Capacity Bounds for Large Wireless Networks under Fading and Node Mobility

S. Toumpis
Forschungszentrum Telekommunikation Wien (ftw.)
Donau-City-Straße, 1/3 Stock.
A-1220 Wien, Austria
toumpis@ftw.at

A. J. Goldsmith
Department of Electrical Engineering
Stanford University
Stanford, CA, 94305
andrea@ee.stanford.edu

Abstract

We establish lower bounds on the capacity of wireless ad hoc networks, which hold with probability approaching unity as the number of nodes n approaches infinity.

We first focus on networks with n immobile nodes, developing a scheme that can achieve an aggregate throughput that grows with n as $n^{\frac{1}{2}-\epsilon}$, for any $\epsilon > 0$, and under a general model of fading.

We then extend our formulation to study the effects of mobility. We develop a scheme that can achieve an aggregate throughput that grows as $n^{\frac{1}{2}-\epsilon}$, for any $\epsilon > 0$, while respecting a constant upper bound on the packet delay which does not depend on n. We then identify a fundamental throughput-delay tradeoff in mobile ad hoc networks. In particular, provided we tolerate packet delays that are upper bounded by n^d, where $0 < d < 1$, the same scheme can achieve an aggregate throughput that grows as $n^{\frac{1+d}{2}-\epsilon}$, for any $\epsilon > 0$. A general model of fading is assumed. Nodes require no global topology or routing information, and only need to coordinate locally.

Keywords: Ad Hoc Network, Capacity, Delay, Fading, Mobility, Throughput, Wireless Communication.

I. Introduction

Ad hoc wireless networks consist of collections of mobile nodes communicating over a wireless channel. Contrary to cellular networks, where the nodes are restricted to communicate with a few strategically placed base stations, in wireless ad hoc networks any two nodes are allowed to communicate directly. However, because of the nature of the wireless channel, each node can effectively communicate with only some of the others, that typically lie in its vicinity. On the other hand, the traffic requirements are taken to be arbitrary, therefore it is necessary that nodes cooperate to forward packets to their final destinations.

The problem of determining fundamental limits on the performance of ad hoc networks has only recently attracted the interest of researchers [1], [2], [3]. In a landmark paper, the authors of [1] investigate the asymptotic behavior of the capacity of a class of two-dimensional random networks as the number of nodes n approaches infinity, under a uniform traffic assumption. Here, nodes are assumed immobile. The authors present a scheme that achieves **with high probability (*w.h.p.*)**, i.e., with probability approaching 1 as n approaches infinity, a communication rate equal to $(n \log n)^{-\frac{1}{2}}$, up to a multiplicative constant, from each node to its randomly chosen destination. The authors also show that, with high probability, the n nodes cannot send data to their destinations with a per-node communication rate greater than $(n \log n)^{-\frac{1}{2}}$, up to a (different) multiplicative constant.

The last result is disheartening since it suggests that, as the number of nodes n goes to infinity, the per-node communication rate will necessarily go to zero. However, node mobility can lead to dramatic improvements: In [2] the authors concentrate on a wireless network with n mobile nodes. They show that, with high probability, in the absence of any constraint on

Work supported by the Office of Naval Research (ONR) under grant numbers N00014-99-1-0698 and N00014-02-1-0003, the Stanford Networking Research Center, and **K***plus* funding for the ftw. project I0 "Signal and Information Processing".

the delay in the delivery of a packet, each node is guaranteed a fixed rate of communication to its destination, which is not a function of the number of nodes n. The downside of this result is that, as the number of nodes increases, so will the expected packet delay.

In this paper, we continue the investigation along the lines of [1] and [2]. In Section II we specify our model for wireless ad hoc networks, we formally define their capacity, and we introduce the concept of the order of a network. In Section III we introduce notation and a couple of simple technical lemmas. In Section IV we start by concentrating on networks with immobile nodes and with no fading. In Section V we extend our formulation to include fading. In Section VI we present a throughput-delay tradeoff for the case of networks with mobile nodes, and in the presence of flat fading. We conclude in Section VII. Throughout the text, terms being defined are set in **boldface**. Some of the proofs are omitted, but all proofs appear in [4].

II. Network Model, Capacity and Order

A. Network model

We consider a collection of n immobile nodes $X_1, X_2, \ldots, X_n$, placed randomly, uniformly and independently, in the two-dimensional area $\{(x, y) : -\frac{1}{2} \leq |x|, |y| \leq \frac{1}{2}\}$.

Regarding the traffic model, we assume that each node is the source of a single data stream, and the destination of a single data stream. A node cannot be the source and destination of the *same* stream. Apart from this restriction, all other combinations of sources and destinations are equally probable. Alternative traffic models can easily be incorporated in our formulation and are studied in [4].

