

UC Santa Cruz

UC Santa Cruz Previously Published Works

Title

Capacity and Delay Analysis for Ad Hoc and Sensor Networks

Permalink

<https://escholarship.org/uc/item/2x9360b3>

Author

Garcia-Luna-Aceves, J.J.

Publication Date

2007

Peer reviewed

Capacity and Delay Analysis for Ad Hoc Networks

Renato M. de Moraes, Hamid R. Sadjadpour, and J.J. Garcia-Luna-Aceves

Abstract

Capacity and delay are tightly related entities in ad hoc and sensor networks. In the past few years, there has been considerable amount of work describing how these entities behave given a set of assumptions. It was shown that the improvement of one is obtained at the expense of the other; that is, there is a trade-off between them. On the other hand, there are other important parameters such as mobility and physical layer properties (e.g., the use of directional antennas) that can change the way these two entities are associated to each other. Furthermore, diversity routing strategies like multi-copy packet forwarding has proven to enhance delay without changing the capacity order of such networks. On this chapter, we present a summary of main results on the trade-off between capacity and delay for ad hoc and sensor networks. First, we review the basic concepts employed throughout the literature, followed by the fundamental results for capacity and delay in ad hoc networks. Then, we show how mobility can be exchanged with capacity and delay, as well as how the physical layer features change such trade-off. Finally, we describe how a multi-copy packet relaying strategy provides an exponential delay reduction in mobile ad hoc networks (MANETs) compared with the traditional single-copy relaying scheme without changing the capacity order of such networks.

This work was supported in part by CAPES/Brazil, by the US Army Research Office under grant W911NF-05-1-0246, and by the Basking Chair of Computer Engineering.

R. M. de Moraes is with Department of Computing Systems, University of Pernambuco, Brazil. Email: renato@dsc.upe.br

H. R. Sadjadpour is with Department of Electrical Engineering, University of California at Santa Cruz (UCSC). Email: hamid@soe.ucsc.edu

J.J. Garcia-Luna-Aceves is with Department of Computer Engineering at UCSC and with Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304, USA. Email: jj@soe.ucsc.edu

I. INTRODUCTION

There are many different scenarios for wireless networks. In a traditional cellular wireless network, the nodes (or users) communicate with each other through base stations (or access points). These stations provide access control in the network and play a central role, such as frequency management, power control, etc. The base stations are fixed and form a network infrastructure while the nodes are either mobile or stationary. We can find such networks in the common cellular phone systems, for example.

Wireless ad hoc and sensor networks require no base station and all the control and access tasks are distributed among nodes acting as peers. This makes them attractive in situations where there is no fixed infrastructure, such as remote sensing, battle fields or catastrophe-relief efforts. However, the lack of centralized control imposes significant challenges for sensor and ad hoc network designers and there are many open problems, questions like throughput, delay, storage trade-offs, capacity bounds, link lifetime, and their interactions. Also, the network modeling issue is a research challenge, and more complete and real models are necessary to help the network analysis. Therefore, the performance analysis of wireless networks is a key issue in the field of network communications and requires careful modeling and attention. In addition, the demand for higher data rates in fixed or mobile wireless sensor and ad hoc networks offer challenging scalability issues. Consequently, there have been considerable efforts [1], [2], [3] [4], [5], [6], [7] to increase the performance of wireless ad hoc networks since Gupta and Kumar [8] showed that the capacity of a fixed wireless network decreases as the number of nodes increases when all the nodes share a common wireless channel. More precisely, they showed that the node throughput ¹ of a fixed wireless network scales as $\Theta\left(1/\sqrt{n \log(n)}\right)^2$ for n total users (or nodes) in the network, while the associated packet delivery delay grows as $\Theta\left(\sqrt{n/\log(n)}\right)$ [9]. This is a disappointing result, because both the capacity and delay degrade as the number of nodes in the network increases. On the other

¹The precise definitions for node throughput and delivery delay will be given in the next section.

²Here we use the notation: (a) $b(n) = O(h(n))$ means there are positive constants c and m , such that $0 \leq b(n) \leq c h(n) \forall n > m$. (b) $b(n) = \Omega(h(n))$ means there are positive constants c_1 , and m_1 , such that $0 \leq c_1 h(n) \leq b(n) \forall n > m_1$. (c) $b(n) = \Theta(h(n))$ means there are positive constants c_2 , c_3 , and m_2 , such that $0 \leq c_2 h(n) \leq b(n) \leq c_3 h(n) \forall n > m_2$. Also, $\log(\cdot)$ stands for the natural logarithm.

hand, Grossglauser and Tse [1], [10] demonstrated that the capacity and delay of wireless mobile ad hoc networks (MANETs) scale as $\Theta(1)$ and $\Theta(n)$ [9], respectively, by utilizing mobility and a multiuser diversity scheme [11] together with a two-phase relaying strategy. The importance of this result is that it proves that trade-offs can be made to substantially improve the capacity of wireless networks.

Accordingly, a review on the main results for performance of wireless ad hoc and sensor networks is the aim of Sections II, III and IV. The scope of this chapter is to revise analytical methods for computing the throughput and delay of these networks. We revisit the main models that are used for the analysis of such networks, and use the Information Theory tools for computation of capacity in network communication systems. Moreover, different schemes to improve the overall behavior of wireless networks is studied. In Section VIII, multi-copy forwarding of packets is presented to reduce delay of MANETs without changing the $\Theta(1)$ capacity scalability order. Also, a mathematical formula is provided for the throughput as a function of the network parameters. In Section IX, the capacity-delay trade-off of ad hoc wireless networks is investigated, where it is found that mobility can be exchanged with capacity and delay. Furthermore, we demonstrate that by changing the physical layer properties of the ad hoc network (e.g., considering restrained mobility or directional antennas for the nodes), the capacity or delay behavior can be improved. Finally, Section X concludes the chapter.

II. BASIC DEFINITIONS

In order to investigate capacity and delay tradeoff of wireless networks, the following definitions are used throughout the chapter.

A *node throughput* (or per source-destination throughput, or simply throughput) of $\lambda(n)$ bits per second is feasible if every node can send information at an average rate of $\lambda(n)$ bits per second to its chosen destination [8], [9].

The *delivery delay* (or simply delay, or latency) of a packet in a network is the time it takes the packet to reach the destination after it leaves the source, where queuing delay at the source is not considered.

The average packet delay for a network with n nodes is obtained by averaging over all packets, all source-destination pairs, and all random network configurations [9].

The *interference* at a node j , when node i transmits to node j through a frequency bandwidth W Hz, is defined as the power of the signals (in units of Watts) from all transmitting nodes in the network in the sub-spectrum W , except node i .

The term *half-duplex* means that a node cannot transmit and receive data simultaneously through the same frequency bandwidth.

The term *cell* denotes the set of nodes located inside a defined area of the network.

Basically, the wireless ad hoc networks can be classified in either static, mobile, or hybrid, according to the movement of the nodes. In general, the mobile networks present more complexity than the static ones, because with mobility the topology of network is constantly changing, which imposes control challenges for communication. In either case, the capacity and delay behavior can result quite different as it will be discussed later.

III. STATIC WIRELESS NETWORKS

In a fixed(or static) wireless network, the nodes are stationary and their locations can be deployed arbitrary, random (i.e., according to a two-dimensional Poisson distribution), or regularly spaced (like in a grid) in the network.

A. The Fundamental Scalability Results

The capacity of static wireless networks when the number of total nodes in the network scale to infinite can be analyzed [8]. Let's consider the network model as a sphere of unit area containing n total fixed nodes with identical properties. Nodes were placed either arbitrarily or randomly in the area. Communication among nodes is obtained through a single wireless channel shared among all nodes, and therefore subject to interference. Packets are sent from source to destination in a multihop fashion following the path close to the straight line that links the source to its destination. Therefore, each node could function as source,

relay and destination of packets. It is shown [8] that there exists a Voronoi tessellation \mathcal{V}_n on the unit sphere surface satisfying the following properties:

- Every Voronoi cell V contains a disk of area $100 \log(n)/n$ and corresponding radius $\epsilon(n) = c_4 \sqrt{\log(n)/n}$, for some positive constant c_4 .
- Every Voronoi cell is contained within a circle of radius $2\epsilon(n)$.

Each Voronoi cell $V \in \mathcal{V}_n$ is simply a cell of the network, and the cells do not have a regular shape because the network is arbitrary or random. With this tessellation, each cell contains at least one node with high probability (whp)³ which meets the connectivity requirement [8]. Furthermore, by choosing the transmission range equal to $8\epsilon(n)$ for each node, it allows direct communication within a cell and between adjacent cells. Accordingly, two cells are interfering neighbors if there is a point in one cell that is within a distance $(2 + \Delta)8\epsilon(n)$ of some point in the other cell, in which $\Delta > 0$ is a given constant modeling condition where a guard zone is required to prevent a neighboring node from transmitting on the same channel at the same time [8]. These are two models for successful communication among nodes. Let X_i denote the location of a node i in the network. The *Protocol Model* establishes that node i transmits successfully to node j if the following condition is satisfied

$$|X_k - X_j| \geq (1 + \Delta)|X_i - X_j|, \quad (1)$$

so that transmit node X_k will not impede X_i and X_j communication. In the *Physical Model* node i transmits successfully to node j if the signal to noise and interference ratio (*SNIR*) at node j satisfies

$$SNIR = \frac{\frac{P_i}{|X_i - X_j|^\alpha}}{N_0 + \sum_{k \neq i} \frac{P_k}{|X_k - X_j|^\alpha}} \geq \beta, \quad (2)$$

where P_i is the transmit power of node i , α is the path loss parameter, N_0 is the noise power, and β is the minimum value of *SNIR* necessary for successful reception. It is known that if $\alpha > 2$ and each node transmits at same power, then the *Protocol* and *Physical* models are equivalent [9].

