Mobility Increases the Capacity of Ad-hoc Wireless Networks

Matthias Grossglauser
AT&T Labs- Research
180 Park Avenue
Florham Park NJ 07932
mgross@research.att.com

David Tse
Department of EECS
University of California
Berkeley CA 94720
dtse@eecs.berkeley.edu

Abstract

The capacity of ad-hoc wireless networks is constrained by the mutual interference of concurrent transmissions between nodes. We study a model of an ad-hoc network where n nodes communicate in random source-destination pairs. These nodes are assumed to be mobile. We examine the per-session throughput for applications with loose delay constraints, such that the topology changes over the time-scale of packet delivery. Under this assumption, the per-user throughput can increase dramatically when nodes are mobile rather than fixed. This improvement can be achieved by exploiting node mobility as a type of *multiuser diversity*.

1 Introduction

A fundamental characteristic of mobile wireless networks is the time variation of the channel strength of the underlying communication links. Such time variation can be due to multipath fading, path loss via distance attenuation, shadowing by obstacles and interference from other users. The impact of such time variation on the design of wireless networks permeates throughout the layers, ranging from coding and power control at the physical layer to cellular handoff and coverage planning at the networking layer.

An important means to cope with the time-variation of the channel is the use of *diversity*. Diversity can be obtained over time (interleaving of coded bits), frequency (combining of multipaths in CDMA systems) and space (multiple antennas). The basic idea is to improve performance by having several independent signal paths between the transmitter and the receiver.

These diversity modes pertain to a point-to-point link. Recent results point to another form of diversity, inherent in a wireless network with multiple users. This *multiuser diversity* is best motivated by an information theoretic result of Knopp and Humblet [6]. They focused on the uplink in the single cell, with multiple users communicating to the basestation via time-varying channels. To maximize the total information theoretic capacity, they showed that the optimal strategy is to schedule at any one time only the user with the best channel to transmit to the basestation. Diversity gain arises from the fact that in a system with many users, there is likely to be a user with a very good channel at any one time. Overall system throughput is maximized by allocating at any time the common channel resource to the user that can best exploit it.

Strategies of this type incur additional delay, because packets have to be buffered until the channel becomes strong relative to other users. Therefore, the time-scale of channel fluctuations that can be exploited through multiuser diversity is limited by the delay tolerance of the user or application. For example, for applications that can tolerate delays on the order of fractions of seconds to several seconds, short time-scale multipath fading can be taken advantage of. In this paper, the focus is on applications that are so asynchronous in nature that they can tolerate end-to-end delays of minutes or even hours. On such a long time-scale, even more diversity gain can be obtained because the *network topology* changes significantly over time due to user mobility. Examples of such applications include electronic mail, database synchronization between a mobile terminal and a central database, and certain types of event notification.

We demonstrate in this paper that these ideas have ramifications to the design of wireless networks beyond classical cellular architectures. We will focus on mobile ad-hoc networks which have no fixed basestations and with multiple pairs of users wanting to communicate with each other. Gupta and Kumar [4] proposed a model for studying the capacity of *fixed* ad-hoc networks, where nodes are randomly located but are immobile. Each source node has a random destination to which it wants to communicate. Their main result shows that as the number of nodes per unit area n increases, the throughput per source-destination (S-D) pair decreases approximately like $1/\sqrt{n}$.

This is the best performance achievable even allowing for optimal scheduling, routing and relaying of packets in the networks, and is a somewhat pessimistic result on the scalability of such networks, as the traffic rate per S-D pair actually goes to zero.

In this paper, we introduce mobility into the model and consider the situation when users move independently around the network. Our main result shows that the average long-term throughput per S-D pair can be kept *constant* even as the number of nodes per unit area n increases. This is in sharp contrast to the fixed network scenario, and the dramatic performance improvement is obtained through the exploitation of the time-variation of the users' channels due to mobility. We observe that our result implies that, at least in terms of growth rate as a function of n, there is no significant loss in throughput per S-D pair when there are many nodes in the network as compared to having just a single S-D pair. A caveat of this result is that the attained long-term throughput is averaged over the time-scale of node mobility, and hence delays of that order will be incurred.

In the fixed ad-hoc network model, the fundamental performance limitation comes from the fact that long-range direct communication between many user pairs is infeasible, due to the excessive interference caused. As a result, most communication has to occur between nearest neighbors, at distances of order $1/\sqrt{n}$, with each packet going through many other nodes (serving as relays) before reaching the destination. The number of hops in a typical route is of order $\sqrt{n}$. Because much of the traffic carried by the nodes are relayed traffic, the actual useful throughput per user pair has to be small.

With mobility, a seemingly natural strategy to overcome the above performance limitation is to transmit only when the source and destination nodes are close together, at distances of order $1/\sqrt{n}$. This is reminiscent of the Infostation architecture [3], where users connect to the infostations only when they are close by. However, this strategy turns out to be too naive in the present situation. The problem is that the fraction of time two nodes are nearest neighbors is too small, of the order of $1/n$. Instead, our strategy is for each source node to distribute its packets to as many different nodes as possible. These then serve as mobile relay nodes and whenever they get close to the final destination, they hand the packets off to the final destination. The basic idea is that since there are many different relay nodes, the probability that at least one is close to the destination is significant. On the other hand, each packet goes through at most one relay node, and hence the throughput can be kept high. Although the basic communication problem is point-to-point, this strategy effectively creates multiuser diversity by distributing packets to many different intermediate nodes which have independent time-varying channels to the final destination.

2 Model

The ad-hoc network consists of n nodes all lying in the open disk of unit area (of radius $1/\sqrt{\pi}$). The location of the ith user at time t is given by $X_i(t)$. Nodes are mobile, and we assume that the process $\{X_i(\cdot)\}$ is stationary and ergodic with stationary distribution uniform on the open disk; moreover the trajectories of different users are independent and identically distributed.