Nodes communicate over a wireless channel of bandwidth W. Half-duplex transmission is assumed, i.e., nodes cannot transmit and receive simultaneously. Each node can transmit with any power P_i, provided it is less than the maximum P_0. When node X_i transmits with power P_i, node X_j receives the transmitted signal with power $G_{ij}P_i$, where $G_{ij} = Kd_{ij}^{-\alpha}$. K is a constant, the same for all nodes, d_{ij} is the distance between nodes X_i and X_j, and $\alpha > 2$ is the **decay exponent**. For now, we do not incorporate fading in our model.

Let $\{X_t : t \in \mathcal{T}\}$ be the set of transmitting nodes at a given time, each node X_t transmitting with power P_t. Let us assume that node X_j, $j \notin \mathcal{T}$ is receiving information from X_i, $i \in \mathcal{T}$. Then the **signal to interference and noise ratio (SINR)** at node X_j will be

$$\gamma_j = \frac{G_{ij}P_i}{\eta + \sum_{k \in \mathcal{T},\ k \neq i} G_{kj}P_k},$$

where η is the thermal noise power at the receiver, which is assumed the same for all nodes. We assume that the transmission of the packet will be successful if and only if the transmission rate used, R_j, satisfies the inequality

$$R_j \leq f_R(\gamma_j) \equiv W \log_2(1 + \frac{1}{\Gamma}\gamma_j) \tag{1}$$

where $\log_2(x)$ denotes the logarithm of x, in base 2. With $\Gamma = 1$, the receiver achieves Shannon's capacity. With $\Gamma > 1$, (1) approximates the maximum rate that meets a given BER requirement under a specific modulation and coding scheme such as coded MQAM [5].

B. Capacity and order

Let there be a communication scheme (that is consistent with the assumptions on the topology, the channel, and the transceivers) under which each source can send traffic to its destination with an end-to-end rate equal to λ. We will say that the rate λ is **uniformly achievable**. We define the **capacity C of the network** to be the supremum of all uniformly achievable rates, multiplied by the number of nodes n. Therefore, the capacity is the supremum of all *aggregate* throughputs that can be carried by the network. Note that, since there is randomness in the creation of the network, the capacity is a random variable. Let $C(n)$

be the sequence of capacities of the network versus the number of nodes n. $C(n)$ may be thought of as a random process.

We are now ready to define the order of a network. Let O be the set of real numbers o for which the following holds:

$$\lim_{n\to\infty} P[C(n) \geq n^o] = 1. \tag{2}$$

The **order** of the network is defined as the supremum of O. In other words, the order is the supremum of all powers of n with which the aggregate throughput can increase, with probability approaching 1. We do not define the order as the maximum of O, as this may not exist. (For example this is the case when O is an open set.) On the other hand, the supremum of a set always exists. (By convention, the supremum of the empty set $\emptyset$ is $-\infty$.) Using the supremum also yields another important gain: to establish a lower bound b on the order, we do not need to come up with a scheme that achieves the lower bound, i.e., its aggregate throughput increases as n^b. Rather, it suffices to exhibit, for all $\epsilon > 0$, a scheme whose aggregate throughput increases as $n^{b-\epsilon}$.

Note that the definitions of the capacity and the order are quite general, and may *in principle* be applied to a broad range of networks, even to those that are not wireless.

III. Notation and Technical Lemmas

We will use the symbols $<_a$, $>_a$, $\leq_a$, $\geq_a$ to denote that the corresponding inequality will only hold **asymptotically**, i.e., for sufficiently large n. For example, $f(n) <_a g(n)$ means that there is a n_0 such that $f(n) < g(n)$ for all $n > n_0$.

Whenever the parameter k appears in a relation without having previously been defined, it will be assumed that the relation will hold for some real and positive k. The same will hold for all parameters of the form k_i.

Unless specified otherwise, all limits we write will be for $n \to \infty$. We will say that $f(n)$ approaches a fixed limit L **exponentially fast with rate** r if $|f(n) - L| \leq_a \exp(-kn^r)$. We then write $f(n) \to_r L$.

Following [1], we say that an event sequence $\{A_n\}$ occurs **with high probability (*w.h.p.*)** if $P[A_n] \to 1$. Note that if $\{A_n\}$ and $\{B_n\}$ occur *w.h.p.*, then $\{A_n \cap B_n\}$ also occurs *w.h.p.* The following lemma, whose proof is a straightforward application of the union bound, shows that the intersection of even polynomially many events also occurs *w.h.p*, provided the probability of these events goes to 1 exponentially fast:

Lemma 1: Let $\{A_{nm}\}$, where $n = 1, \ldots$ and $1 \leq m \leq M(n)$, be a collection of events for which $P[A_{nm}] = P[A_{n1}]$ for all $m = 1, \ldots, M(n)$, and $P[A_{n1}] \to_r 1$. Also let $M(n) \leq_a n^p$. Then $P[\cap_{m=1}^{M(n)} A_{nm}] \to_r 1$.