³With high probability means with probability $\geq 1 - \frac{c_5}{n}$, for some positive constant c_5 [12].

By using the previous two communication models, the node throughput of fixed wireless networks scale as $\Theta(1/\sqrt{n})$ for the arbitrary placement of nodes, and as $\Theta\left(1/\sqrt{n \log(n)}\right)$ for the random placement of nodes [8]. In either case, the capacity of each node decreases as the number of total nodes n in the network increases. This poor performance is mainly due to interference among nodes.

B. The Deterministic Approach

The Gupta and Kumar's results can also be obtained using different techniques. For example, Kulkarni and Viswanath [13] employed a deterministic approach to throughput scaling in static wireless networks where they divided the network area in equal square cells (the squarelets) which size were function of n . By defining an equivalent class of squarelets such that nodes sufficiently far apart can transmit without causing interference to each other, they applied fundamental statistics concepts such as Chernoff bounds and Borel-Cantelli Lemma to find a stronger (almost sure) version of the $1/\sqrt{n \log(n)}$ throughput for nodes located randomly in the network.

C. Wireless Backbone Approach

Franceschetti et al [14] proved that a node throughput of $\Theta(1/\sqrt{n})$ can be achieved in a static random network, if a wireless backbone is rich enough in crossing paths such that it transports all traffic of the static network, although it does not cover all the nodes. The authors proved their approach by using percolation theory arguments.

D. Effects of Physical Layer Properties

The above results obtained for wireless ad hoc networks can be substantially different if the physical layer assumptions are changed, for example, by adopting very large bandwidth with a minimum power routing for communication or directional antennas for transmission and reception.

1) *The Minimum Power Routing:* Shepard [15] was the first to note that if the minimum power route for packets is considered along with an efficient distributed channel access technique, then a scalable static

wireless network is feasible. Later, Negi and Rajeswaran [16] provided an ultra wide band (UWB) multiple access technique, using the minimum power routing idea, and they proved that a wireless network can indeed have a capacity that grows with the total number of nodes. They showed that appropriate change in the physical layer assumptions, for example, using an integrated CDMA MAC and UWB, can improve the throughput of wireless ad hoc and sensor networks significantly. More specifically, they found that uniform per node throughput of $O\left((n \log(n))^{\frac{\alpha-1}{2}}\right)$ (upper bound) and $\Omega\left(\frac{n^{(\alpha-1)/2}}{(\log(n))^{(\alpha+1)/2}}\right)$ (lower bound) with associated bandwidth expansion of $\Theta(n(n^2 \log(n))^{\frac{\alpha}{2}})$ can be achieved.

2) *Directional Antennas*: Channel access techniques for directional antennas has been proposed for ad hoc wireless networks [17], [18]. Yi et al [2] investigated the capacity scalability for an extended model from Gupta and Kumar [8] where the nodes were endowed with directional antennas. Consider a directional antenna model (illustrated in Fig. 1) as a sector characterized by (i) the transmission (Tx)/reception (Rx) range r and (ii) the beamwidth γ for transmission or β for reception. There is also the omni-directional antenna case.

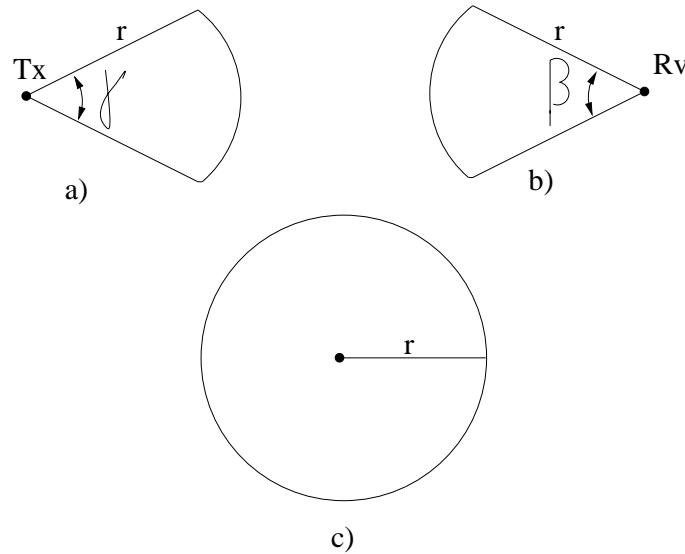


Fig. 1. a) Directional transmission (Tx) antenna model. b) Directional receiver (Rv) antenna model. c) Omni-directional antenna for Tx or Rv.

Using such antenna models, the following Theorem is obtained [2].

Theorem 1 For static random wireless networks, there is a deterministic constant k not depending on n , Δ , or W , such that

$$\lambda(n) = \begin{cases} \frac{kW}{(1+\Delta)^2 \sqrt{n \log(n)}}, & \text{Omni Tx Omni Rv,} \\ \frac{2\pi}{\gamma} \frac{kW}{(1+\Delta)^2 \sqrt{n \log(n)}}, & \text{Dir Tx Omni Rv,} \\ \frac{2\pi}{\beta} \frac{kW}{(1+\Delta)^2 \sqrt{n \log(n)}}, & \text{Omni Tx Dir Rv,} \\ \frac{4\pi^2}{\gamma\beta} \frac{kW}{(1+\Delta)^2 \sqrt{n \log(n)}}, & \text{Dir Tx Dir Rv,} \end{cases} \quad (3)$$

bits per second is feasible whp. Here we use the notations: *Omni* for omnidirectional, *Dir* for directional, *Tx* for transmission mode, and *Rv* for reception mode. Therefore, they found that constant gains in capacity are possible when compared to [8].

On the other hand, Peraki and Servetto [3] employed a network flow analysis to study the capacity of wireless networks implementing directional antennas and they showed that it is indeed possible to obtain gains of $\Theta(\log(n))$ compared to Gupta and Kumar's results [8]. Similar results are shown in Section IX-D using different analysis.

IV. MOBILE WIRELESS NETWORKS

The second type of wireless network is formed by mobile nodes. The movement of the nodes are function of their velocity and direction. There are many models for movements. In this chapter, we are going to consider the *uniform mobility model* [7] and the *random waypoint mobility model* [19], [20].

A. Mobility Models

In the *uniform mobility model* [7], the nodes are initially uniformly distributed, and move at a constant speed v and the directions of motion are independent and identically distributed (iid) with uniform distribution in the range $[0, 2\pi)$. As time passes, each node chooses a direction uniformly from $[0, 2\pi)$ and moves in that direction at speed v for a distance z , where z is an exponential random variable with mean μ . After reaching z the process repeats. This model satisfies the following properties [7]:

- At any time t , the position of the nodes are independent of each other.
- The steady-state distribution of the mobile nodes is uniform.
- Conditional on the position of a node, the direction of the node movement is uniformly distributed in $[0, 2\pi)$.

The uniform mobility is the theoretical model implemented for analysis of MANETs in this chapter.

An equivalent mobility pattern is obtained if the properties above are maintained, but the node moves in the random direction at speed v for a random chosen time that is exponentially distributed.

In the *random waypoint mobility model* [19], [20], the nodes are initially randomly distributed in the network area. A node begins its movement by remaining in a certain position for some fixed time, called *pause time*, distributed according to some random variable, and when it expires the node chooses a random destination point in the network area and begins to move toward that point with a constant speed uniformly distributed over $[v_{min}, v_{max}]$, where v_{min} and v_{max} stands for minimum and maximum velocity respectively. Upon arrival at the destination, the node pauses again according to the pause time random variable and the process repeats. Nodes move independently of each other. This is the model implemented in the simulations presented in this chapter for MANETs.

B. Multiuser Diversity

Knopp and Humblet [11] established a new communication scheme for wireless systems called *multiuser diversity* which was shown to maximize the capacity of a single-cell multiuser communication. Their basic idea is that the base station reserves the communication link for the user which presents the best channel. Because in a mobile wireless network the communication channel to each user is governed by independent fading, multiuser diversity provides a form of randomization of the channel access which one can take advantage of.

C. Mobility Increases the Capacity of Wireless Networks

In 2001, Grossglauser and Tse [1], [10] presented a two-phase packet relaying technique for mobile ad hoc networks that attains $\Theta(1)$ per source-destination throughput when n tends to infinity.

The scheme is based on *multiuser diversity* [11] where each source node transmits a packet to the nearest neighbor; that is, using the simple path propagation model, the source reserves its channel for a receiver that can best exploit it. This neighbor node functions as a relay and delivers the packet to the destination when this destination becomes the closest neighbor of the relay.

The network model consists of a normalized unit area disk containing n mobile nodes. They considered a time-slotted operation to simplify the analysis. The position of node i at time t is indicated by $X_i(t)$. The process $\{X_i(\cdot)\}$ is stationary and ergodic with stationary uniform distribution on the disk, which yields node trajectories that are iid.

At each time step, a scheduler decides which nodes are senders, relays, or destinations, in such a manner that the source-destination association does not change with time. Accordingly, A fraction of the total number of nodes n in the network, n_S , is chosen randomly by the scheduler as senders, while the remaining nodes, n_R , operate as possible receiving nodes [1]. A sender density parameter θ is defined as $n_S = \theta n$, where $\theta \in (0,1)$, and $n_R = (1 - \theta)n$. Each node can be a source for one session and a destination for another session. Packets are assumed to have header information for scheduling and identification purposes.

Suppose that a source i has data for a certain destination $d(i)$ at time t . Because nodes i and $d(i)$ can have direct communication only $1/n$ of the time on the average, a relay strategy is proposed to deliver data to $d(i)$ via relay nodes. They assume that each packet can be relayed at most once.

Therefore, according to the above communication scheme, each node sends data to its destination in a two phase process (see Fig. 2). Packet transmissions from sources to relays (or destinations) occur during *Phase 1*, and packet transmissions from relays (or sources) to destinations happen during *Phase 2*. Both phases occur concurrently, but *Phase 2* has absolute priority in all scheduled sender-receiver pairs.