We first describe the session model. We assume that each of the n nodes is a *source* node for one session, and a *destination* node for another session. Let us stipulate that the source node i has data intended for destination node $d(i)$. We assume that each source node has an infinite reservoir of packets to send to its destination. The source-destination association does not change with time, although the nodes themselves move.

We next describe the transmission model. At (slotted) time t, let $P_i(t)$ be the transmit power of node i, and $\gamma_{ij}(t)$ be the channel gain from node i to node j, such that the received power at node j is $P_i(t)\gamma_{ij}$. At time t, node i transmits data at rate R packets/sec to node j if

$$\frac{P_i(t)\gamma_{ij}(t)}{N_0 + \frac{1}{L}\sum_{k\neq i} P_k(t)\gamma_{ij}(t)} > \beta, \qquad (1)$$

where β is the signal-to-interference ratio (SIR) requirement for successful communication, N_0 is the background noise power, and L is the *processing gain* of the system. For a narrowband system, $L = 1$, while for a spread-spectrum CDMA system, L is larger than 1. In this paper, we only consider large-scale path loss characteristics in the fading channel model. The channel gain is given by

$$\gamma_{ij}(t) := \frac{1}{|X_i(t) - X_j(t)|^\alpha},$$

where α is a parameter greater than 2.

We will consider two models. In the first model, packets are only transmitted directly from the source to the destination, without relaying through other nodes. In other words, $j(i) = d(i)$. In the second model, nodes can serve as *relays* for packets intended for other nodes. We assume that each node has an infinite buffer to store such packets.

At any time t, a *scheduler* chooses which nodes will be senders, and the power levels $P_i(t)$ for these senders. The objective of the scheduler for both models is to ensure a high long-term throughput for each S-D pair. More precisely, consider a scheduling and relay policy π. Let $M_i^\pi(t)$ be the number of source node i packets that destination $d(i)$ receives at time t under policy π. Given the random trajectories of the users, we shall say a long term throughput of $\lambda(n)$ is feasible if there is a policy π such that for every S-D pair i,

$$\liminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} M_i^\pi(t) \geq \lambda(n).$$

We note that the throughput $\lambda(n)$ is a random quantity as it depends on the random locations of the users. The indexing by the system size n emphasizes that we are interested in studying the asymptotic behavior as n becomes large.

3 Results

3.1 Fixed Nodes

First we review results of Gupta and Kumar [4]. They focus on a static model, where nodes are not mobile. The node positions $\{X_i\}$ are i.i.d. and uniformly distributed in the open disk of unit area, but fixed over time. The destination for each source node is chosen to be the node closest to a randomly chosen point on the disk. The destinations are all chosen independently. Relaying is allowed in their model. The following results yield upper and lower bounds on the asymptotically feasible throughput.

Theorem 3.1 *(Main result 4 in [4]) There exists constants c and c′ such that*

$$\lim_{n\to\infty} Pr\left\{\lambda(n) = \frac{cR}{\sqrt{n\log n}} \quad \textit{is feasible}\right\} = 1,$$

and

$$\lim_{n\to\infty} Pr\left\{\lambda(n) = \frac{c'R}{\sqrt{n}} \quad \textit{is feasible}\right\} = 0,$$

Thus, within a factor of $\sqrt{\log n}$, the throughput per S-D pair goes to zero like $R/\sqrt{n}$ in the case when the nodes are fixed.

3.2 Mobile Nodes Without Relaying

The reason why the throughput for fixed nodes goes to zero is that the number of relay nodes a packet has to go through scales as $\sqrt{n}$. However, in our model of mobile nodes, any two nodes can be expected to be close to each other every from time to time. This suggests that we may be able to improve the capacity by not relaying at all, and only letting sources transmit directly to destinations. We now show that without relaying, there is no way to achieve an $0(1)$ throughput per O-D pair.

We first need the following Lemma. This fact is already established in the proof of Theorem 2.1(ii) in [4], but we include the proof here for completeness.

Lemma 3.2 *Consider a scheduling policy that schedules direct transmissions only. Fix an arbitrary time t. Let $\mathcal{S}(t)$ be the set of source nodes that are scheduled successful transmission to their respective destinations. Then*

$$\sum_{i\in\mathcal{S}(t)} |X_i(t) - X_{j(i)}(t)|^\alpha \leq B,$$

where

$$B := 2^\alpha \pi^{-\alpha/2}\frac{\beta+L}{\beta}.$$

Proof: Writing down the SIR inequalities, we get for every $i \in \mathcal{S}(t)$,

$$\frac{P_i(t)\gamma_{i,j(i)}(t)}{N_0 + \frac{1}{L}\sum_{k\in\mathcal{S}(t),k\neq i} P_k(t)\gamma_{k,j(i)}(t)} \geq \beta.$$

This is equivalent to:

$$\frac{P_i(t)\gamma_{i,j(i)}(t)}{N_0 + \frac{1}{L}\sum_{k\in\mathcal{S}(t)} P_k(t)\gamma_{k,j(i)}(t)} \geq \frac{\beta L}{\beta+L}.$$

