that
\[ \lambda(n) = \frac{p_cW}{60Le} n^{2\beta - \alpha - 1} \]

bits per second is feasible, and hence
\[ \lambda(n) = \Omega(n^{2\beta - \alpha - 1}). \] (5)

Notice that in (4) and (5), we have
\[ (\beta - \alpha - \frac{1}{2}) - (2\beta - \alpha - 1) = \frac{1}{2} - \beta > 0, \]

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.

B. 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 \(\Pi\). We have the following definition [24].

**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 \(\Pi\).

Besides, we have the following result on the first hitting time for a single state in case of a two-dimensional torus containing \(n\) nodes uniformly distributed in \(\sqrt{n} \times \sqrt{n}\) squares [24]:

4A similar approach is employed in [10] [24].

**Lemma 4:** Let \(T^A\) denote the first hitting time for a single state \(A\) on a two-dimensional 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 find a lower bound on the average packet delay by using a routing protocol such that the cells a packet goes through approximately compose a straight line connecting the two cells where the source node and the destination node are located, respectively. One such an example is shown in Fig. 3.

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 delays as in [24]. Thus, 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 in 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. Besides, we ignore the queuing delay since Lin et al. [17] have shown that queueing delay at source nodes and relay nodes can be bounded by a constant under certain conditions.

Denote the time needed for the relay node in source cell to firstly 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
\[ p_0 = 1 - (1 - n^{-2\beta})^n = 1 - e^{-n \log(1 - n^{-2\beta})} = 1 - \frac{O(1)}{en^{1-2\beta}}, \]

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 relay cells, and \(E[T_R]\) the expectation. Similarly, we can also get
\[ E[T_R] = \Omega(n^{2\beta - 2\alpha \log n}). \]

Finally, let \(T_D\) denote the time spent in destination cell, and \(E[T_D]\) the expectation. Since we assume 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 \(L = \Theta(1)\). Denote by \(H\) the average number
satisfies assumption (A2), a packet is relayed from one cell to another.

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

\[
E[D] = E[T_S] + E[H]E[T_R] + E[T_D]
\]

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

Next, we attempt to find an upper bound on \( E[D] \). Chen et al. [2] show that the cover time\(^5\) of a 2-dimensional 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 into the boundary squares of the current cell. Thus, we can have

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

Similar to (3), we can also have

\[
\lambda(n)T \cdot n \cdot E[H] \cdot \frac{1}{4} n^{-\beta} \leq WT \sqrt{\frac{2n}{\pi \Delta^2}},
\]

i.e.,

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

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.,

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

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

\[
(\beta - \alpha + \frac{1}{2}) - (2\beta - \alpha) = \frac{1}{2} - \beta > 0,
\]

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.

\(^5\)Cover time is the expected time taken by a simple random walk as we assume here on a graph \( G \) to visit all nodes in \( G \).

C. Extreme Cases of Special Interests

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

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

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

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 [22] 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 Section IV-A and Section IV-B, we can obtain

\[
\lambda(n) = O(n^{-\frac{1}{2}}), \ E[D] = \Omega(n^{\alpha-1}),
\]

and

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

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

Along the line in Section IV-A and Section IV-B, we can obtain that

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

Note that in this case, our mobile network is just like a static network as studied in [5], [10]. Gupta and Kumar [10]
have shown that a static random network can achieve per-node throughput of $\sqrt{n \log n}$, while Franceschetti et al. [5] 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: $\alpha = 0$, $\beta = 1/2$.

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 [7].

Combining with the results obtained in Section IV-A and Section IV-B, 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

$$\lambda(n) = \begin{cases} 
\Omega(n^{2\beta - \alpha - 1}) & \\
O(n^{\beta - \alpha - \frac{1}{2}}) \end{cases} \quad \text{if } \alpha < \beta \leq \frac{1}{2},$$

with $\alpha, \beta$ being bounded as

$$\lambda(n) = \begin{cases} 
\Omega(n^{1/2}) & \\
O(n^{\beta - \alpha - 1/2}) \end{cases} \quad \text{if } 0 \leq \alpha = \beta \leq \frac{1}{2}.$$

D. Trade-offs Among Capacity, Delay, and Mobility

Recall that two parameters $\alpha$ and $\beta$ together define a mobility pattern, and $\beta$ 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 $\beta$ while decreasing $\alpha$, 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 protocol. So, from Theorem 1 and Theorem 2, we can obtain that when the per-node throughput of $\lambda(n) = n^{2\beta - \alpha - 1}$ bits/sec 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 $\beta \geq (x + 1)/2$. Then, the exponential order of the upper 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 $0 \leq \alpha = \beta = \frac{1}{2}$.

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

$$E[D] = \begin{cases} 
\Omega(n \lambda(n) \log n) & \\
O(n \sqrt{\lambda(n) \log^2 n}) \end{cases} \quad \text{when } 0 \leq \alpha < \beta \leq \frac{1}{2},$$

and being bounded as

$$E[D] = \begin{cases} 
\Omega(n \lambda(n)) & \\
O(n \sqrt{\lambda(n)}) \end{cases} \quad \text{when } 0 \leq \alpha = \beta \leq \frac{1}{2}.$$

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 $\beta \geq y - 1/2$. 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 that 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

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

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

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

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

$$\lambda(n) = \begin{cases} 
\Omega \left( \frac{E[D]}{n^4 \log^2 n} \right) & \\
O \left( \frac{E[D]}{n \log^2 n} \right) \end{cases} \quad \text{when } 0 \leq \alpha < \beta \leq \frac{1}{2},$$

and that of the lower bound on the expected average packet delay is

$$2\beta - \alpha = x + 1.$$
when $0 \leq \alpha < \beta \leq \frac{1}{2}$, and as

$$\lambda(n) = \begin{cases} \Omega(\frac{E[D]^2}{n^2}) & \text{if } \alpha < \beta < \frac{1}{2} \\ \Omega(E[D]) & \text{if } \alpha = \beta = \frac{1}{2} \\ O(\frac{E[D]}{n}) & \text{if } \alpha = \beta > \frac{1}{2} \end{cases}$$

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

We further illustrate the trade-offs among capacity, delay, and mobility more explicitly in Fig. 4, ignoring the logarithmic terms as in [16]. Our results show that by controlling nodes’ mobility pattern $(\alpha, \beta)$, we can have smooth trade-offs between throughput and delay.

V. CONCLUSION

In this paper, we have investigated 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 $(\alpha, \beta)$.

REFERENCES