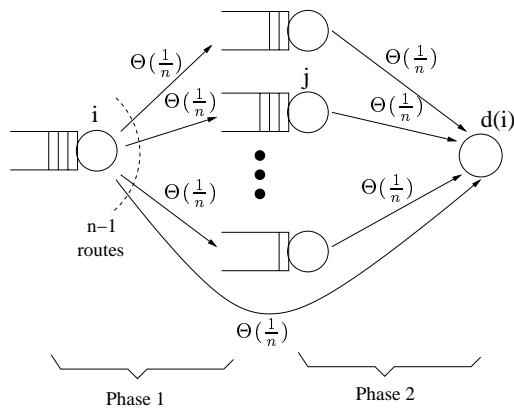


Fig. 2. Two-phase process for packet delivery.

Because node trajectories are iid and the system is in steady-state, the long-term throughput between any two nodes equals the probability that these two nodes are selected by the scheduler as a feasible sender-receiver pair. According to [1] this probability is $\Theta(\frac{1}{n})$. Also, there is one direct route and $n - 2$ two-hop routes passing through one relay node for a randomly chosen source-destination pair. Thus, the service rate is $\lambda_j = \Theta(\frac{1}{n})$ through each actual relay node, as well as the direct route. Consequently, the total throughput per source-destination pair λ_T is

$$\lambda_T = \sum_{j=1, j \neq i}^n \lambda_j = \sum_{j=1, j \neq i}^n \Theta\left(\frac{1}{n}\right) = \Theta\left(\frac{n-1}{n}\right) \xrightarrow{n \rightarrow \infty} \Theta(1). \quad (4)$$

Thus, this scheme attains $\Theta(1)$ per source-destination throughput when n tends to infinity. However, the delay experienced by packets under this strategy was shown to be large and it can be even infinite for a fixed number of nodes (n) in the system, which has prompted more recent work presenting analysis of capacity and delay tradeoffs [6], [7], [21], [9], [22]. Given that the number of nodes in real MANETs is finite, delay is an important performance issue.

D. One-Dimensional Mobility is Enough

In [23], Diggavi et al proved that by having nodes executing more restricted movements it is still possible to recover the Grossglauser and Tse's results [1]. They assume that mobile nodes have iid movements over segments of lines on a sphere, or more specifically, each node is constrained to move on a single-dimensional great circle. It is shown that if the locations of the great circles are chosen randomly and

independently, then the per source destination pair throughput can remain constant as the number of nodes increases, for almost all configurations of such great circles.

V. RANDOM LOAD BALANCING

The works by Azar et al [24] and Mitzenmacher [25] developed a powerful idea in the field of computer performance. They showed that a small amount of choice can lead to drastically different results in random load balancing. More specifically, if n balls are randomly placed into n bins, the size of the maximum bin is given approximately by $\frac{\log(n)}{\log(\log(n))}$ whp. However, if each ball is placed sequentially into the least full of K bins chosen uniformly and independently at random, the maximum load is then only $\frac{\log(\log(n))}{\log(K)} + O(1)$ whp, which represents an exponential reduction in maximum load. This principle seems to be related to the delay reduction in MANETs by using multiple copies of same packet following distinct random routes to the destination described in Section VIII.

VI. TRADE-OFF RESULTS

In wireless ad hoc networks, the capacity-delay trade-off is a key issue. The assumptions for attaining a given capacity is always associated with the delivery delay of packets. Perevalov and Blum [6] presented an analysis for delay limited capacity of ad hoc networks in which they computed the ensemble average of the probability that two nodes come close to each other. Their approach can be used to obtain delay for the multi-copy forwarding scheme presented in this chapter. Neely and Modiano [21] provided an interesting study on capacity and delay trade-off for mobile ad hoc networks using the theory of queues. By assuming that the network has a cell partitioned structure, the nodes move according to an iid model, and using a modified two-hop relaying strategy, they found that the throughput and delay must satisfy $delay/rate \geq O(n)$. El Gamal et al [9] provided a detailed trade-off analysis for throughput and delay in wireless networks, where they recover the results (as outlined below) by Gupta and Kumar [8], as well as Grossglauser and Tse [1], using simpler techniques.

A. Outline of Results

1) *Static random wireless network*: The relationship between delivery delay $D(n)$ and node throughput $\lambda(n)$ is given by [9]

$$D(n) = \Theta(n\lambda(n)), \text{ for } \lambda(n) = O(1/\sqrt{n \log(n)}). \quad (5)$$

From this result, the following implications are implied. (i) The maximum node throughput in a static wireless network is $O(1/\sqrt{n \log(n)})$, as shown by Gupta and Kumar [8]. The average delivery delay associated to this throughput is $\Theta(n\lambda(n))$ as illustrated by the point M in Fig. 3. (ii) The average number of traversed hops from source to destination can be reduced by increasing the transmission radius, which increases the interference. In such case, the throughput is lower. Segment VM in Fig. 3 shows the dependence between $D(n)$ and $\lambda(n)$ for Eq. (5).

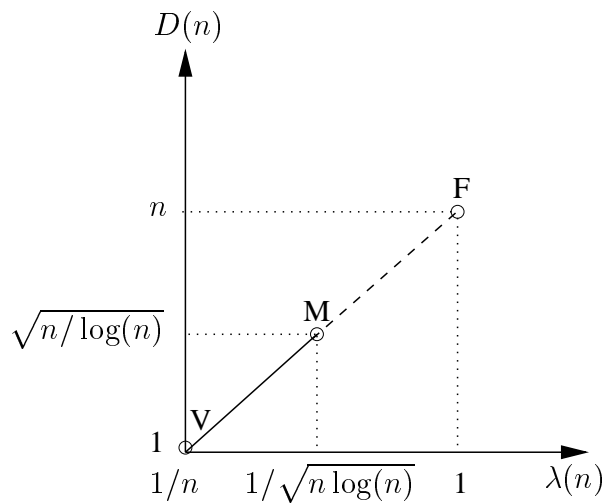


Fig. 3. Throughput delay trade-off for a wireless network assuming $v(n) = \Theta(1/\sqrt{n})$. The values shown on the axes represent the orders asymptotically in n .

2) *Delay in mobile wireless network for $\lambda(n) = \Theta(1)$* : In a mobile network where nodes move with speed $v(n)$ according to independent Brownian motions and employ a single relay as in Grossglauser and Tse [1], El Gamal et al [9] showed that

$$D(n) = O\left(\frac{\sqrt{n}}{v(n)}\right) \text{ when } \lambda(n) = \Theta(1), \quad (6)$$

using results from random walks and queuing theory. Making $v(n) = \Theta(1/\sqrt{n})$, the point F in Fig. 3 is obtained.

3) *Throughput delay trade-off for a mobile network*: Let $\sqrt{a(n)}$ represents the average distance traveled by a mobile node. $a(n)$ ranges from $\Theta(\log(n)/n)$ (which corresponds to grossglauser and Tse model [1], point F in Fig. 3) to $\Theta(1)$ (corresponding to Gupta and Kumar model [8], point M). For the range between $\Theta(1/\sqrt{n \log(n)})$ and $\Theta(1/\log(n))$, the trade-off is given by [9]

$$\lambda(n) = \Theta\left(1/\sqrt{na(n) \log(n)}\right), \text{ and } D(n) = \Theta\left(1/v(n)\sqrt{a(n)}\right). \quad (7)$$

This relationship is illustrated by the segment MF in Fig. 3.

VII. HYBRID NETWORKS

Another possible scenario arises when there are nodes and base stations in a wireless network [26]. Liu et al [4] study the capacity scalability for these types of networks. They considered that some static nodes are represented as base stations which are interconnected by a broadband wired infrastructure. These base stations can forward the messages from the user nodes across the network. They showed that the number of base stations should grow faster than \sqrt{n} , where n is the number of mobile nodes, in order to obtain a capacity that scale linearly with the number of base stations. Otherwise, the benefit of adding base stations on capacity is insignificant. In another hybrid scenario, Agarwal and Kumar [27] emphasized that by choosing appropriate transmission power levels and communicating with the closest node or base station in a multihop fashion can provide good capacity performance for static ad hoc wireless network.

VIII. THROUGHPUT-DELAY ANALYSIS OF MANETs WITH MULTI-COPY FORWARDING

This section summarizes an improved two-phase packet forwarding strategy for MANETs [28] that attains the $\Theta(1)$ capacity of the basic scheme by Grossglauser and Tse [1] (see Section IV), but provides better delay behavior than the single-copy technique. The main objective is to decrease the delay incurred by the packet to reach its destination in steady-state⁴ while maintaining the capacity of the network at the

⁴That is, after averaging over all possible starting random network topologies so that transient behaviors are removed.

same order of magnitude as that attained in [1]. The basic idea is to give a copy of the packet to multiple one-time relay nodes that are within the transmission range of the sender. By doing so, the time within which a copy of the packet reaches its destination can be decreased. The first one-time relay node that is close enough to the destination delivers the packet.

We find an enormous reduction in delay by having a packet more than one possible random route to the destination. This result is analogous to the problem in which Azar et al [24] and Mitzenmacher [25] showed that a small amount of choices can lead to drastically different result in randomized load balancing. More specifically, having just two random choices yields an exponential reduction in maximum loading over having one choice, while each additional choice beyond two decreases the maximum load by just a constant factor. In this case, by multi-copy forwarding a packet and having only the fastest copy being delivered, it is analogous to having the packet taking one of the shortest random path to the destination from multiple random routes (or choices).

In the presented multi-copy relay strategy, each sender transmits to all the nodes in the feasible transmission range, these additional copies follow different random routes and find the destination earlier compared to the single-copy strategy of Grossglauser and Tse [1].

An interesting feature of the multi-copy relaying approach is that the additional relaying nodes carrying that same copy of the packet can be used as backups to protect against node failures which improves the reliability of the network [29].