Substituting

$$\gamma_{ij}(t) = \frac{1}{|X_i(t) - X_j(t)|^\alpha},$$

we get the bound:

$$\begin{aligned} & |X_i(t) - X_{j(i)}(t)|^\alpha && (2)\\ \leq\ & \frac{\beta+L}{\beta L}\frac{P_i(t)}{N_0 + \frac{1}{L}\sum_{k\in\mathcal{S}(t)} \frac{P_k(t)}{|X_k(t)-X_{j(i)}(t)|^\alpha}} \\ \leq\ & \frac{\beta+L}{\beta L}\frac{P_i(t)}{N_0 + \frac{1}{L}(\frac{\pi}{4})^{\frac{\alpha}{2}}\sum_{k\in\mathcal{S}(t)} P_k(t)} && (3)\end{aligned}$$

since $|X_k(t) - X_{j(i)}(t)| \leq \frac{2}{\sqrt{\pi}}$. Summing over all active $S-D$ pairs at time t, we get

$$\begin{aligned} & \sum_{i\in\mathcal{S}(t)} |X_i(t) - X_{j(i)}(t)|^\alpha \\ \leq\ & \frac{\beta+L}{\beta L}\frac{\sum_{i\in\mathcal{S}(t)} P_i(t)}{N_0 + \frac{1}{L}(\frac{\pi}{4})^{\frac{\alpha}{2}}\sum_{k\in\mathcal{S}(t)} P_k(t)} \\ \leq\ & 2^\alpha\pi^{-\alpha/2}\frac{\beta+L}{\beta},\end{aligned}$$

which proves the Lemma upon setting

$$B := 2^{\alpha}\pi^{-\alpha/2}\frac{\beta + L}{\beta}.$$

■

This Lemma shows that the number of simultaneous long-range communication is limited by interference. Since the distance between the source and destination is $O(1)$ most of the time, this limitation in turn puts a bound on the performance of any strategy which uses only direct communication.

Theorem 3.3 *Assuming that the policy is only allowed to schedule direct transmission between the source and destination nodes, and no relaying is permitted. If c is any constant satisfying*

$$c > \left[2^{\alpha}\left(1 + \frac{2}{\alpha}\right)\pi^{-\alpha/2}\frac{\beta + L}{\beta}\right]^{\frac{1}{1+\alpha/2}},$$

then

$$Pr\left\{\lambda(n) = cn^{-\frac{1}{1+\alpha/2}}R \quad \textit{is feasible}\right\} = 0$$

for sufficiently large n.

This result says that without relaying, the achievable throughput per S-D pair goes to zero at least as fast as $n^{-\frac{1}{1+\alpha/2}}$.

Proof: We will argue by contradiction. Fix a $c > 0$ and a policy π that schedules direct transmission only, and suppose a throughput of $\lambda(n) = cn^{-\frac{1}{1+\alpha/2}}R$ is feasible. Focus on a source node i, and let $\mathcal{A}_T(i)$ be the set of time instants up until time T where node i is scheduled successful transmission to the destination $d(i)$. By definition of feasible throughputs,

$$\liminf_{T\to\infty}\frac{|\mathcal{A}_T(i)|}{T} \geq cn^{-\frac{1}{1+\alpha/2}}. \tag{4}$$

Consider the process

$$D_i(t) := |X_i(t) - X_{j(i)}(t)|^{\alpha}, \qquad t = 1, 2, \ldots$$

By stationarity and ergodicity of this process, (4) implies that almost surely,

$$\liminf_{T\to\infty}\frac{1}{T}\sum_{t\in\mathcal{A}_T(i)} D_i(t) \geq \int_0^{F^{-1}(cn^{-\frac{1}{1+\alpha/2}})} zdF(z)$$

where F is the cdf of the random variable $D_i(t)$. This holds for all source nodes i. Summing over all i, we have

$$\liminf_{T\to\infty}\frac{1}{T}\sum_{i=1}^{n}\sum_{t\in\mathcal{A}_T(i)} D_i(t) \geq n\int_0^{F^{-1}(cn^{-\frac{1}{1+\alpha/2}})} zdF(z),$$

which is equivalent to:

$$\liminf_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{i\in\mathcal{S}(t)} D_i(t) \geq n\int_0^{F^{-1}(cn^{-\frac{1}{1+\alpha/2}})} zdF(z).$$

Here $\mathcal{S}(t)$ is the set of source nodes which are scheduled successful transmission by the policy at time t. The last inequality in turn implies that there must exist a time τ, such that

$$\sum_{i\in\mathcal{S}(\tau)} D_i(\tau) \geq n\int_0^{F^{-1}(cn^{-\frac{1}{1+\alpha/2}})} zdF(z). \tag{5}$$

Conditional on $X_{j(i)}(t) = x$ in the open disk D, it holds that for $z^{1/\alpha} < |\pi^{-1/2} - x|$,

$$\Pr\left\{D_i(t) < z|X_{j(i)}(t) = x\right\} = \pi z^{2/\alpha},$$

the probability that node i is within a neighborhood of radius z from node $d(i)$. Hence,

$$\begin{aligned}
& \lim_{z\to 0} F(z)/z^{2/\alpha} \\
= & \lim_{z\to 0} z^{-2/\alpha}\int_{x\in D}\Pr\left\{D_i(t) < z|X_{j(i)}(t) = x\right\}dx \\
= & \int_{x\in D}\lim_{z\to 0} z^{-2/\alpha}\Pr\left\{D_i(t) < z|X_{j(i)}(t) = x\right\}dx \\
= & \pi
\end{aligned}$$

where the interchange of limit and integration follows from the Dominated Convergence Theorem.

Substituting this into the integral in (5), we get

$$\lim_{n\to\infty} n\int_0^{F^{-1}(cn^{-\frac{1}{1+\alpha/2}})} zdF(z) = \frac{c^{1+\alpha/2}}{\pi^{\alpha}(1 + 2/\alpha)}.$$

If

$$c > \left[2^{\alpha}\left(1 + \frac{2}{\alpha}\right)\pi^{\alpha/2}\frac{\beta + L}{\beta}\right]^{\frac{1}{1+\alpha/2}},$$

then

$$\lim_{n\to\infty} n\int_0^{F^{-1}(cn^{-\frac{1}{1+\alpha/2}})} zdF(z) > B$$

where

$$B := 2^{\alpha}\pi^{-\alpha/2}\frac{\beta + L}{\beta}.$$

Hence, for sufficiently large n, inequality (5) contradicts Lemma 3.2. For sufficiently large n, the probability that $cn^{-\frac{1}{1+\alpha/2}}R$ is a feasible throughput is zero. ■

3.3 Mobile Nodes With Relaying

The problem of only using direct transmission between the source and destination is that too much of the communication is long-range, and the resulting interference limits the number of concurrent transmissions. This suggests that the total throughput of the network can be increased if we constrain transmission to nearest neighbors. Indeed, Theorem 3.4 below demonstrates that $O(n)$ *concurrent successful transmissions* per time slot are possible. The problem with this approach is that the source and the destination node of a session are nearest neighbors only for a very small fraction of time, on the order of $1/n$, vanishing as the user density increases. Therefore, although we can schedule $O(n)$ communicating sender-receiver pairs per time slot, the fraction of pairs that actually has a packet to transmit vanishes for large n.