This result will be used repeatedly in the following, sometimes without explicit mention. Event sequences that go to 1 (or 0) exponentially fast often appear in relation to large deviations of random variables. An example appears in the following lemma, which is a direct application of Chernoff's bounds [6] (alternatively Sanov's theorem [7]):

Lemma 2: Let $B(n)$ be a sequence of binomially distributed random variables, with number of attempts $a(n)$ and probability of success (per attempt) $p(n)$, such that $0 <_a p(n) <_a 1$ and $a(n)p(n) \geq_a kn^c$, where $c > 0$. Let $\beta > 1$. Then $P[\frac{1}{\beta}a(n)p(n) < B(n) < \beta a(n)p(n)] \to_c 1$.

IV. A Constructive Lower Bound on the Order

A. The cell lattice

We denote by $\lfloor\lfloor x \rfloor\rfloor$ the greatest odd multiple of 3 that is less than or equal to x. Let $0 < b < 1$ and $\lfloor\lfloor n^{\frac{b}{2}} \rfloor\rfloor = 2r + 1$. We divide the space occupied by the network into a regular lattice of $g(n) = (2r+1)^2$ cells. The boundaries of the cells are formed by the lines

$$x = -\frac{1}{2} + \frac{i}{2r+1},\ i = 0, \ldots, 2r+1, \quad y = -\frac{1}{2} + \frac{j}{2r+1},\ j = 0, \ldots, 2r+1.$$

Points on the boundary lines are assigned arbitrarily to one of the cells forming the boundary. It is straightforward to show that

$$\forall \epsilon > 0,\ (1-\epsilon)n^b <_a g(n) \leq n^b,$$

We denote the cells by $c_1,\ c_2,\ \ldots, c_{g(n)}$. Each cell can be identified by its coordinates (v_1, v_2) in the lattice, where $-r \leq v_1,\ v_2 \leq r$.

Let m_i be the number of nodes in cell c_i. Since $b < 1$, the number of cells increases *polynomially* slower than the number of nodes. This has the very nice implication that the nodes become uniformly distributed in the cells exponentially fast:

Lemma 3: For all $\beta > 1$, $P[\frac{1}{\beta}n^{1-b} < m_i < \beta n^{1-b}\ \forall i] \rightarrow_{(1-b)} 1$.

Proof: Pick $\beta > 1$, and select $\epsilon > 0$ such that $\beta_1 = \beta(1-\epsilon) > 1$. The number of nodes m_i that appear in a particular cell c_i follows the binomial distribution, with a number of tries equal to $a(n) = n$ and probability of success $p(n) = \frac{1}{g(n)}$. Since $a(n)p(n) = \frac{n}{g(n)} \geq n^{1-b}$, Lemma 2 applies. Therefore, $P[\frac{1}{\beta_1}a(n)p(n) < m_i < \beta_1 a(n)p(n)] \rightarrow_{(1-b)} 1$. Since $\beta_1 a(n)p(n) <_a \beta n^{1-b}$ and $\frac{1}{\beta_1}a(n)p(n) > \frac{1}{\beta}n^{1-b}$, we have that $P[\frac{1}{\beta}n^{1-b} < m_i < \beta n^{1-b}] \rightarrow_{(1-b)} 1$. Noting that the number of cells grows only polynomially fast, the result follows by Lemma 1. □

B. The routing rules

By Lemma 3, all cells are guaranteed to have at least one node (in fact many more) in the limit of a large number of nodes. We are therefore justified to define the following rules that govern the routing of data to their final destination:

(i) Direct transmission is allowed only between nodes that lie in the same cell, or in neighboring cells (cells are called **neighbors** if they share a common edge, so that each cell has at most four neighbors).

(ii) Nodes that do not lie in the same or neighboring cells communicate by using nodes in intermediate cells as relays. The message is first transmitted along cells whose x-coordinate is the same as the x-coordinate of the source cell, until it arrives at a cell whose y-coordinate is the same as the y-coordinate of the destination cell. Then, the message is transmitted along cells whose y-coordinate is the same as the y-coordinate of the destination cell. In each of the intermediate cells, one of the nodes in the cell is arbitrarily selected to act as the relay of the packet.

Since the streams that are routed through a particular cell create load for the nodes in that cell, it is important to have an estimation of their number. We have the following lemma:

Lemma 4: Let s_i *be the number of streams arriving at (and possibly going through) cell* c_i. *Then* $P[s_i < 3n^{1-\frac{b}{2}}\ \forall i] \rightarrow_{(1-\frac{b}{2})} 1$.

C. The time division

By the construction of the cell lattice, if n is the number of nodes in the network, there will be $g(n) = \lfloor\lfloor n^{\frac{b}{2}}\rfloor\rfloor^2 = (2r+1)^2$ cells, $g(n)$ being a multiple of 9. Therefore, we may divide the $g(n)$ cells perfectly into nine sub-lattices. We index the nine sub-lattices by the pairs (i, j), where $-1 \leq i, j \leq 1$. The cells belonging to sub-lattice (i, j) are all those whose coordinates are $(i + 3k_1, j + 3k_2)$, for some $k_1, k_2 \in \mathbf{Z}$. In Fig. 1(a) we have shaded the cells belonging to one of the 9 sub-lattices.