A. *Basic Assumptions*

The network model is similar to the one introduced by Grossglauser and Tse [1]. Nodes are assumed to move according to the *uniform mobility model* [7], in which the steady-state distribution of the mobile nodes in the network is uniform.

B. Multi-Copy One-time Relaying

As discussed in Section IV, Grossglauser and Tse [1] consider a single-copy forwarding scheme consisting of two phases. Packet transmissions from sources to relays (or destinations) occur during *Phase 1*, and packet transmissions from relays (or sources) to destinations happen during *Phase 2*. Both phases occur concurrently, but *Phase 2* has absolute priority in all scheduled sender-receiver pairs. This scheme allows multiple copies of the same packet during *Phase 1*, as described below.

1) *Packet Forwarding Scheme*: Phase 1 allows more than one relay node to receive a copy of the same packet. Thus, the chance that a copy of this packet reaches its destination in a shorter time is increased, compared to using only one relay node as in [1]. Also, if for some reason a relaying node fails to deliver the packet when it is within the transmission range of the destination, the packet can be delivered when another relaying node carrying a copy of the same packet approaches the destination.

In Fig. 4, three copies of the same packet are received by adjacent relay nodes j , p , and k during *Phase 1*. All such relays are located within a distance r_o from sender i . At a future time t , in *Phase 2*, node j reaches the destination before the other relays and delivers the packet. Note that relay node j need not be the closest node to the source during *Phase 1*.

If the density of nodes in the disk is $\rho = \frac{n}{\text{total area}} = \frac{n}{1} = n$, then, for a uniform distribution of nodes, the radius for one sender node is given by

$$1 = \theta \rho \pi r_o^2 = \theta n \pi r_o^2 \implies r_o = \frac{1}{\sqrt{\theta n \pi}}. \quad (8)$$

Thus, the radius r_o defines a cell (radius range) around a sender.

2) *Enforcing One-Copy Delivery*: There are several ways in which the delivery of more than one copy of the same packet to a destination can be prevented. For example, each packet can be assigned a sequence number (SN) [21]. Before a packet is delivered to its destination, a relay-destination handshake can be established to verify that the destination has not received a copy of the same packet and to inform the relay to delete a packet copy that has already been delivered.

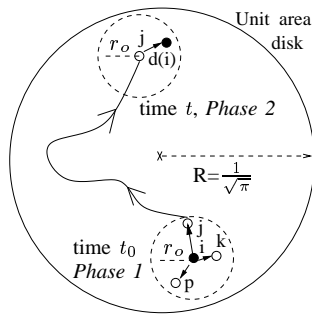


Fig. 4. Three copies of the same packet are transmitted to different receivers at *Phase 1*. Node j is the first to find the destination, and delivers the packet at *Phase 2*. The other copies are deleted later after p and k nodes handshake with destination.

In Fig. 4, node j finds the destination node $d(i)$ first and delivers the packet. The other copies are dropped from the queues at p and k at a later time after handshaking with destination, and only one node out of the three potential relays actually delivers the packet to the destination.

Because we address the network capacity for any embodiment of the multi-copy relaying strategy, we assume for the rest of this chapter that the overhead of the relay-destination handshake is negligible.

C. Per Source-Destination Throughput

The throughput per source-destination pair with multi-copy relaying approach remains the same order of magnitude as the original single-copy relaying scheme, which is $\Theta(1)$ [1].

Referring to the recent work by El Gamal et al [9], each cell in this strategy has area $a(n) = \frac{1}{n_S} = \frac{1}{\theta n}$. By applying random occupancy theory [12, Chapter 3], the fraction of cells containing L senders and K receivers is obtained by

$$\begin{aligned}
 \mathbb{P}\{\text{senders} = L, \text{receivers} = K\} &= \mathbb{P}\{\text{senders} = L\} \mathbb{P}\{\text{receivers} = K \mid \text{senders} = L\} \\
 &= \binom{n}{L} \left(\frac{1}{n_S}\right)^L \left(1 - \frac{1}{n_S}\right)^{n-L} \binom{n-L}{K} \left(\frac{1}{n_S}\right)^K \left(1 - \frac{1}{n_S}\right)^{n-L-K} \\
 &\approx \frac{1}{L!} \left(\frac{1}{\theta}\right)^L e^{-1/\theta} \frac{1}{K!} \left(\frac{1}{\theta}\right)^K e^{-1/\theta} \quad (\text{for large } n). \tag{9}
 \end{aligned}$$

For example, for $L = 1$, $K \geq 2$, and $\theta = \frac{1}{3}$, the fraction of the cells containing one sender and at least two receivers equals $\frac{1}{\theta} e^{-1/\theta} (1 - e^{-1/\theta} - \frac{1}{\theta} e^{-1/\theta}) \approx 0.12$. Therefore, for $K \geq 2$, approximately 12% of the

cells can forward packets using the multi-copy scheme in *Phase 1*.

In addition, for $\theta = \frac{1}{3}$, the fraction of the cells with one sender and one receiver equals $(\frac{1}{\theta}e^{-1/\theta})^2 \approx 0.02$. In this case, the scheduler does not select these cells for packet transmission, because the delivery delay incurred can be significantly high, as will be shown subsequently.

In the case of multi-copy forwarding, only one copy is delivered to the destination and the other copies are dropped from the additional relaying nodes because the handshake between relay and destination informs the relay that the packet has been delivered. Therefore, only one node out of K nodes actually functions as a relay. Accordingly, only one copy passes successfully through the two-phase process, as in Fig. 2. Because node trajectories are iid and the system is in steady-state, the long-term throughput between any two nodes equals the probability that these two nodes are selected by the scheduler as a feasible sender-receiver pair, which has been shown to be $\Theta(\frac{1}{n})$ [1]. Also, there is one direct route and $n - 2$ two-hop routes passing through one relay node for a randomly chosen source-destination pair. Thus, the service rate is $\lambda_j = \Theta(\frac{1}{n})$ through each relay node, as well as the direct route. Accordingly, from Eq. (4), the total throughput per source-destination pair λ_T is $\Theta(1)$.

From the description given above and from Eq. (9), the exact total throughput per source-destination pair is given by the fraction of cells that successfully forward packets (i.e, the cells that are selected by the scheduler containing feasible sender-receiver pairs). Then, for one sender and at least K receivers per cell, we have

$$\lambda_T = \mathbb{P}\{\text{senders } (L) = 1, \text{ receivers are at least } K\} \approx \frac{1}{\theta} e^{-1/\theta} \left(1 - \sum_{k=0}^{K-1} \frac{1}{k!} \left(\frac{1}{\theta}\right)^k e^{-1/\theta} \right). \quad (10)$$

Hence, for at least two receivers per cell and $\theta = \frac{1}{3}$, $\lambda_T = \frac{1}{\theta}e^{-1/\theta}(1 - e^{-1/\theta} - \frac{1}{\theta}e^{-1/\theta}) \approx 0.12 = \Theta(1)$. Therefore, the multi-copy forwarding strategy attains the same throughput order as in [1].

Also, for at least one receiver per cell and $\theta = \frac{1}{3}$, $\lambda_T = \frac{1}{\theta}e^{-1/\theta}(1 - e^{-1/\theta}) \approx 0.14$. Hence, for the case in which $K \geq 1$, Eqs. (9) and (10) give the same throughput value obtained by Tse and Grossglauser [1], as well as Neely and Modiano [21]. The single-copy forwarding strategy [1] selects only the nearest neighbor amongst the K potential receiver nodes.

D. Delay Equations

This section describes the relationship between the delay value d for the case of only one copy relaying [1], and the new delay d_K for $K \geq 2$ copies transmitted during *Phase I* in steady-state behavior. Obviously, we have $d_K \leq d$. A naive guess would be to take $d_K = \frac{d}{K}$. However, the correct answer is obtained considering the random movement of the nodes. K is a small integer, much smaller than n .

1) *Single-Copy Forwarding Case*: Because we have iid node trajectories, we focus on a given relay node labeled as *node 1*, and without loss of generality assume that *node 1* received a packet from the source during time $t_0 = 0$. Let $\mathbb{P}\{|X_1(s) - X_{dest}(s)| \leq r_o \mid s\}$ denote the probability that relay *node 1* at position $X_1(s)$ is close enough to the destination node *dest*, given that the time interval length is s , where r_o is the radius distance given by Eq. (8), so that successful delivery is possible. The time interval length s is the delivery-delay random variable. Perevalov and Blum [6] obtained an approximation for the ensemble average with respect to all possible uniformly-distributed starting points, $(X_1(0), X_{dest}(0))$, where they considered the nodes moving on a sphere. Their result can be extended to nodes moving in a circle by projecting the sphere surface movement in the sphere equator and thus have trajectories described in a circle and have

$$E_U[\mathbb{P}\{|X_1(s) - X_{dest}(s)| \leq r_o \mid s\}] = 1 - e^{-\lambda s} \left(1 - \lambda e^{-\lambda \int_0^s h_{X'}(t) dt} \int_0^s e^{\lambda \int_0^t h_{X'}(u) du} h_{X'}(t) dt \right) = \mathbb{P}\{S \leq s\} = F_S(s), \quad (11)$$

where $E_U[\cdot]$ means the ensemble average over all possible starting points that are uniformly distributed on the disk. $F_S(s)$ can be interpreted as the cumulative density function of the delay random variable S . The function $h_X(t)$ is the difference from the uniform distribution, such that $h_X(0) = 0$ and $|h_X(t)| < 1$ for all t , and X' is a point at distance r_o from the destination. The parameter λ is related to the mobility of the nodes in the disk and can be expressed by [6]

$$\lambda = \frac{2 r_o v}{\pi R^2} = \frac{2 r_o v}{1} = 2 r_o v, \quad (12)$$

which results from evaluating the flux of nodes entering a circle of radius r_o during a differential time interval, considering the nodes uniformly distributed over the entire disk of unit area and traveling at speed v . From Eq. (8), we see that the radius r_o decreases with $\frac{1}{\sqrt{n}}$. To model a real network in which a node would occupy a constant area, if the network grows, the entire area must grow accordingly. Therefore,

because in the present analysis the total area is maintained fixed, the speed of the nodes must be scaled down [9]. Consequently, the velocity of the nodes also must decrease with $\frac{1}{\sqrt{n}}$. Then

$$\lambda = \frac{1}{\Theta(n)}. \quad (13)$$

Now, $h_X(t)$ has to be taken according to the random motion of the nodes. If we consider the *uniform mobility model* [7], then a steady-state uniform distribution results as the random motion of the nodes in the disk. In such a case, $h_X(t) = 0 \forall t \geq 0$. Applying this result in Eq. (11) we have

$$E_U [\mathbb{P}\{|X_1(s) - X_{dest}(s)| \leq r_o \mid s\}] = 1 - e^{-\lambda s} = \mathbb{P}\{S \leq s\} = F_S(s). \quad (14)$$

Thus, the delay behaves exponentially with mean $\frac{1}{\lambda}$ and variance $\frac{1}{\lambda^2}$.