The idea now is to spread out packets to a large number of intermediate *relay nodes* that temporarily buffer packets until final delivery to the destination is possible. For a source-destination pair S-D, all the other $n-2$ nodes can serve as relay nodes. The goal is that in steady-state, the packets of every source node will be distributed across all the nodes in the network, hence ensuring that every node in the network will have packets buffered destined to every other node (except itself). This ensures that a scheduled sender-receiver pair always has a packet to send, in contrast to the case of direct transmission.

The question is how many times a packet has to be relayed in order to spread traffic uniformly to all nodes. In fact, as the node location processes $\{X_i(t)\}$ are independent, stationary and ergodic, it is actually sufficient to *relay only once*. This is because the probability for an arbitrary node to be scheduled to receive a packet from a source node S is equal for all nodes and independent of S. Each packet then makes two hops, one from the source to its random relay node, and one from that relay node to the destination. As no packet is transmitted more than twice, the achievable total throughput is $O(n)$.

We now make the above argument rigorous. We first exhibit a scheduling policy π to select random sender-receiver pairs in each time slot t, such that all the pairs can successfully transmit in time slot t. We will then use this policy as a building block to achieve $O(1)$ throughput per S-D pair for large n.

The scheduling policy π is as follows. Let us focus on a particular time slot t. To simplify notation, we will drop the time index t in the following discussion. We fix a *sender density* parameter $\theta \in (0,1)$. We randomly designate $n_S = \theta n$ of the nodes as senders in each time slot, and the remaining n_R nodes as *potential* receivers. Specifically, we randomly pick one out of $\binom{n_S}{n}$ equally likely partitions of the n nodes into the set of senders $\mathcal{S}$ and the set of potential receivers $\mathcal{R}$. Each sender node transmits packets to its nearest neighbor *among all nodes in* $\mathcal{R}$, using unit transmit power ($P_i = 1$). Among the n_S sender-receiver pairs, we retain those for which the interference generated by the other senders is sufficiently small that transmission is possible. Let N_t be the number of such pairs. Theorem 3.4 below shows that the number of feasible sender-receiver pairs N_t is $O(n)$. Note that the set of sender-receiver pairs is random and that it depends only on the node locations $\{X_i\}$.

Theorem 3.4 *For the scheduling policy π, the expected number $E[N_t]$ of feasible sender-receiver pairs is $O(n)$, i.e.,*

$$\lim_{n\to\infty} \frac{E[N_t]}{n} = \phi > 0. \tag{6}$$

Furthermore, for two arbitrary nodes i and j, the probability that (i,j) is scheduled as a sender-receiver pair is $O(1/n)$.

We can now apply this scheduling policy π to our basic problem. The overall algorithm is divided into two phases: (1) scheduling of packet transmissions from sources to relays (or the final destination), and (2) scheduling of packet transmissions from relays (or the source) to final destinations. These two phases are interleaved: in the even time-slots, phase 1 is run; in the odd time-slots, phase 2 is run.

In phase 1, we can apply the scheduling policy π to transmit packets from sources to relays or destinations. In phase 2, we again apply the policy π, but this time to transmit packets from relays to final destinations (or, as in phase 1, from a source directly to the destination). More specifically, when a receiver is identified for a sender under π, the sender checks if it has any packets for which the receiver is the destination; if so, it will transmit it. It should be noted that every packet goes through at most two hops: it is transmitted once in phase 1 from its source to an intermediate relay, and once in phase 2 from a relay to the final destination. We allow for packets to be directly transmitted from their source to their destinations in both phases, if a sender-receiver pair happens to be a source-destination pair as well.

Let us analyze the throughput per S-D pair under this two-phased scheme. As π only depends on node locations and because the node locations $\{X_i(t)\}$ are i.i.d., stationary, and ergodic, the long-term throughput between any two nodes is equal to the probability that these two nodes are

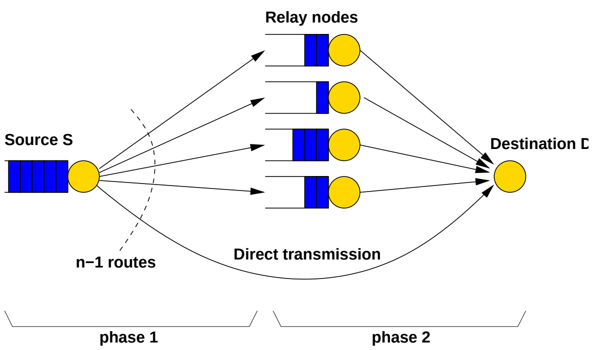

Figure 1: The two-phase scheduling policy viewed as a queuing system, for a source-destination pair: in phase 1, a packet at S is served by a queue of capacity $O(1)$, and is forwarded either to the destination or to one of $n-2$ relay nodes with equal probability. The service rate at each relay node R is $O(1/n)$, for a total session rate of $O(1)$.

selected by π as a feasible sender-receiver pair. According to Theorem 3.4, this probability is $O(1/n)$. Now, for a given S-D pair, there is one direct route and $n-2$ two-hop routes which go through one relay node R. The throughput over the direct route is $O(1/n)$. For each two-hop route, we can consider the relay node R as a single server queue (cf. Fig. 1). Applying Theorem 3.4, we see that both the arrival rate and the service rate of this queue is the same and $O(1/n)$. Summing over the throughputs of all the $n-1$ routes, it can be seen that the total average throughput per S-D pair is $O(1)$. We have proved the following Theorem, which is the main result of this paper.