We divide time into frames, and each frame into nine slots. Each slot corresponds to a sub-lattice. At any time during that slot, only one node from each cell of the sub-lattice is allowed to *receive* (but many nodes in that cell may receive consecutively). As specified by the routing protocol, the transmitter of that transmission will have to lie in the same cell, or in one of the four neighboring cells. The transmitter will be transmitting with the maximum power P_0.

We will clearly need a lower bound on the Signal to Interference and Noise Ratio (SINR) at each receiver. Intuitively, such a bound exists, as the time division scheme spaces out the interferers. In fact, we can prove the following:

Lemma 5: In the absence of fading, the SINR γ_i at any node X_i, where $i = 1, \ldots, n$, is asymptotically lower bounded by a fixed constant γ_{min}, which is not a function of n.

D. A lower bound on the order

We will refer to the algorithm specified in Sections IV-A to IV-C as the **basic scheme**.

Theorem 1: The basic scheme achieves, with probability approaching 1 as $n \to 1$, an aggregate throughput $T(n) = k_1 n^{\frac{b}{2}}$ for any $b < 1$. Therefore, the order o of the wireless ad hoc network is lower bounded by $o \geq \frac{1}{2}$.

Proof: By Lemma 5, each receiver is asymptotically guaranteed a reception rate of at least $R_{min} = f_R(\gamma_{min}) = W \log_2(1 + \gamma_{min})$. In addition, by Lemma 4 each cell will need to serve less than $3n^{1-\frac{b}{2}}$ streams with high probability. By Lemma 3, there will be, with high probability, a node in every cell to forward the packets of these streams. Noting that, due to the time division scheme, each cell will be able to receive packets during only one out of nine slots, each stream is guaranteed, *w.h.p.*, a rate equal to $\frac{W \log_2(1+\gamma_{min})}{27 n^{1-\frac{b}{2}}}$. Multiplying by n we see that an aggregate rate $T(n) = \frac{W}{27} \log_2(1 + \gamma_{min}) n^{\frac{b}{2}}$ is asymptotically achievable. The bound on the order follows trivially. □

V. FADING

Until now, we have assumed that the transmissions are not subject to any type of fading, so that the received power is a deterministic function of the distance between the transmitter and the receiver. Here, we introduce flat fading, and show that the scheme introduced in Section IV can be modified so that its aggregate throughput is reduced by a factor smaller than n^{ϵ}, for any $\epsilon > 0$.

A. Fading model

We assume that when node X_i transmits with power P_i, node X_j receives the transmitted signal with power $G_{ij} P_i$, where now $G_{ij} = K d_{ij}^{-\alpha} f_{ij}$. The extra factor f_{ij} is the **fading coefficient**, a non-negative random variable that models fading, and does not change with time. We assume that $E[f_{ij}] = 1$, and that $f_{ij} = f_{ji}$. We take the distinct $\frac{n(n-1)}{2}$ fading coefficients to be independent and identically distributed (iid). We also assume that their complementary cumulative distribution function $F^c(x)$ has a thin, exponentially decaying tail. Formally:

$$F^c(x) \equiv P[f_{ij} > x] \leq \exp[-qx] \ \forall x > x_1, \tag{3}$$

for some real and positive parameters $q, \ x_1$. In addition, we make the very mild assumption that there is a $f_{median} > 0$ such that $P[f_{ij} \geq f_{median}] \geq \frac{1}{2}$. Both of these assumptions are satisfied by most distributions used to model fading. In fact, they are satisfied by the Nakagami, Rayleigh, and Ricean distributions [4].

B. A lower bound on the order

Under the basic scheme, each node with a packet waiting to be relayed through a neighboring cell will arbitrarily pick one of the nodes in that cell as the relay. In the absence of fading, the choice of relay is not critical. With fading, the choice is clearly important, as the power gains from one node to various nodes of the neighboring cell may be drastically different. Fortunately, by Lemma 3 there are many potential receivers for a transmitter to choose from.

We therefore use the basic scheme with a single modification on the routing rules. Specifically, when it is the turn of a node to transmit, the transmitter will pick as the receiver any of the nodes in the target cell, among those whose link with the transmitter has a fading coefficient greater than or equal to the median of the distribution, f_{median}. (By the definition of the median, the probability that a node in the cell will satisfy this requirement is at least $\frac{1}{2}$.) If such node does not exist, the packet is dropped. This rule is extremely intuitive:

nodes plainly avoid transmitting to nodes with poor fading coefficient. There is, however, a complication. Once a packet arrives to its destination *cell*, it may not find itself at its destination *node*. This can easily be amended as follows: Let the packet be in node X_1, and have a destination node X_2, both being in the same cell c_i. The packet will be transmitted two more times, once to an intermediate node X_3, and then from the intermediate node to the destination node X_2. The intermediate node is chosen among the nodes in cell c_i, such that the fading coefficients of both links are greater than the median of the fading distribution. If no such node exists, the packet is dropped. We will refer to this algorithm as the **spatial-diversity scheme**.