From now on, we replace s by d to indicate the delay for single-copy forwarding at *Phase I*.

2) *Multi-Copy Forwarding Case*: Now consider that K copies of the same packet were successfully received by adjacent relaying nodes during *Phase I* (where $1 < K \ll n$). Let $\mathbb{P}_D(s)$ be the probability of having the first (and only) delivery of the packet at time interval length s . Hence, given that only one-copy delivery is enforced (see Section VIII-B.2), and all K relays are looking for the destination, we have that

$$\mathbb{P}_D(s) = \mathbb{P}\left\{\bigcup_{i=1}^K [|X_i(s) - X_{dest}(s)| \leq r_o \mid s]\right\} \implies \mathbb{P}_D(s) \leq \min\left[\sum_{i=1}^K \mathbb{P}\{|X_i(s) - X_{dest}(s)| \leq r_o \mid s\}, 1\right], \quad (15)$$

where the union bound was used considering that $\mathbb{P}_D(s)$ can be at most equal to 1, and Eq. (14) holds for each individual relay i because all the K nodes have independent and identically distributed movements and one can use the results in [6] for a single relay. However, when one attempts to compute the probabilities of multiple relays, since all these nodes start moving from the same area to search for destination (within a circle of radius r_o), their probability distributions are not mutually exclusive. If the time necessary for all these nodes to uniformly spread in the disk is equal to t_{spread} , since each node has a speed $v = \Theta(\frac{1}{\sqrt{n}})$, then in general, $t_{spread} = \Theta(\sqrt{n})$. As we will show later, the maximum delay $d_K^{max} = \Theta(n)$ given that $K \ll n$ whp. Therefore, $t_{spread} \ll d_K^{max}$ for large values of n , and consequently one can approximate

all K probabilities using Eq. (14). This approximation for Eq. (15) results in

$$\mathbb{P}_D(s) \leq \min [K \cdot \mathbb{P}\{|X_1(s) - X_{dest}(s)| \leq r_o \mid s\}, 1]. \quad (16)$$

Equation (16) describes two cases. The first case is when $\mathbb{P}_D(s)$ is less than 1 while the second case is when the union bound is greater than 1. Obviously, we can derive a meaningful relationship between d_K and d only for the first case and that is the basis of the remaining analysis. From Eqs. (14) and (16), and replacing s by d_K ⁵ to indicate the delay for K -copies forwarded during *Phase I*, we have for the *uniform mobility model*,

$$E_U[\mathbb{P}_D(s)] = E_U \left[\mathbb{P} \left\{ \bigcup_{i=1}^K [|X_i(s) - X_{dest}(s)| \leq r_o \mid s = d_K] \right\} \right] = \mathbb{P}\{D_K \leq d_K\} = F_{D_K}(d_K) \approx K(1 - e^{-\lambda d_K}), \quad (17)$$

for a uniform steady-state distribution resulting from the random motion of the nodes. Exact computation of probability of d_K is a tedious task. Computer simulation for MANETs demonstrates that this approximation can show the asymptotic behavior of d_K reasonably well. $F_{D_K}(d_K)$ can be interpreted as the cumulative density function of the delay random variable D_K for K copies transmission at *Phase I*.

3) Relationship between Delays: The throughput of the multi-copy scheme is of the same order as the one-copy scheme [1]. Indeed, we showed that $\lambda_T \approx 0.14$ for single-copy and $\lambda_T \approx 0.12$ for multi-copy ($K > 1$), for $\theta = \frac{1}{3}$. This capacity is proportional to the probability of a packet reaching the destination. Hence, because only one copy of the packet is actually delivered to the destination for single-copy or multi-copy, their total probabilities can be approximated at their respective delivery time, and so their ensemble averages are

$$E_U \left[\mathbb{P} \left\{ \bigcup_{i=1}^K [|X_i(s) - X_{dest}(s)| \leq r_o \mid s = d_K] \right\} \right] \approx E_U [\mathbb{P}\{|X_1(s) - X_{dest}(s)| \leq r_o \mid s = d\}], \quad (18)$$

whose solution must be obtained by substituting Eq. (11) (for $s = d_K$ and $s = d$ respectively) on both sides of Eq. (18) and solving for d_K for the particular model of random motion of nodes. For a steady-state

⁵To be more accurate, we should use \tilde{d}_K instead of d_K for the rest of the chapter because of the approximation. In order to make the paper easy to read, we will continue to use the same notation for simplicity.

TABLE I

AVERAGE DELAY AND VARIANCE FOR SINGLE-COPY [1] AND MULTI-COPY ($1 < K \ll n$) TRANSMISSION OBTAINED FROM EQS. (14),

AND (17), AND RESPECTIVE ASYMPTOTIC DELAY VALUES d_K^{max} FROM EQ. (20), FOR FINITE n .

Copies	Mean	Variance	d_K^{max}
Single-copy	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	∞
$K = 2$	$0.307 \frac{1}{\lambda}$	$0.039 \frac{1}{\lambda^2}$	$\frac{\log(2)}{\lambda}$
$K = 3$	$0.189 \frac{1}{\lambda}$	$0.014 \frac{1}{\lambda^2}$	$\frac{\log(3/2)}{\lambda}$
$K = 4$	$0.137 \frac{1}{\lambda}$	$0.007 \frac{1}{\lambda^2}$	$\frac{\log(4/3)}{\lambda}$

uniform distribution for the motion of the nodes, a simplified solution is obtained by substituting Eqs. (14) and (17) in Eq. (18), i.e.,

$$K(1 - e^{-\lambda d_K}) \approx 1 - e^{-\lambda d} \implies d_K \approx \frac{1}{\lambda} \log \left(\frac{K}{K - 1 + e^{-\lambda d}} \right). \quad (19)$$

The last column of Table I shows statistics of this asymptotic delay for the single-copy and multi-copy ($2 \leq K \leq 4$) cases, expressed as a function of the mobility parameter λ , for a finite number of nodes n .

This last equation reveals some properties for the strategy of transmitting multiple copies of a packet during *Phase 1*. If $K = 1$, then obviously $d_K = d$. If $d \rightarrow \infty$, with n finite, and because $K \ll n$, then we have

$$d_K^{max} \approx \lim_{d \rightarrow \infty} \frac{1}{\lambda} \log \left(\frac{K}{K - 1 + e^{-\lambda d}} \right) = \frac{1}{\lambda} \log \left(\frac{K}{K - 1} \right) \stackrel{\text{if } K \geq 1}{=} c_6. \quad (20)$$

Fig. 5 shows curves for Eq. (19), where $\lambda = 0.0052$. The case of single-copy ($K = 1$) is also plotted. Except for the case of single-copy, the delay d_K tends to a constant value as d increases. *Hence, for a finite n , the multi-copy relay scheme can reduce a delay of hours of the single-copy relay scheme to a few minutes or even a few seconds, depending on the network parameter values.* Fig. 5 also shows that, a packet having two random routes to reach its destination produces an exponential reduction in delay compared to having only one random route, whereas having three (or more) random routes leads only to

incremental improvements from having two random routes. This result is analogous to the power of two random choices studied by Azar et al [24] and Mitzenmacher [25].

4) *Simulation Results:* Equation (16) and all the results derived from it simply approximate the behavior of the delay d_K . In order to demonstrate if this approximation and the associated results are justified, a MANET with 1000 nodes is used for simulation. The simulations compare the behavior of the multi-copy packet forwarding strategy with the single-relay strategy. The *BonnMotion* simulator [30] is used, which creates mobility scenarios that can be used to study MANETs characteristics.

For simulations, the simplified version of the *random waypoint mobility model*, where no pause was used and $v_{min} = v_{max} = v$. Fig. 5 shows the averaged pairs of points (d, d_K) obtained for $K = 2$ and $K = 4$ for 1000 seconds of simulations for $n = 1000$ nodes, $v = 0.13$ m/s, $r_o = 0.02$ m, and a unit area disk as the simulation area, which results $\lambda = 0.0052$. To obtain a solution close to the steady-state behavior, 40 random topologies are created and averaged them as follows. In each run, a node is randomly chosen with $K = 2$ and $K = 4$ neighbors within r_o , respectively, and measured the time that each of these K nodes reached each of the other $n - K$ nodes in the disk (i.e., except the sender and its other $K - 1$ neighbors) considering each of them as a destination. The delay of the sender's nearest node reaching each destination is d by definition, and d_K is the minimum time among all the K nodes that reach the destination. It is observed that the averaged random waypoint mobility curves follow the steady-state uniform distribution predicted by the theory. Note that they do not match exactly each other, because they are slightly different mobility models and the delay analysis for multi-copy forwarding scheme is only an approximation.