Theorem 3.5 *The two-phased algorithm achieves a throughput per S-D pair of $O(1)$, i.e. there exists a constant $c>0$ such that*

$$\lim_{n\to\infty} Pr\left\{\lambda(n) = cR \quad \textit{is feasible}\right\} = 1.$$

Note that the largest possible throughput is $c = \phi/2$. We now prove Theorem 3.4.

Proof: We consider a fixed time t. Let $U_1, \ldots U_{n_S}$ be the random positions of the senders in $\mathcal{S}$. Let $V_1, \ldots, V_{n_R}$ be the positions of nodes in the receiver set $\mathcal{R}$. These random variables are i.i.d. uniformly distributed on the open disk of unit area. For each node $s \in \mathcal{S}$, let its intended receiver $r(s) \in \mathcal{R}$ be the node which is nearest to s among all nodes in $\mathcal{R}$.

We now analyze the probability of successful transmission for each chosen sender-receiver pair. By symmetry, we can just focus on one such pair, say $(1, r(1))$. The event of successful transmission depends on the positions $U_1, \ldots U_{n_S}$ and $V_1, \ldots, V_{n_R}$. Let Q_i be the received power from sender node i at receiver node $r(1)$, and

$$Q_i = |U_i - V_{r(1)}|^{-\alpha}.$$

The node $r(1)$ satisfies:

$$r(1) = \text{argmin}_j |U_1 - V_j|.$$

The total interference at node $r(1)$ is given by $I = \sum_{i\neq 1} Q_i$. The signal-to-interference ratio for the transmission from sender 1 at receiver $r(1)$ is given by:

$$\text{SIR} = \frac{Q_1}{N_0 + \frac{1}{L} I}.$$

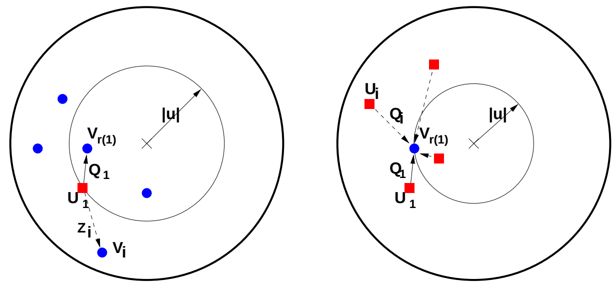

Figure 2: An illustration of random variables used in the proof: sender location U_1, receiver location $V_{r(1)}$, received signal power Q_1, scaled distance to random receiver Z_i, and scaled interfering sender distance Q_i.

We now analyze the asymptotics of Q_1 and I as $n \to \infty$. Now,

$$Q_1 = \max_{j=1,\ldots,n_R} Z_j,$$

where $Z_j = |U_1 - V_j|^{-\alpha}$. Let us first condition on $U_1 = u$ for some u in the open disk. A disk centered at u and of radius $r < (\pi^{-1/2} - |u|)$ lies entirely inside the unit disk (cf. Fig. 2). Then for every $z > r^{-\alpha}$ and for all j,

$$\begin{aligned} \Pr\left\{Z_j > z | U_i = u\right\} &= \Pr\left\{|V_j - u| < z^{-\frac{1}{\alpha}}\right\} \\ &= \pi z^{-\frac{2}{\alpha}} \end{aligned} \tag{7}$$

Conditional on $U_1 = u$, the random variables Z_j's are i.i.d. By a standard result on the asymptotic distribution of extremum of i.i.d. random variables [1, p.258-260], the extremum Q_1 of n_R i.i.d. random variables whose cdf satisfies

$$\lim_{n_R\to\infty} \frac{1 - F_Z(x)}{1 - F_Z(kx)} = k^b \tag{8}$$

satisfies

$$\lim_{n_R\to\infty} \Pr\left\{Q_1 \le a_{n_R} x\right\} = \exp\left(-x^{-b}\right), \tag{9}$$

where a_{n_R} is given by $F_Z^{-1}(1-1/n_R) = (\pi n_R)^{\alpha/2} = [(1-\theta)\pi n]^{\alpha/2}$. Thus, the asymptotic distribution of Q_1 conditional on $U_1 = u$ depends only on the tail of the distribution of the Z_j's, and is given by:

$$\lim_{n\to\infty} \Pr\{Q_1 < a_{n_R} x | U_1 = u\} = F_{Q_\alpha^*}(x) \quad (10)$$

where Q_α^* has a cdf:

$$F_{Q_\alpha^*}(x) = \begin{cases} \exp\left(-x^{-2/\alpha}\right) & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Hence, for every $x > 0$,

$$\begin{aligned} & \lim_{n\to\infty} \Pr\{Q_1 < a_{n_R} x\} \\ = & \lim_{n\to\infty} \int_{u\in D} \Pr\{Q_1 < a_{n_R} x | U_1 = u\}\, du \\ = & \int_{u\in D} \lim_{n\to\infty} \Pr\{Q_1 < a_{n_R} x | U_1 = u\}\, du \\ = & F_{Q_\alpha^*}(x). \end{aligned}$$

The interchange of limit and integration follows from the Dominated Convergence Theorem. We conclude that

$$[(1-\theta)\pi n]^{-\alpha/2} Q_1 \stackrel{\mathcal{D}}{\to} Q_\alpha^*. \quad (11)$$