For a given number of nodes n, there is a positive probability that the packets of some of the streams will have to be dropped, since at some point along the way no node may be found to satisfy the restriction we place on the fading coefficients. However, as the number of nodes increases, this probability goes to zero, as we now show:

Let $\beta > 1$, and let p_{if}^{Uh} be the probability that the packets of stream i will be dropped, conditioned on the event that the source is h cells away from the destination (i.e., $h+2$ hops away, counting the extra 2 hops in the last cell), and on the event $U = \{\frac{1}{\beta}n^{1-b} < m_i < \beta n^{1-b}\}$. Note that, by Lemma 3, $P[U] \rightarrow_{(1-b)} 1$. Then:

$$\begin{aligned} p_{if}^{Uh} &= \sum_{l=1}^{h+1} P(\text{packet is dropped right before } l\text{-th hop}) \\ &\leq h(\frac{1}{2})^{\frac{1}{\beta}n^{1-b}} + (\frac{3}{4})^{\frac{1}{\beta}n^{1-b}-2} \;\; \text{(using the conditioning on } U) \\ &\leq 2n^{\frac{b}{2}} \exp[-(\log 2)\frac{1}{\beta}n^{1-b}] + (\frac{4}{3})^2 \exp[-\log(\frac{4}{3})\frac{1}{\beta}n^{1-b}] \;\; \text{(since } h < 2n^{\frac{b}{2}}) \\ &\leq \exp[-k_1 n^{1-b}]. \end{aligned}$$

Noting that this upper bound does not depend on the number of hops, we can remove the conditioning on h. Since $P[U] \rightarrow_{(1-b)} 1$, we can also dispense with the conditioning on U: Let p_{if} be the probability that the packets of stream i will be dropped. Then:

$$p_{if} \leq P[U]\exp[-k_1 n^{1-b}] + (1 - P[U]) \leq_a \exp[-k_2 n^{1-b}].$$

Combining this result with the union bound, we have that the probability that the packets of *any* stream are dropped, p_f, goes to 0 exponentially fast:

$$p_f \leq np_{if} \leq n\exp[-k_2 n^{1-b}] \leq \exp[-\frac{k_2}{2}n^{1-b}] \rightarrow_{(1-b)} 0. \quad (4)$$

Equation (4) implies that, by taking advantage of the existing spatial diversity, all nodes will be using strong paths and no packet will be dropped with high probability. However, a bound on the interference experienced by the receivers is also required. Such a bound can be easily derived, by virtue of the exponentially thin tail of the fading distribution. Indeed, focusing on an arbitrary fading coefficient, we have, using (3), that $P[f_{ij} \leq n^\epsilon] \geq 1 - \exp[-qn^\epsilon]$. Noting that there are $\frac{n(n-1)}{2}$ fading coefficients (i.e, polynomially many), we can use Lemma 1 to arrive at:

$$P[\max\{f_{ij}\} \leq n^\epsilon] \rightarrow_\epsilon 1. \quad (5)$$

We now are now ready to state the main result of the section:

Theorem 2: In the presence of fading, the spatial-diversity scheme can achieve, with probability approaching 1 *as* $n \rightarrow \infty$, *an aggregate throughput* $T(n) = k_2 n^{\frac{b}{2}-\epsilon}$ *for any* $b < 1$ *and* $\epsilon > 0$. *Therefore, the order* o *is lower bounded by* $o \geq \frac{1}{2}$.

Proof: Pick an arbitrary $\epsilon > 0$. Let us concentrate on a receiving node X_i. Because of the restriction we place on which paths may be used, if S_i is the received power of the useful signal without fading and S_i^F is the received power with fading, we will have $S_i^F \geq f_{median} S_i$. In addition, by (5), and *w.h.p.*, all fading coefficients will be smaller than n^ϵ. Therefore, if I_i is the interference power at X_i with no fading and I_i^F is the interference power at X_i with fading, we will have that $I_i^F \leq n^\epsilon I_i$. These bounds will hold uniformly, i.e. for all receptions, and *w.h.p.*. Noting that the thermal noise power is not affected by the presence of fading, all nodes are guaranteed, *w.h.p.*, a minimum SINR γ_{min}^F equal to

$$\gamma_{min}^F(n) = \frac{f_{median}}{n^\epsilon}\ \gamma_{min}, \tag{6}$$

where γ_{min} is the minimum SINR in the absence of fading of Lemma 5. Therefore, *w.h.p.*, each receiver is guaranteed a rate $R_{min}^F(n) = f_R(\gamma_{min}^F(n)) = W \log_2(1 + \gamma_{min}^F(n))$.