IX. MOBILITY-CAPACITY-DELAY TRADE-OFF IN WIRELESS AD HOC NETWORKS

In this section, we present new network models to show that mobility can also be varied as a resource together with capacity and delay [31], [32]. The idea is to allow the nodes to execute *restricted* movements, i.e., each node moves only inside some given area in the network. By allowing transmissions to closest neighbor nodes only, each node overcomes interference from other transmitting nodes. Given that nodes

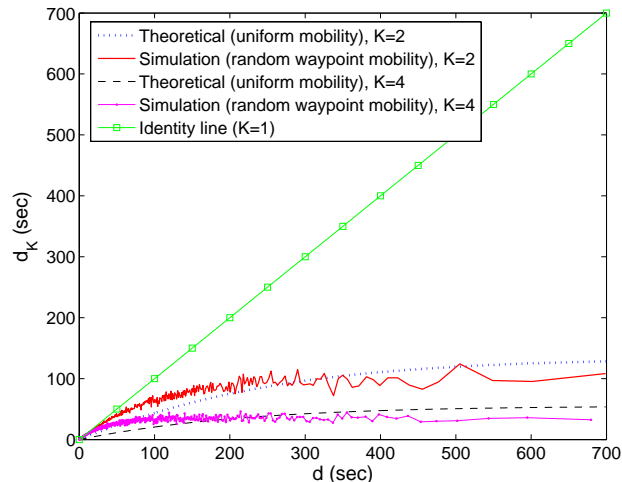


Fig. 5. Theoretical and simulation results for the *uniform mobility model* and *random waypoint mobility model*, respectively, for $K = 2$ and $K = 4$, with $\lambda = 0.0052$. For simulations, the delay measured is averaged over 40 random topologies.

have restrained mobility, the delivery from source to destination is done across multiple hops obtained by relaying packets along the path linking the source to the destination. Diggavi et al [23] considered one-dimensional mobility model in which nodes were allowed to execute movements on circles on a sphere. They showed that a constant throughput is still feasible; however, they did not present the corresponding trade-offs associated with mobility, capacity and delay.

Note that restrained mobility patterns have potential practical applications in cases in which nodes are not allowed to leave a given region like a room, a hallway, or a predefined region covered by a sensor network, and has to rely on multiple hops (i.e., relays) to send a packet to farther destinations. Restrained mobility also provides insights on the performance of a network covering geo areas that are too vast for a single node to cover. Therefore, restricted mobility models are important to the study of ad hoc and sensor networks.

A. Basic Assumptions

The model considered here is that of a wireless ad hoc network with nodes assumed either fixed or mobile. The network consists of a normalized unit area torus containing n nodes [8], [10], [9].

For the case of *fixed nodes*, the position of node i is given by X_i . A node i is capable of transmitting at

a given transmission rate of W bits/sec to j if the protocol model condition for successful transmission is satisfied as given in Eq. (1).

For the case of *mobile nodes*, the network model follows the assumptions given in Section IV. A successful transmission is governed again by the protocol model, where the positions of the nodes are time dependent.

B. Scheme 1

The first restricted mobility scheme attains a capacity gain of $\Theta(\log(n))$ compared to the static network model [8]. The throughput still decreases as the number of nodes n in the network grows to infinity. However, it models cases in which nodes can move around only a fraction of the geo region covered by the network, and serves as a building block for the scheme presented in the next section, which attains non-zero asymptotic throughput capacity in a dense network.

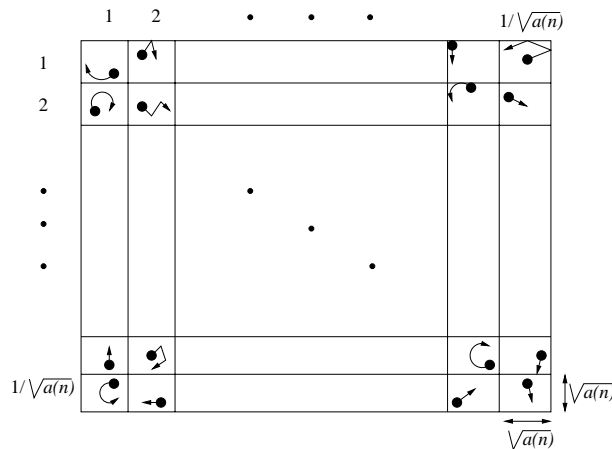


Fig. 6. Unit area torus network divided into $1/a(n)$ cells, each with size of $a(n)$.

The node limited movement model is illustrated in Fig. 6. The network is a unit torus divided into square cells, each of area $a(n)$ as in [9], in which they showed that, if $a(n) \geq \frac{2 \log(n)}{n}$, then each cell has at least one node whp. This condition guarantees connectivity whp [8], [9].

Let's consider the additional assumption that each node has its movement confined to only one cell. This means that a node cannot cross the cell edge and percolate to a neighbor cell. By doing so, each

cell is composed by at least one node whp, and such a node moves with speed $v(n)$, and no preferential direction of movement within the cell. Nodes move independently of each other, and once they hit the cell boundaries they are bounced back (with relation to the edge normal).

Each node only communicates with another node from an adjacent cell, and this happens only when the nodes are close enough to each other (i.e., both are near to the common edge that separates the cells) so that the effect of interference can be minimized. Thus, a source node relies on relays across several cells to have a packet delivered to a destination. Each packet travels via multiple relays from source to destination following the path close to the straight line linking source and destination. Each source-destination pair is chosen uniformly and independently from different cells. Fig. 7 illustrates a packet whose source and destination nodes are in cells i and d respectively, separated by an average distance \bar{L} . Possible cell paths for this packet are $\{i \rightarrow j \rightarrow f \rightarrow g \rightarrow c \rightarrow d\}$, $\{i \rightarrow j \rightarrow f \rightarrow g \rightarrow h \rightarrow d\}$, $\{i \rightarrow e \rightarrow f \rightarrow g \rightarrow c \rightarrow d\}$, $\{i \rightarrow e \rightarrow f \rightarrow g \rightarrow h \rightarrow d\}$, for example.

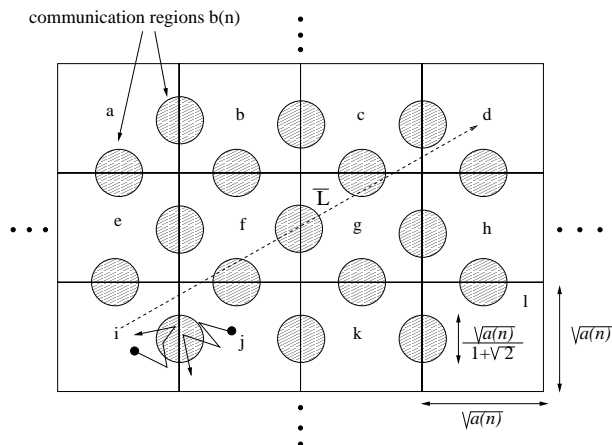


Fig. 7. Region $b(n)$ where communication between nodes from adjacent cells is possible.

Grossglauser and Tse [10] showed that transmission to the nearest node is possible, even when the number of interferers in the network scale to infinity. This allows a node to schedule transmission to a neighbor node from an adjacent cell when Eq. (1) is satisfied. In addition, we assume that both nodes are so close that communication is successful during the entire time slot (or session). The transmission is half-duplex so that each node uses half of the communication time slot to transmit at a rate of W

bits/sec, and the other half to receive at the same rate. Thus, the average available bit rate is $\frac{W}{2}$ *bits/sec*. Each time two nodes communicate with each other, they exchange packets, and these exchanges can be source-relay, relay-relay, or relay-destination transmissions.

The area in which successful communication can occur is shown in Fig. 7. Basically, it is a semi-circumference $b(n)$ of radius $\frac{\sqrt{a(n)}}{2+2\sqrt{2}}$ where two nodes from adjacent cells can come close to each other so that Eq. (1) is satisfied, i.e., no other node from the other cells will be closer to them than themselves. For the case in which more than one node in the same cell are simultaneously traveling inside $b(n)$, only one of these nodes is allowed to communicate with a node from the adjacent cell. Accordingly, from Fig. 7, the two adjacent nodes in cells i and j are able to communicate during the time they simultaneously travel inside their respective regions $b(n)$'s in their cells as shown. We have that

$$b(n) = \frac{1}{2} \pi \left(\frac{\sqrt{a(n)}}{2 + 2\sqrt{2}} \right)^2 = \frac{\pi a(n)}{24 + 16\sqrt{2}}. \quad (21)$$

The probability of finding a node traveling inside $b(n)$ is $\frac{b(n)}{a(n)}$, because the node has no preferential direction of movement in the cell and tends to move uniformly inside the cell. In addition, because the nodes have iid movements, the probability that both nodes come to the communication region simultaneously, denoted by \mathbb{P}_{comm} , equals

$$\mathbb{P}_{comm} = \left[\frac{b(n)}{a(n)} \right]^2 = \left(\frac{\pi}{24 + 16\sqrt{2}} \right)^2 = c_7. \quad (22)$$

Hence, \mathbb{P}_{comm} does not depend on n .

Because \bar{L} is the mean distance between two uniformly and independently chosen source-destination nodes in the network, the average path distance across cells traversed by a packet from source to destination is $\Theta(\bar{L})$. Accordingly, each cell hop has an average size of $\sqrt{a(n)}$. Thus, the mean number of hops traversed by a packet is $\frac{\Theta(\bar{L})}{\sqrt{a(n)}}$.