We now turn to the interference $I = \sum_{i=2}^{n_S} Q_i$. Conditional on $V_{r(1)} = u$, we observe that for $i \neq 1$, $Q_i's$ are i.i.d. and have the same distribution as the Z_i's conditional on $U_1 = u$. Hence, the distribution of Q_i conditional on $V_{r(1)} = u$ has the same tail as given in (7). From the theory of stable random variables [2, pp.448, Theorem 2], it follows that, conditional on $V_{r(1)} = u$,

$$\left[\pi\Gamma\left(1-\frac{2}{\alpha}\right) n_S\right]^{-\alpha/2} I = \left[\pi\Gamma\left(1-\frac{2}{\alpha}\right)\theta n\right]^{-\alpha/2} I \stackrel{\mathcal{D}}{\to} I_\alpha^*, \quad (12)$$

where I_α^* has the stable distribution with characteristic exponent $\frac{2}{\alpha}$, and does not depend on u.

Again, the asymptotic limit above depends only on the tail of the conditional distributions of the individual Z_i's, which does not depend on u. Using a similar argument as above for Q_1, we conclude that (12) in fact holds unconditionally.

Finally, we claim that the signal power Q_1 and the total interference I are asymptotically independent (although they are in general not independent for finite n.). The argument is as follows. Eqn. (12) implies that the total interference I is asymptotically independent of $V_{r(1)}$, since the limiting distribution of I conditional on $V_{r(1)} = u$ does not depend on u. Note also that conditional on $V_{r(1)}$, U_1 and I are independent. Hence, in fact, I is asymptotically independent of the pair $(U_1, V_{r(1)})$. But the signal power Q_1 is a continuous function of U_1 and $V_{r(1)}$, and hence by the Continuous Mapping Theorem, I and Q_1 are asymptotically independent.

Combining this last fact with (11) and (12), we get the result on the probability of successful transmission from node 1 to node $r(1)$:

$$\Pr\{\mathrm{SIR} > \beta\} = \Pr\left\{\frac{Q_1}{N_0 + \frac{1}{L}I} > \beta\right\} \to \Pr\left\{\frac{Q_\alpha^*}{I_\alpha^*} > \beta^*\right\} > 0, \quad (13)$$

where

$$\beta^* = \frac{\beta}{L}\left[\frac{\theta}{1-\theta}\Gamma\left(1-\frac{2}{\alpha}\right)\right]^{\alpha/2}, \quad (14)$$

where $\Gamma(s) = \int_0^\infty x^{s-1}e^{-x}dx$ is the standard gamma function. The last inequality follows from the fact that Q_α^* and I_α^* can be chosen independent and Q_α^* has infinite support.

Therefore, as there are $n_S = \theta n$ senders attempting to transmit, the expected number of feasible sender-receiver pairs is $\mathrm{E}[N_t] = \theta n \cdot \Pr\{\mathrm{SIR} > \beta\}$, i.e., $\phi = \theta \cdot \Pr\{\mathrm{SIR} > \beta\}$. Furthermore, as π only depends on node locations, and as the node locations $\{X_i\}$ are i.i.d., the probability of success of any specific sender-receiver pair is equal, and thus $O(1/n)$. This completes the proof.

■

The essence of the proof of Theorem 3.4, and the fundamental reason why we can have $O(n)$ concurrent nearest neighbor transmission, is the fact that the received power at the nearest neighbor is of the same order as the total interference from $O(n)$ number of interferers. A similar phenomenon has been observed in [5] in the cellular setting, where they have shown that, provided $\alpha > 2$, the capture probability of the nearest transmitter to the basestation does not go to zero as the number of interferers become large. Although these results may seem surprising on first sight, they are all based on the following property: if $W_1, \ldots, W_n$ are i.i.d. random variables such that the cdf $F(w)$ decays slower than w^{-1} as $w \to \infty$, then the largest of them is of the same order as the sum. In the context of our problem, W_i's are the received powers from the transmitting nodes.

The technical complication in the proof of Theorem 3.4 is due to the fact that both the distribution of the received power from the sender and the distribution of the interference depends on the location of the receiver. This is primarily due to the edge effects of the disk, and this dependency would not be present if for example the nodes are randomly located on the surface of a sphere. Fortunately, in the regime we are interested in, the asymptotic distributions depends only on what happens in the local neighborhood around the receiver, and this is independent of where the receiver is in the open disk.

3.4 Sender-Centric vs. Receiver-Centric Approach

In the proof of Theorem 3.4, we have used a *sender-centric* approach, in that it is the senders that select the closest receiver to send to. We could also have considered a *receiver-centric* approach, where each receiver selects the closest sender from which to receive. It might seem that the situation is symmetric, and that a similar proof would carry through to arrive at the same result. However, this is not the case.

In the sender-centric approach, several senders may select the same receiver. This is not problematic from a technical point of view. By analogy, in the receiver-centric approach, it is possible that several receivers select the same sender. We can either assume that a sender is indeed able to generate signals for several receivers, or we can assume that the sender has to select only one receiver to which to send to. Both assumptions lead to difficulties in an analogous proof. Under the former assumption, we have to account for the elimination of sender-receiver pairs because the sender has to be unique; simple worst-case bounds can be found, but turn out to be too crude to improve upon the sender-centric capacity. Under the latter assumption, we have to account for the fact that a single sender can generate several unit-power interference signals (or analogously, the fact that the desired signal is only a fraction of unit power). We have not found an elegant way to integrate these complications into the above proof.

However, note that the receiver-centric approach is preferable in terms of the signal-to-interference ratio for a single receiver. The reason is that in the receiver-centric approach, the signal from the selected sender is always the strongest. If $\{Q_i\}$ are the received powers from the n_S senders, then the received signal power is $\max(Q_i)$, while the remaining $n_S - 1$ signals are interference. On the other hand, in the sender-centric approach used in our proof, the designated receiver is selected as the maximum of an *independent* set $\{Z_i\}$ of n_R random variables, where Z_i has identical distribution as Q_i[1]. The received signal power is $\max(Z_i)$, and the interference power is $\sum Q_i$ (where the sum is over $n_S - 1$ terms).