By Theorem 4, a maximum of $3n^{1-\frac{b}{2}}$ streams will be arriving at each cell. For a few of these, specifically those whose destination lies in the cell, three receptions will be required, according to the routing rules. Therefore, a maximum of $3 \times (3n^{1-\frac{b}{2}})$ receptions will have to take place in each cell. Noting that each cell will be receiving in only one of every nine slots, we see that each stream is guaranteed, *w.h.p.*, a rate equal to $\frac{W \log_2(1+\gamma_{min}^F(n))}{81 n^{1-\frac{b}{2}}}$. Multiplying by the number of streams n, an aggregate rate $T(n) = \frac{W}{81} \log_2(1+\gamma_{min}^F(n)) n^{\frac{b}{2}}$ is asymptotically achievable. Substituting $\gamma_{min}^F(n)$ from (6), and using the limit $\lim_{x \to 0} \frac{\log_2(1+x)}{x} = \log_2(e)$ we arrive at the result. □

VI. Node Mobility

Before formally presenting the capacity results under node mobility, it is worthwhile discussing the basic idea behind them, which in fact is very simple: It was shown in [2] that a number of simultaneous transmissions on the order of n is possible, with a transmission rate that does not decrease with n, if all nodes transmit to their nearest neighbors. In addition, because of the mobility of the nodes, two transmissions are enough for a packet: once to a relay (that happens to be the nearest neighbor of the source) and once more to the final destination, whenever this happens to be the nearest neighbor of the relay. (One transmission would also be enough, but the number of source-destination pairs that are also nearest neighbors is not on the order of n.) Therefore, an aggregate throughput on the order of n is possible. However, under this scheme each packet will have to remain in its relaying node for a time increasing also like n, since the chance of the relay being the closest neighbor with the destination is $\frac{1}{n}$ (of course, to make this argument formal we need to specify the mobility model in detail). The motivation for our schemes is simple: If more nodes could receive the packet (and act as potential relays), rather than just one, maybe the packet would not have to wait that much time. But for more nodes to receive the packet, it is necessary that fewer nodes transmit, so that the transmitted signals experience less interference and reach further. However, this will reduce throughput. In other words, it is clear that, in principle, there can be a tradeoff between delay and throughput.

Fortunately, we already have developed the mathematical structure required to capture this tradeoff, in the form of the cell lattice: With the proper time division, there can be one transmission per cell, and all nodes in the cell will correctly decode it. If we decrease the number of cells (by decreasing b), there will be fewer transmissions (around n^b), but more recipients per transmission (around n^{1-b}). So the throughput will decrease but so will the average delay per packet, since more nodes will be acting as relays, and this is exactly the tradeoff we want to capture.

A. Network model

All the assumptions of Sections II and V continue to hold. However, nodes are no longer immobile, but rather move inside the square region according to a stationary and ergodic process that has the following properties:

(i) There is a duration of time s, such that within this time interval, all nodes remain immobile, so that the power gains of all links remain constant, even in the presence of fading.
(ii) There is a duration of time S, such that after the passing of an interval of length S, the positions of all nodes become perfectly reshuffled. In other words, after the lapse of an interval equal to S, the nodes are again randomly, uniformly and independently redistributed. For convenience, we take S to be a large integer multiple of s, i.e., $S = Ns$.

Finally, we assume that the whole communication period will last for a time interval of length equal to $2n^D S$, where D is an arbitrary integer, greater than unity.

B. Constant delay constraint

As in Section IV, we divide the area of the network into a regular lattice of $g(n) = \lfloor \lfloor n^{\frac{b}{2}} \rfloor \rfloor^2$ cells. Now, however, we require that $0 < b < \frac{1}{2}$. The cells are again divided into the nine sub-lattices (i, j) where $-1 \leq i, j \leq 1$.

As shown in Fig. VII(a), we divide time into identical frames, with each frame consisting of N mini-frames. Each mini-frame will have a duration equal to s, and will consist of 9 slots, each of duration $\frac{s}{9}$. In each of the slots, only nodes lying in a corresponding sub-lattice are allowed to *transmit* (and only with maximum power). The rest will have to remain silent.

In addition to this time division, which is on the 9 slots of each mini-frame, we superimpose another, distinctly different time-division, which is on the N mini-frames of each frame. Specifically, the N mini-frames within the frame are logically independent: packets that are transmitted, received, created, etc., in a given mini-frame are stored in the memory of the nodes for $N - 1$ mini-frames, and then brought back forward for the corresponding mini-frame of the next frame. We thus create N independent virtual channels, with the nodes communicating using each of them independently. As a result, the topology in each of these virtual channels becomes totally decorrelated every other mini-frame, as opposed to every N mini-frames.

In the following, we suppress all reference to the "outer" time division, and we concentrate on a single virtual channel. Therefore, our "working model" becomes that of Fig. VII(b). It should be understood, though, that the nodes will be executing the same algorithm independently, and concurrently, in each of the N virtual channels.