According to the definition of throughput, each source generates $\Lambda(n)$ bits per second⁶ and there are n sources in the network. Also, each bit needs to be relayed by $\frac{\Theta(\bar{L})}{\sqrt{a(n)}}$ nodes on the average. Thus, the

⁶Here we use the notation $\Lambda(n)$ (instead of $\lambda(n)$ given in Section II) to indicate that this throughput is related to restrained mobility models, or nodes using directional antennas.

total average number of bits per second served by the entire network equals $\frac{\Theta(\bar{L})n\Lambda(n)}{\sqrt{a(n)}}$. To ensure that all required traffic is carried, we need that

$$c_9 n \frac{W}{2} \mathbb{P}_{comm} \leq \frac{\Theta(\bar{L})n\Lambda(n)}{\sqrt{a(n)}} \leq c_{10} n \frac{W}{2} \mathbb{P}_{comm} \implies c_{11} \sqrt{a(n)} \leq \Lambda(n) \leq c_{12} \sqrt{a(n)}. \quad (23)$$

We have just proved the following Theorem.

Theorem 2 *For Scheme 1 with $a(n) = \frac{k \log(n)}{n}$ and $k \geq 2$, to guarantee connectivity, we have*

$$\Lambda_1(n) = \Theta \left(\sqrt{\frac{\log(n)}{n}} \right).$$

Compared to the capacity result obtained by Gupta and Kumar [8] which is $\Theta(1/\sqrt{n \log(n)})$, the result of Theorem 2 represents a gain of $\Theta(\log(n))$. *Thus, a gain in throughput over the static network model is obtained by allowing the nodes to execute a restricted mobility pattern.*

The average delay incurred by a packet to reach the destination in *Scheme 1* is the sum of the average time a packet spends in each hopping cell in the path to its destination. A node travels around the cell boundary on average every $t(n)$ time-slots that is proportional to

$$t(n) \propto \frac{\Delta S \cdot \mathbb{P}_{comm}}{v(n)} \implies t(n) = \Theta \left(\frac{\sqrt{a(n)}}{v(n)} \right), \quad (24)$$

where $\Delta S = \Theta(\sqrt{a(n)})$ is the average distance in one-round trip inside a cell. Note also that the total number of hops is $\Theta(\bar{L}/\sqrt{a(n)})$, and that the speed of each node $v(n)$ must decrease with $1/\sqrt{n}$. Combining all this information, the average delay (D_1) in *Scheme 1* is

$$D_1(n) = (\# \text{ of hops}) \cdot t(n) = \Theta \left(\frac{1}{v(n)} \right) = \Theta(\sqrt{n}). \quad (25)$$

This delay is larger than that obtained by Gupta and Kumar [8], which was shown to be $\Theta(1/\sqrt{a(n)}) = \Theta(\sqrt{n/\log(n)})$ [9]. This is a direct consequence of the throughput-delay trade-off property [9]. *The capacity improvement is obtained at the cost of increase in delay.*

C. Scheme 2

In the previous section we saw that, by having an infinite number of relays (or hops), the capacity of the network decreases as the number of nodes increases. Here, we show that, by having a finite number of relays and using local transmission to overcome interference, we can attain constant throughput as n increases, but we can also trade-off the number of hops with capacity and delay, i.e., we can exchange mobility with capacity and delay, which is a generalization of the results by Grossglauser and Tse [10].

Now the network area is divided into l square cells and l is a network design parameter that does not depend on n . Hence, each cell has area of size $\frac{1}{l}$. Again, we assume that the n nodes are uniformly distributed over the entire network, but each node is restricted to move only inside of its cell (one of the l cells). Among the total number of nodes n , a fraction of them, n_S , are randomly chosen as senders, while the remaining nodes, n_R , function like possible receiving nodes [10]. A sender density parameter θ is defined as $n_S = \theta n$, where $\theta \in (0,1)$, and $n_R = (1 - \theta)n$ as in Section IV. Here, we consider that each node can communicate with its closest neighbor within the transmission range r_o , whether this neighbor is inside its own cell or from an adjacent cell (when it is traveling around the cell boundary). For a uniform distribution of the nodes, $r_o = 1/\sqrt{\theta\pi n}$ (see Section VIII-B.1). Communication between two nodes from the same cell can only be a source-destination, or a relay-destination packet exchange. A relay-relay communication only happens between nodes from different neighboring cells.

A source-destination pair is uniformly chosen among the n nodes, so that the destination does not have to be necessarily in the same cell as its source. A packet may traverse relays to reach its destination. We assume that, once a packet is relayed to a cell, it is not relayed again for another node in the same cell. Instead, the node keeps the packet in its queue until it reaches the neighborhood of an adjacent cell in the path toward the destination, so that it forwards the packet to the closest receiver node in the neighboring cell. In this model there is no fixed communication region as in the previous model. Once the node moves close enough around the cell boundary and there is a neighbor receiver node from the adjacent cell moving within the transmission range r_o , then it relays the packet to this neighbor if there is a packet to forward in

that direction. Thus, it can be either a source-relay, or relay-relay, or relay-destination transmission. The communication is simplex, so that each sender node uses the entire communication time slot to transmit at rate W bits/sec.

Using an analogous analysis as in previous subsection, the following Theorem can be easily proved [32].

Theorem 3 *For Scheme 2, for finite l and sufficiently large n , we have*

$$\Lambda_2(n) = \frac{1}{\sqrt{l}} \Theta(1).$$

Theorem 3 is a generalization of the results by Grossglauser and Tse [10], given that we have divided the network into l equal cells. If we set $l = 1$, Theorem III.5 in [10] follows.

The delivery delay can also be easily derived and shown to be approximated for large n by [32]

$$D_2(n) \approx \Theta\left(\frac{n}{l}\right). \quad (26)$$

Comparing $D_2(n)$ to the delay attained in the scheme by Grossglauser and Tse [10], whose delay was shown to be $\Theta(n)$ [21], [9], we conclude that, as we expected, the delay in *Scheme 2* is smaller by a factor of l .

From Theorem 3, Eq. (26), and comparing with [10], we conclude that we can trade-off mobility as a resource with capacity and delay. By restraining the nodes to move inside cells of size area $\frac{1}{l}$, the $\Theta(1)$ throughput obtained in [10] is reduced by a factor of \sqrt{l} , while the delivery delay is decreased by a factor of l . Thus, *Scheme 2* is a generalization of the network model by Grossglauser and Tse [10].

D. Fixed Nodes with Directional Antennas

In this section, we present a model where nodes are static, but endowed with directional antennas. Previous work [2], [3] has considered capacity analysis for static networks using directional antennas, where they showed that no scheme using directed beams can circumvent the constriction on capacity

in dense networks. In this study, a slightly different modeling approach is presented compared to these previous directional antenna analysis. The communication is constrained to occur only between closest neighbors by using very narrow beams. The network model is shown in Fig. 8. A source-destination pair of nodes is randomly chosen so that a packet is sent from cell a to cell t , for example, relying on multiple relays (or hops) using directional antenna transmission along close neighbors in the path to the destination. The nodes are deployed uniformly in the network area torus. As in *Scheme 1*, the network is divided in $1/a(n)$ cells, each with an area $a(n)$. Let's assume $a(n) \geq 2 \log(n)/n$, so that each cell has at least one node whp [9]. In each cell a node is chosen to relay the traffic of the cell. Fig. 8 shows a source node in cell a that has destination at a node in cell t separated by a distance \bar{L} . Accordingly, the cell path along the closest neighbors is $\{a \rightarrow f \rightarrow g \rightarrow h \rightarrow m \rightarrow n \rightarrow o \rightarrow t\}$.

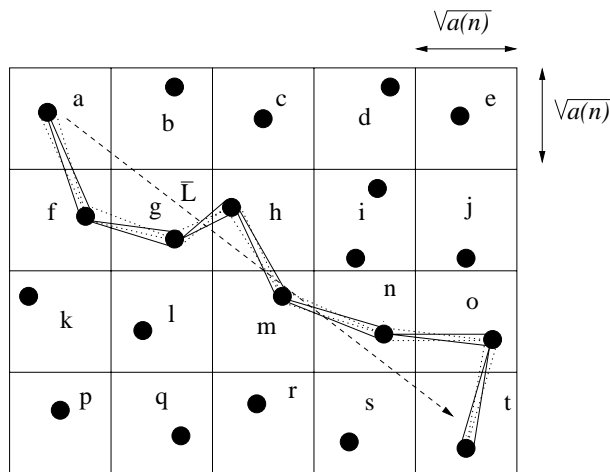


Fig. 8. Unit area torus network divided into $1/a(n)$ cells each with size area of $a(n)$. Transmissions are employed using bi-directional antennas, with very narrow beams, between closest neighbors from adjacent cells along the path to destination.

We want to obtain the average throughput for a source-destination pair uniformly chosen among all n nodes, as well as the delay behavior. The relay transmissions are scheduled at regular time intervals so that each node is assigned a time slot to transmit successfully to its closest neighbor in the path to the chosen destination. This is a time schedule constraint because a node can only point its antenna to a close neighbor at consecutive time intervals. For the example shown in Fig. 8, each node has eight neighbors, given that we assume a torus net, so that it can communicate to each of them at regular eight

slot time interval respectively, i.e., a time division multiple access (TDMA) with bi-directional beam transmission. Each time two nodes point their antennas toward each other, they exchange packets, so that each of these exchanges can involve either source-relay, relay-relay, or relay-destination transmissions. Interference is overcome by the use of directional beams to the nearest neighbor, so that Eq. (1) is satisfied. The transmissions are half duplex, i.e., the communication time slot is divided in two equal parts. Each node transmits at W bits/sec. Hence, the average available bit rate is $W/2$ bits/sec.

Using similar analysis as in subsection IX-B, the following Theorem can also be easily proved [32].

Theorem 4 *For a given node using directional antenna transmission to closest neighbor along the path to destination, with $a(n) = \frac{k \log(n)}{n}$, for $k \geq 2$, to guarantee connectivity, we have*

$$\Lambda_D(n) = \Theta \left(\sqrt{\frac{\log(n)}{n}} \right).$$

This result represents a better bound on throughput capacity than what Gupta and Kumar [8] obtained which was $\Theta(1/\sqrt{n \log(n)})$, and the results by Yi et al [2]. Indeed, it is a gain of $\Theta(\log(n))$ and is similar to Peraki and Servetto's results [3] obtained for a single directed beam, where they use a different approach applying networking flow analysis to calculate the network transport capacity (i.e., maximum stable throughput). This is the same capacity scalability obtained for *Scheme 1*. We see that capacity is still constrained in dense networks. This is due to the wasting of the available bandwidth to forward the same packet over multiple hops by an amount of time that scales with n .

