Delay Limited Capacity of Ad hoc Networks: Asymptotically Optimal Transmission and Relaying Strategy

Eugene Perevalov
Lehigh University
Bethlehem, PA 18015
Email: eup2@lehigh.edu

Rick Blum
Lehigh University
Bethlehem, PA 18015
Email: rblum@eecs.lehigh.edu

***Abstract*— The delay limited capacity of an ad hoc wireless network confined to a finite region is investigated. A transmission and relaying strategy making use of the nodes' motion to maximize the throughput is constructed. An approximate expression for the capacity as a function of the maximum allowable delay is obtained. It is found that there exists a critical value of the delay such that: (1) for values of the delay d below critical, the capacity does not benefit appreciably from the motion, (2) for moderate values of the delay d above critical, the capacity that can be achieved by taking advantage of the motion increases as $d^{2/3}$, (3) the dependence of the critical delay on the number of nodes is a very slowly increasing function ($n^{1/14}$) . Finally, asymptotic optimality of the proposed strategy in a certain class is shown.**

I. Introduction

Ad hoc wireless networks [1],[2] represent a promising new technology in communications that is currently receiving significant attention due to several successful research programs. In particular, one early effort was the DARPA packet-radio network program [3],[4]. Important applications include range from rescue operations to collaborative computing in mobile environments to distributed control and command systems to ubiquitous personal communication systems. Many researchers in the field of wireless communications believe that, due to their unique features, ad hoc networks will play an increasingly important role in the near future. Several attractive features of ad hoc networks are: (1) ease of deployment due to the absence of required infrastructure; (2) potentially low cost due to omission of large-scale hardware such as base stations; (3) very high degree of flexibility. An ad hoc network is a collection of (in general mobile) nodes that can exchange information via a wireless channel characterized by the absence of any fixed infrastructure or hierarchy. Present and future designers of such networks have to meet multiple challenges created by the networks' very nature. Significant difficulties arise at all levels of such networks: physical, MAC (medium access control), and the network layer. The main source of such difficulties is the same as the source of their potential advantages: essential lack of inherent organization in the network due to the absence of infrastructure. The latter feature implies that all questions of control, due to lack of any coordinating center, have to be addressed by the participating nodes themselves. Mobility makes the situation especially complex since, under such conditions, the topology of the network is constantly changing, and all the control decisions have to reflect that.

In this paper, we study the maximum information transport capacity of an ad hoc wireless network. More precisely, we are interested in the relationship between the end-to-end delay and the capacity. This work is motivated by the results of Gupta and Kumar [5] and Grossglauser and Tse [6] (see also [8] for a different approach). In [5], it was shown that, in an ad hoc network of size n, the capacity per node goes down with n thus making large networks impractical. In [6], it was demonstrated that, if the nodes' mobility is taken advantage of, the effect of decreasing capacity can be overcome. The price one has to pay for such a dramatic increase in capacity is an end-to-end delay no smaller than the time scale characterizing the nodes' motion. In this paper, we make an attempt to quantify the relationship between the maximum allowable delay and the transport capacity.

First, we establish an upper bound on the delay limited capacity within the class of "one relay" strategies in the spirit of [6]. We find an analytic expression for the upper bound and use it to get a simple approximate result valid for moderate values of the maximum delay. We then construct a transmission and relaying strategy that achieves the above upper bound asymptotically. Our approach is based on the combination of the *diversity routing* idea of Grossglauser and Tse [6] and the multipath routing methodology of Tsirigos and Haas [10] that relies on the *diversity coding* approach from [9]. Namely, just as in [6] a source node transmits to its current nearest neighbor at each time slot allocated for transmission. The difference is that, in our approach, we do not send a packet to its destination via one relay node. Instead, after adding redundant information, we split the resulting "enlarged" packet into many blocks and send the blocks to the destination via different relay nodes.

As a result, in order to achieve the desired level of service (measured as the probability of correct reconstruction of the message by the destination node within time d from the moment of the message origination), one needs to employ a certain redundancy level which in turn directly affects the maximum capacity. We calculate the required redundancy level approximately.

The main result that we obtain is that there exists a critical value of the delay such that for delays below the critical value, the gain in the capacity that can be achieved by making use of the motion is negligible. In other words, for such delays the result of [5] applies: the capacity of the network goes down roughly as $\frac{1}{\sqrt{n}}$. For the delays d above the critical, the capacity benefits from the motion, and, for not very long delays, increases approximately as $d^{2/3}$. It is interesting to note that the value of the critical delay increases only very slowly (as $n^{1/14}$) with the number of nodes n which is a welcome feature.

II. Model and Previous Results

The model we adopt is similar to to those used in [5] and [6]. The network consists of n nodes located on a sphere of area A. All nodes are mobile, and we assume that the motion of any node is described by the same stationary ergodic random process such that at each time there is no preferred direction. The trajectories of all nodes are assumed to be independent and identically distributed. We assume that every node i has an infinite amount of data for its destination $f(i)$, and that the source-destination association does not change with time.

In the transmission model we use (same as in [6]), a node i is capable of transmitting W bits/sec to node j at time t if

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

where P_i is the transmit power of node i, γ_{ij} is the channel gain from node i to node j, N_0 is the background noise power, L is the processing gain of the system, and β is the SIR requirement for successful communication. The channel gain is assumed to depend only on the shortest distance D_{ij} between the respective nodes as in

$$\gamma_{ij}(t) = \frac{1}{D_{ij}^{\alpha}(t)},$$

where α is a parameter greater than 2.