The algorithm executed in each of the N virtual channels is as follows: All nodes pick a common $\beta > 1$. In the first, third, and generally all odd mini-frames, nodes wait for the slot in which they are allowed to transmit (this depends on which sub-lattice the node lies). They then transmit a packet, with rate $R_{min}(n) = f_R(C^{\gamma}_{min}(n))$, where $\gamma^F_{min}(n)$ is given by (6), for a time duration equal to $\frac{s}{9\beta n^{1-b}}$. Nodes that lie in the same cell coordinate among themselves, and transmit their packets consecutively. If there are more than βn^{1-b} nodes in any cell, then some of the packets will not be transmitted and will be dropped. We call this a type-I error. The transmitted packets will possibly be received successfully by some of the other nodes lying in the same cell. These will act as relays in the subsequent mini-frame.

In the second, fourth, and generally all even-numbered mini-frames, each destination will wait for the slot in which nodes in its cell can transmit, and then will receive the packet intended for itself, by one of the relays lying in the same cell, that successfully received the packet in the previous mini-frame, and has a sufficiently strong path to the destination. We say that a type-II error occurs if there are more than βn^{1-b} nodes in a cell, so that some of them will not have the time to receive their packets. We say that a type-III error occurs when, for a specific packet, there is no relay with sufficiently strong paths between both itself and the source, and itself and the destination, so that the packet never makes it to the destination. At the end of each even-numbered mini-frame, nodes discard all packets remaining in their buffers. We will refer to this algorithm as the **mobility scheme**.

Regarding the packet delay, we assume that packets are created just before it is their node's turn to transmit. (Delays due to random arrival times and queuing are beyond the scope of this work – such issues are discussed in [8].) We then note that each packet will arrive at its destination within an interval $(N+1)s$ after its transmission from the source. Therefore, packet delays are smaller than a maximum $d_{max} = (N+1)s$. This bound does not depend on the number of nodes n, but rather on the underlying mobility model: it is slightly longer than the decorrelation time of the topology.

We are now ready to present our first main result on the capacity of networks with node mobility, whose proof appears in [4]:

Theorem 3: Under the mobility scheme, no errors of any type occur, with probability going to 1 *as the number of nodes* n *goes to infinity. In addition, the mobility scheme achieves an aggregate throughput equal to* $T(n) = k_3 n^{b-\epsilon}$*, for any* $b < \frac{1}{2}$ *and* $\epsilon > 0$*, with an upper bound on the acceptable packet delay equal to* $d_{max} = (N+1)s$*. Therefore, the order of the wireless ad hoc network, when there is an upper bound on the acceptable packet delay equal to* $d_{max} = (N+1)s$*, is lower bounded by* $o \geq \frac{1}{2}$*.*

Theorem 3 states that a non-trivial aggregate throughput, that increases roughly like $n^{\frac{1}{2}}$, is possible when nodes are mobile, even with a constant upper bound on the acceptable packet delay and without the nodes using global routing and topology information. This aggregate throughput happens to be the aggregate throughput that nodes would roughly achieve if they were using the spatial-diversity scheme, properly modified to handle node mobility, and assuming that they had access to global topology information. However, for large networks and high node mobilities, such information comes at a prohibitive cost, or is just impossible to acquire. But network designers need not despair: Theorem 3 states that the same aggregate throughput is achievable, with no need for routing or topology information, provided a packet delay roughly equal to the topology decorrelation time is tolerated. To summarize, the mobility scheme looks far more appropriate than traditional route discovery protocols, in large networks with high node mobility.

C. Polynomially increasing delay constraint

In Section VI-B it was shown that a non-trivial aggregate throughput is possible even with a constant upper bound on the delay. It is natural to ask how much better we can do if larger delays are tolerated, in particular delays that increase with the number of nodes. Looking back at the mobility scheme, it is clear that the aggregate throughput is limited by our requirement that the number of relays per packet (which is around n^{1-b}) be larger than the number of cells (which is around n^b). If this requirement is relaxed, packets may need to spend more than one slot at a relay, but since there are now more cells, and hence more simultaneous transmissions, the aggregate throughput increases. In fact, the following holds:

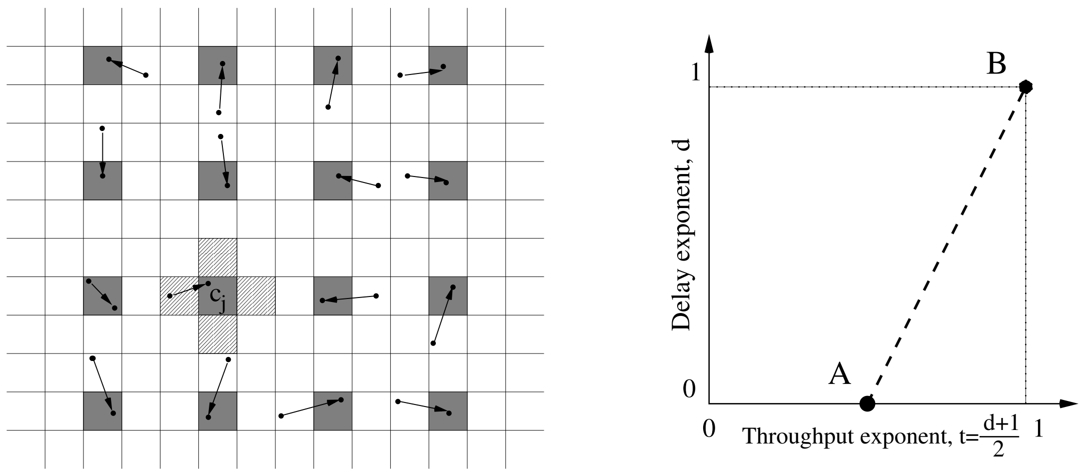
Theorem 4: The mobility scheme can achieve an aggregate throughput equal to $T(n) = k_4 n^{\frac{d+1}{2}-\epsilon}$ *for any* $\epsilon > 0$*, with an upper bound on the acceptable packet delay equal to* $d_{max} = (4Ns)n^d$*, that holds for all created packets,* w.h.p.. *Therefore, under this bound on the delay, the order of the wireless ad hoc network is lower bounded by* $o \geq \frac{1+d}{2}$*.*