Let us assume first that $\theta = 1/2$, i.e., $n_S = n_R = n/2$. The power of the received signal is the maximum of n_S i.i.d. random variables in both cases; hence, they are distributed equally. However, the interference in the receiver-centric case is stochastically smaller than in the sender-centric case: in the former, the interference is the sum of $n_S - 1$ random signal powers, whereas in the latter, it is the sum of n_S random signal powers *minus the strongest of these signals*. Therefore, the SIR for the receiver-centric approach is smaller on average than in the sender-centric approach. Note that the probability of capture $\Pr\{\text{SIR} > \beta\}$ for a single receiver decreases with increasing θ for the sender-centric approach, while it does not depend on θ in the receiver-centric approach. Hence, the relative advantage of the receiver-centric approach holds at least for $\theta \geq 1/2$.

4 Numerical Results

We have examined the throughput capacity both through numerical evaluation of the asymptotic probability of capture developed in the preceding section, and through simulation of random network topologies.

We have evaluated the asymptotic fraction of feasible pairs ϕ for the special case $\alpha = 4$, because for this case, the normalized interference I_α^* has Lévy distribution[2] [7], with cdf

$$F_{I_\alpha^*}(x) = 2\left[1 - Q\left(\sqrt{\frac{\sigma}{x}}\right)\right], \qquad (15)$$

where $Q(.)$ is the standard Gaussian cdf, with $\sigma = 1/2$[3]. It is therefore straightforward to numerically evaluate (13) through Monte-Carlo simulation.

We have compared the fraction of feasible pairs ϕ for $\beta =$ 6dB and $L = 1$ predicted by our model with simulations based on $n = 1000$ nodes (cf. Fig. 3). The simulation results are averaged over 20 random topologies. Figure 4 shows the simulated normalized throughput for $\alpha = 2$, 3, and 4, and the throughput predicted by the asymptotic model for $\alpha = 4$. There is very good agreement between the analytical model and simulation results.

It is evident from the figure that given α, there exists an optimal sender density θ that maximizes the throughput. If θ is too small, then we do not exploit the potential for spatial channel reuse. If θ is too large, then the interference power becomes too dominant. The optimal θ obviously depends on α. For small α, interference limits the spatial channel reuse. Hence, the sender density has to be small.

[1] Ignoring edge effects.

[2] There is no closed form for the distribution or density function of I_α^* for general α; only the Laplace transform of its density is known explicitly [2, 7], and is given by $\psi_{I_\alpha^*}(s) = \exp(-s^{2/\alpha})$.

[3] This can be seen by comparing the Laplace transform of the density of non-negative strictly stable random variables in [2, page 448] with the expression for the characteristic function of general stable random variables in [7, page 5].

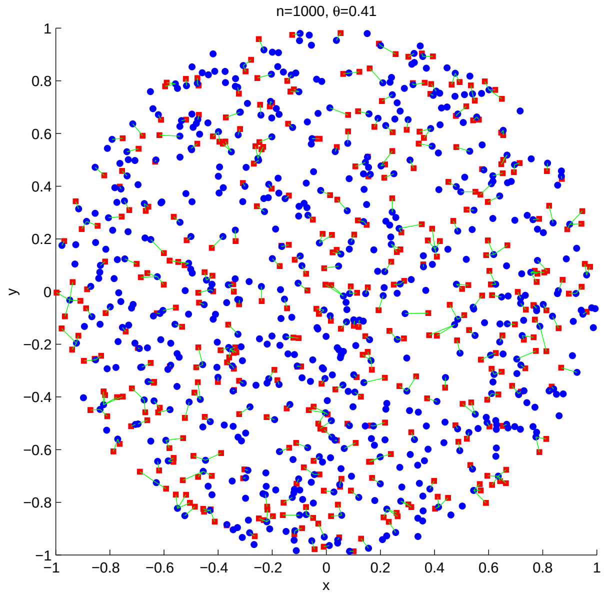

Figure 3: An example of a random topology with $n = 1000$ nodes, for sender density $\theta = 0.41$. Senders are depicted as red squares, receivers as blue circles. A line connects each sender to its closest receiver.

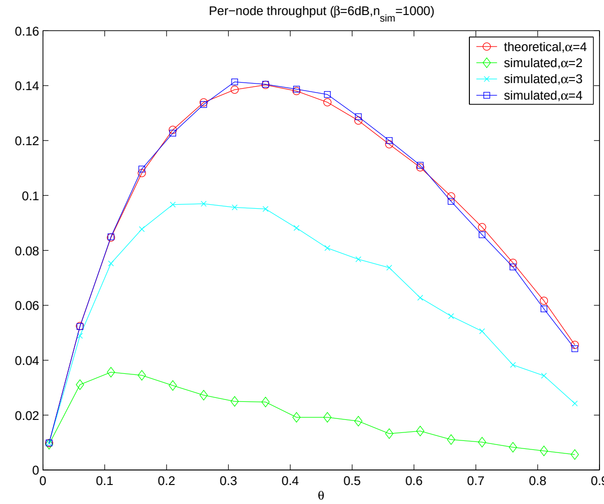

Figure 4: The normalized per-node throughput, as a function of the sender density θ, for different values of α. For $\alpha = 4$, the throughput predicted by the model is also shown.

For large α, interference is more localized, and the optimal θ and the maximum throughput are larger.