The average delay incurred by a packet to reach the destination is the sum of the average time a packet spends hopping along the path to its destination. The total number of hops to reach destination is $\Theta(\bar{L}/\sqrt{a(n)})$. Accordingly, the delay using directional antenna transmission to nearest neighbor is given by

$$D_D(n) = (\# \text{ of hops}) \Delta t = \Theta \left(\frac{1}{\sqrt{a(n)}} \right) = \Theta \left(\sqrt{\frac{n}{\log(n)}} \right). \quad (27)$$

TABLE II

THROUGHPUT GAIN AND DELAY INCREASE OBTAINED FROM COMPARING PREVIOUS WORKS [8], [10] WITH RESTRICTED MOBILITY SCHEMES AND DIRECTIONAL ANTENNA TRANSMISSION.

Schemes comparisons	Throughput gain	Delay increase
<u>Scheme 1</u> Gupta & Kumar	$\log(n)$	$\sqrt{\log(n)}$
<u>Grossglauer & Tse</u> Scheme 2	\sqrt{l}	l
<u>Directional antenna</u> Gupta & Kumar	$\log(n)$	none
<u>Scheme 1</u> Directional antenna	none	$\sqrt{\log(n)}$

Compared to Eq. (25) this represents a delay reduction of $\Theta(1/\sqrt{\log(n)})$. Thus, the use of directional antenna with fixed nodes offers a smaller delay on average than the restricted mobility case, while attaining the same throughput scalability as *Scheme 1*.

Therefore, employing directional antenna transmissions between closest nodes along the path to a destination is equivalent, in terms of throughput performance, to nodes executing restricted mobility as in *Scheme 1*, while providing a smaller packet delivery delay.

E. Performance Comparisons

To obtain a benchmark of throughput and delay for wireless ad hoc networks, we compare in Table II the schemes studied with the previous works by Gupta and Kumar [8], and Grossglauer and Tse [10]. The results suggest that using mobility or enhanced physical layer properties (directional antennas in this case) can improve throughput or delay.

X. CONCLUSIONS

In this chapter, we have studied the performance of wireless networks. We revisited some of the main results in the literature for capacity and delay in ad hoc and sensor networks.

We reviewed a multi-copy forwarding scheme of packets, such that the delivery delay is substantially reduced without changing the order of the capacity of the network. Using the results of random occupancy,

we provided a mathematical formula for computation of the node throughput as a function of the network parameters. The delay analysis was corroborated by simulation results showing the exponential delay reduction obtained. We also showed that interference can be overcome if transmission is restricted among close neighbors, where an analytical method for computation of the interference was presented.

Then we revised a study on mobility-capacity-delay trade-off for wireless ad hoc networks. By restricting the movements of the nodes inside a cell, we showed that mobility is an entity that can be exchanged with capacity and delay. We employed two different schemes for such analysis. In the first case, the area of the cells in the network is a function of the total number of nodes n . We found that a throughput gain of $\Theta(\log(n))$ is obtained, compared to the case of Gupta and Kumar [8], where a penalty in delay is the cost for such an improvement. In the second case, we assumed the area of the cells to be independent of n . We showed that a constant asymptotic throughput is obtained and that our results are a generalizations of the results by Grossglauser and Tse [1]. In addition, we observed that by changing the physical layer properties of the wireless network (e.g., using directional antennas) the throughput or delay performance can be improved and the result obtained agrees with previous work which employs network flow analysis.

REFERENCES

- [1] M. Grossglauser and D. Tse, "Mobility increases the capacity of wireless ad-hoc networks," in *Proc. of IEEE Infocom*, Anchorage, Alaska, March 2001.
- [2] S. Yi, Y. Pei, and S. Kalyanaraman, "On the capacity improvement of ad hoc wireless networks using directional antennas," in *Proc. of ACM MobiHoc*, Annapolis, Maryland, June 2003.
- [3] C. Peraki and S. D. Servetto, "On the maximum stable throughput problem in random networks with directional antennas," in *Proc. of ACM MobiHoc*, Annapolis, Maryland, June 2003.
- [4] B. Liu, Z. Liu, and D. Towsley, "On the capacity of hybrid wireless networks," in *Proc. of IEEE Infocom*, San Francisco, California, March 2003.
- [5] W. Yuen, R. Yates, and S.-C. Mau, "Exploiting data diversity and multiuser diversity in noncooperative mobile infostation networks," in *Proc. of IEEE Infocom*, San Francisco, California, March 2003.
- [6] E. Perevalov and R. Blum, "Delay limited capacity of ad hoc networks: asymptotically optimal transmission and relaying strategy," in *Proc. of IEEE Infocom*, San Francisco, California, March 2003.

- [7] N. Bansal and Z. Liu, "Capacity, delay and mobility in wireless ad-hoc networks," in *Proc. of IEEE Infocom*, San Francisco, California, March 2003.
- [8] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, March 2000.
- [9] A. E. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-delay trade-off in wireless networks," in *Proc. of IEEE Infocom*, Hong Kong, March 2004.
- [10] M. Grossglauser and D. Tse, "Mobility increases the capacity of wireless ad hoc networks," *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, pp. 477–486, August 2002.
- [11] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. of IEEE ICC*, Seattle, Washington, June 1995.
- [12] R. Motwani and P. Raghavan, *Randomized Algorithms*. Cambridge Univ. Press, 1995.
- [13] S. R. Kulkarni and P. Viswanath, "A deterministic approach to throughput scaling in wireless networks," *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1041–1049, June 2004.
- [14] M. Franceschetti, O. Dousse, D. Tse, and P. Tiran, "Closing the gap in the capacity of random wireless networks," in *Proc. of IEEE ISIT*, Chicago, Illinois, June 2004.
- [15] T. J. Shepard, "A channel access scheme for large dense packet radio networks," in *Proc. of ACM SigComm*, San Francisco, California, August 1996.
- [16] R. Negi and A. Rajeswaran, "Capacity of power constrained ad-hoc networks," in *Proc. of IEEE Infocom*, Hong Kong, March 2004.
- [17] L. Bao and J. J. Garcia-Luna-Aceves, "Transmission scheduling in ad hoc networks with directional antennas," in *Proc. of ACM Mobicom*, Atlanta, Georgia, September 2002.
- [18] R. R. Choudhuri, R. R. X. Yang, and N. H. Vaidya, "Using directional antennas for medium access control in ad hoc networks," in *Proc. of ACM Mobicom*, Atlanta, Georgia, September 2002.
- [19] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Communication & Mobile Computing (WCMC): Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications*, vol. 2, no. 5, pp. 483–502, 2002.
- [20] C. Bettstetter, H. Hartenstein, and X. Pérez-Costa, "Stochastic properties of the random waypoint mobility model: epoch length, direction distribution, and cell change rate," in *Proc. of ACM MSWiM*, Atlanta, Georgia, September 2002.
- [21] M. J. Neely and E. Modiano, "Improving delay in ad-hoc mobile networks via redundant packet transfers," in *Proc. of Conference on Information Sciences and Systems*, Baltimore, Maryland, March 2003.
- [22] S. Toumpis and A. J. Goldsmith, "Large wireless networks under fading, mobility, and delay constraints," in *Proc. of IEEE Infocom*, Hong Kong, March 2004.
- [23] S. N. Diggavi, M. Grossglauser, and D. Tse, "Even one-dimensional mobility increases ad hoc wireless capacity," in *Proc. of IEEE ISIT*, Laussane, Switzerland, June 2002.

- [24] Y. Azar, A. Z. Broder, A. R. Karlin, and E. Upfal, "Balanced allocations," in *Proc. of ACM STOC*, Montreal, Canada, May 1994.
- [25] M. Mitzenmacher, "The power of two choices in randomized load balancing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 12, no. 10, pp. 1094–1104, October 2001.
- [26] O. Dousse, P. Thiran, and M. Hasler, "Connectivity in ad-hoc and hybrid networks," in *Proc. of IEEE Infocom*, New York, New York, June 2002.
- [27] A. Agarwal and P. R. Kumar, "Capacity bounds for ad hoc and hybrid wireless networks," *ACM SIGCOMM Computer Communications Review*, vol. 34, no. 3, pp. 71–81, July 2004.
- [28] R. M. de Moraes, H. R. Sadjadpour, and J. J. Garcia-Luna-Aceves, "Throughput-delay analysis of mobile ad-hoc networks with a multi-copy relaying strategy," in *Proc. of IEEE SECON*, Santa Clara, California, October 2004.
- [29] D. Nain, N. Petigara, and H. Balakrishnan, "Integrated routing and storage for messaging applications in mobile ad hoc networks," in *Proc. of WiOpt*, Antipolis, France, March 2003.
- [30] C. de Waal and M. Gerharz, "Bonnmotion: A mobility scenario generation and analysis tool," <http://www.cs.uni-bonn.de/IV/BonnMotion/>, 2003.
- [31] R. M. de Moraes, H. R. Sadjadpour, and J. J. Garcia-Luna-Aceves, "On mobility-capacity-delay trade-off in wireless ad hoc networks," in *Proc. of IEEE/ACM MASCOTS*, Volendam, The Netherlands, October 2004.
- [32] R. M. de Moraes, H. R. Sadjadpour, and J. J. Garcia-Luna-Aceves, "Mobility-capacity-delay trade-off in wireless ad hoc networks," *Elsevier Ad Hoc Networks Journal*, vol. 4, no. 5, pp. 607–620, September 2006.