Theorem 4 suggests a fundamental tradeoff between the acceptable packet delay and the achievable per-node throughput: If we are willing to tolerate larger delays, our throughput will increase. In particular, ignoring constant factors and arbitrarily small powers of n, if we tolerate delays on the order of n^d, where the **delay exponent** $d \in (0,1)$, we can achieve aggregate throughputs on the order of n^t, where the **throughput exponent** t is equal to $t = \frac{d+1}{2}$. Therefore, we can achieve any point on the open line interval AB of Fig. 1(b). Theorem 3 shows that point A is also achievable. Point B can be achieved by a scheme very similar to that of [2].

VII. CONCLUSIONS

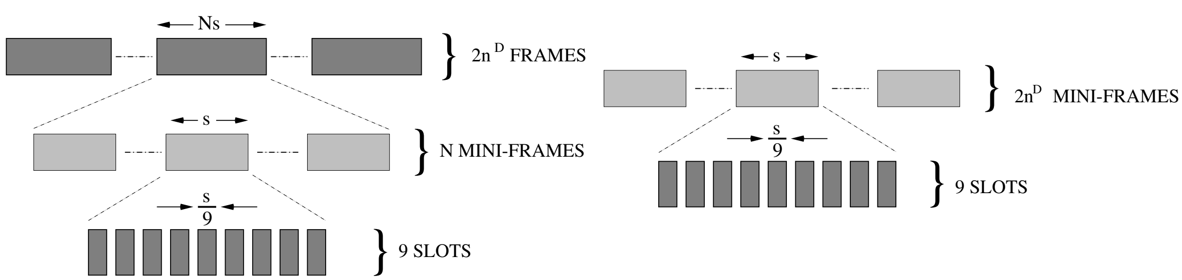
We derive lower bounds on the capacity of ad hoc wireless networks with a large number of nodes. Following [1] and [2], our approach is to construct schemes that achieve a given per-node throughput with probability approaching unity as the number of nodes increases. We first study networks with immobile nodes, possibly in the presence of fading. We then study networks where the nodes are mobile, the channel exhibits fading, and delay constraints are placed on the delivery of packets. Our investigation leads to the establishment of a fundamental tradeoff between packet delay and throughput.

We take the strategic decision, very early on, to concentrate on the *exponent* of n. This is plain from our definition of order. Therefore, we sacrifice the ability to discern any factor of the capacity that is slower than polynomial (notably logarithmic), but our proofs become simpler. Tighter bounds, that maintain logarithmic factors but are somewhat lengthier, appear in [4].



(a) (b)

Fig. 1. **(a)** One of the 9 sub-lattices of cells appears shaded. Only nodes in the sub-lattice are allowed to receive in the corresponding slot, and only from nodes in the same or neighboring cells. The neighbors of cell c_j appear striped. **(b)** A fundamental tradeoff between the packet delays and the per-node achievable throughput.



(a) (b)

Fig. 2. **(a)** The frame structure used in the mobility scheme. **(b)** The frame structure of each of the N virtual channels.

References

[1] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[2] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad-hoc wireless networks," in *Proc. IEEE INFOCOM*, vol. 3, Anchorage, AL, Apr. 2001, pp. 1360–1369.

[3] S. Toumpis and A. J. Goldsmith, "Capacity regions for wireless ad hoc networks," in *Proc. IEEE ICC*, vol. 5, New York, NY, Apr. 2002, pp. 3168–3173.

[4] ——, "On the capacity of large wireless networks under fading, mobility, and delay constraints," Engineering Library, Stanford University, Tech. Rep. WSL-2003-07-23, July 2003.

[5] A. J. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1218–1230, Oct. 1997.

[6] R. Motwani and P. Raghavan, *Randomized Algorithms*, 1st ed. Cambridge, UK: Cambridge University Press, 1995.

[7] T. Cover and J. Thomas, *Elements of Information Theory*, 1st ed. New York: Wiley Interscience, Aug. 1991.

[8] D. Shah, J. Mammen, A. E. Gamal, and B. Prabhakar, "Throughput-delay trade-off in wireless networks," submitted.