5 Discussion

5.1 Distributed Implementation

Although in our problem formulation, we allow central coordinated scheduling, relaying and routing, it should be noted that the algorithm obtained above can be implemented in completely distributed manner. At each time instant, each node can randomly and independently decide whether it wants to participate in phase 1 or phase 2, and whether it wants to be a sender or a potential receiver. Each sender then seeks out a potential receiver nearest to it, and attempts to send data to it. The access is uncoordinated; in fact, multiple senders may attempt to transmit to the same receiver. Whether a sender is "captured" is a random event, much like standard MAC random access protocols. What our analysis showed is that the probability of success is reasonable even in a network with many users.

Note that the two-phased algorithm used in the proof was chosen for mathematical convenience. As the capacity in both phases is identical, the expected delay experienced by a packet from source to destination would actually be infinite even for a finite number of nodes n. It is straightforward to fix this problem, e.g., by allowing both S-R and R-D transmissions to occur concurrently, but giving absolute priority to R-D (phase 2) transmission in all scheduled sender-receiver pairs. A detailed study of local scheduling strategies and their impact on end-to-end delay is the subject of future work.

5.2 Related Work

Frenkiel *et al.*, in [3], have introduced the concept of an *Infostation* for delay-insensitive data applications. An Infostation is a high-speed wireless basestation that does away with ubiquitous coverage. The motivation is that if delay is unimportant, then capacity is maximized by using the entire transmit power budget when the user is close to the basestation, and no power when the user is far away. While the work on Infostations has motivated us to look at delay-insensitive applications, our focus is different. We are not chiefly concerned about maximizing the capacity of a point-to-point channel subject to a power budget, but instead we are interested in the degree of spatial reuse in an entire network of mobile nodes that all desire to communicate, and we propose a new type of multiuser diversity to achieve high overall capacity.

Hajek, Krishna and LaMaire have studied a related interference model [5]. They examine the probability of capture for a single receiver for an asymptotically large number of senders. They show that in the limit, this probability only depends on the roll-off exponent α, but not on other channel effects such as fading and shadowing. While their results are not directly applicable to our setting, we nevertheless believe that this robustness property to other channel effects carries over.

In their numerical experiments, they report that for $\alpha = 4$ and $\beta = 6$dB, the asymptotic probability of capture is approximately 0.319. In our setting, as can be seen from Fig. 4, the probability of capture is approximately 0.14. The reason for this difference is that we study a sender-centric scheduling policy, as discussed earlier in Section 3.4, while Hajek *et al.* study the probability of capture of a single receiver.

Finally, as discussed earlier, Gupta and Kumar [4] have studied the capacity of random, but fixed ad-hoc networks. Their main insights are that communication must be limited to "near" neighbors in order to permit dense spatial channel reuse. However, the throughput per session goes to zero as the number of nodes n scales up, because each session requires on the order of $\sqrt{n}$ hops. A similar problem setup has also been considered by Shepard [8].

6 Conclusion

In this paper, we have examined the asymptotic throughput capacity of wireless ad-hoc networks. We have identified a new type of multiuser diversity that exploits long time-scale fluctuations in the fading process due to node mobility. Our results show that direct communication between sources and destinations is not sufficient to exploit this diversity, because they are too far apart most of the time. We propose to spread traffic randomly to intermediate relay nodes to take advantage of additional "routes" between a source and a destination. A single relay node is sufficient to use the entire throughput capacity of the network within the limits imposed by the interference model. This explains the dramatic performance improvement over a fixed ad-hoc network, where $O(\sqrt{n})$ intermediate relay nodes are necessary.

A key assumption underlying this result is the complete mixing of the trajectories of the nodes in the network, so that every node can get close to any other node. In practice, this assumption may not hold, or the delay to wait for such events to happen may be too long. As such, the result should be viewed as a theoretical one. What the theory does suggest is that there is ample opportunity to trade off delay and throughput in mobile wireless networks. The result of this paper can be considered as an extreme point in the tradeoff.

The ideas in this paper are not relevant to real-time applications such as voice communications. However, wireless data services are expected to grow quickly over the next few years. A subset of these services, such as email and database synchronization, do indeed possess very loose delay constraints (on the order of hours). Also, wireless devices are bound to become smaller and more pervasive in the future; they will not only be carried by humans, but integrated into physical objects (such as cars, electrical appliances, etc.) It is unlikely that the density of base-stations will keep pace, due to regulatory and environmental hurdles in deploying them. Thus, there is a clear opportunity for wireless ad-hoc networks to extend the reach of wireless communication. Our results suggest that delay-tolerant applications can take advantage of node mobility to significantly increase the throughput capacity of such networks.

References

[1] H. A. David. *Order Statistics (2nd ed.)*. Wiley, New York, 1981.

[2] William Feller. *An Introduction to Probability Theory and Its Applications (Volume II, 2nd ed.)*. John Wiley & Sons., New York, 1971.

[3] Richard Frenkiel, B. R. Badrinath, Joan Borras, and Roy Yates. The Infostations challenge: Balancing cost and ubiquity in delivering wireless data. *Preprint*, August 1999.

[4] Piyush Gupta and P. R. Kumar. The Capacity of Wireless Networks. *IEEE Transactions on Information Theory*, 46(2):388–404, March 2000.

[5] B. Hajek, A. Krishna, and R.O. LaMaire. On the capture probability for a large number of stations. *IEEE Transactions on Communications*, 45(2):254–260, February 1997.

[6] R. Knopp and P. A. Humblet. Information Capacity and power Control in Single-Cell Multiuser Communications. In *Proc. Int. Conf. on Communications*, Seattle, June 1995.

[7] Gennady Samorodnitsky and Murad S. Taqqu. *Stable Non-Gaussian Random Processes*. Chapman & Hall, 1994.

[8] Timothy J. Shepard. A Channel Access Scheme for Large Dense Packet Radio Networks. In *Proc. ACM SIGCOMM '96*, San Francisco, Calif., August 1996.