At any time a scheduler decides which nodes transmit bits and the corresponding power levels. The objective is to ensure a high average throughput for every source-destination pair. Let us denote by $M_i(t)$ the number of bits that the destination $f(i)$ receives in time slot t. We say that an average throughput of C is feasible if for every source-destination pair i,

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

Gupta and Kumar [5] demonstrated that, if a node can transmit W bits per second over a common wireless channel, there exist constants c and c' such that

$$\lim_{n\to\infty} Pr\left\{C(n) = \frac{cW}{\sqrt{n\log n}} \text{ is feasible}\right\} = 1,$$

and

$$\lim_{n\to\infty} Pr\left\{C(n) = \frac{c'W}{\sqrt{n}} \text{ is feasible}\right\} = 0,$$

i.e. up to a factor of $\log n$ the throughput per source-destination pair goes to zero as $\frac{W}{\sqrt{n}}$.

Grossglauser and Tse [6] constructed a scheduling policy according to which, in any time slot t, $n_S = \theta n$ (where θ is the sender density parameter to be determined) nodes are designated as senders and the remaining n_R nodes as potential receivers. Each sender node then transmits packets to to its nearest neighbor among potential receivers using unit transmit power. Among the n_S sender-receiver pairs the policy retains those for which the interference generated by the other senders is low enough so that a successful transmission is possible. If N_t is the number of such pairs then, as was shown in [6],

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

III. Upper Bound on Delay Limited Capacity

A. Maximum achievable capacity

Let us denote by C_∞ the per node capacity of the network achieved by the one relay node approach [6] in the absence of end-to-end delay constraints. We now find an upper bound on the capacity in the presence of a uniform end-to-end delay constraint.

Theorem 1: In the class of relaying strategies where each packet goes through at most one relay node, the maximum capacity C_d of an ad hoc network under the constraint that the end-to-end delay not exceed d is upper bounded by

$$C_d \leq C_d^{(u)} \equiv C_\infty \cdot \gamma\langle p(d)\rangle, \tag{1}$$

for sufficiently large n, where $\langle p(d)\rangle$ is the ensemble average of the probability that two nodes come within range of direct transmission in a time not exceeding d, and γ is the corresponding capture probability.

Proof: Since the relative contribution of direct source to destination transmissions to the total capacity is negligible (of the order $\frac{1}{n}$), we will ignore it in the proof. Let us concentrate on a fixed source node i with its associated destination $f(i)$. Suppose node k is the current potential recipient of a packet intended for node $f(i)$. If it were known in advance that node k would come within transmission range of $f(i)$ in time d, and the transmission to $f(i)$ would be successful, the transmission to k would have to take place. Otherwise, the transmission from i to k would be useless – the packet would not reach its destination in time. So, in the ideal case of complete knowledge of all the trajectories the capacity of $C_\infty \cdot \gamma\langle p(d)\rangle$ would be achieved.

To end the proof, we note that stated expression for the upper bound formally approaches γC_∞ as $d \to \infty$. In the following, we will see that, for an optimally chosen transmission range form relay to destination, $\gamma \to 1$ as $d \to \infty$. Moreover for large values of allowable delay, the two given nodes come close on multiple occasions, thus making the effective probability of capture approach unity even faster. ■

B. Probability of close range transmission

We would like to find an approximate expression for the ensemble average $\langle p(t) \rangle$ of the probability of a close range transmission within time t. More precisely, we wish to find the ensemble average of the probability that a given node comes within distance r or less from another fixed node in an interval of time of length t. It turns out to be easier to find the probability of the opposite event: $\langle q(t) \rangle = 1 - \langle p(t) \rangle$. Let us denote by $q_{\vec{x},\vec{y}}(t)$ the probability of two nodes not coming within range r provided that at time 0 the two nodes in question are located at points $\vec{x}$ and $\vec{y}$ respectively. Then the the ensemble average $\langle q(t) \rangle$ can be defined as the expectation of $q_{\vec{x},\vec{y}}(t)$ with respect to all possible (uniformly distributed on the sphere) starting points $\vec{x}$ and $\vec{y}$ which we denote as

$$\langle q(t) \rangle \equiv E_{\mathbf{U}_{\vec{x}},\mathbf{U}_{\vec{y}}}[q_{\vec{x},\vec{y}}(t)]. \tag{2}$$

Due to symmetry of the sphere, we can concentrate on the relative motion of the nodes by assuming that the second node is at rest at the north pole at all times. Then the above definition can be replaced by

$$\langle q(t) \rangle \equiv E_{\mathbf{U}_{\vec{x}}}[q_{\vec{x}}(t)], \tag{3}$$

where $q_{\vec{x}}(t)$ is the probability that the two nodes will not come within range r in time t provided that at time 0 the first node is at point $\vec{x}$ and the second is at the north pole.

Now fix a point in time s such that $0 < s < t$. Then provided the nodes did not come within r between 0 and s and the first node is at the (relative to the north pole) position $\vec{x}_1$, the probability of not coming within range r between 0 and t can be written as

$$q_{\vec{x}}(t) = q_{\vec{x}}(s) q_{\vec{x}_1}(t-s), \tag{4}$$

and taking expectations of both sides in (4) over all initial points and assuming the motion is independent over non-overlapping increments (an independent increment process, a common assumption) we obtain

$$E_{\mathbf{U}_{\vec{x}}}[q_{\vec{x}}(t)] = E_{\mathbf{U}_{\vec{x}}}[q_{\vec{x}}(s)] E_{\mathbf{P}_{\vec{x}_1}}[q_{\vec{x}_1}(t-s)], \tag{5}$$

where $E_{\mathbf{P}_{\vec{x}_1}}$ stands for the expectation with respect to the distribution of the position $\vec{x}_1$ of the first node at time s provided the nodes did not come within range between 0 and s. Now let $f_{\vec{x}}(s)$ be the probability density function characterizing the above probability distribution. We will write it as

$$f_{\vec{x}}(s) = \frac{1}{A}(1 + h_{\vec{x}}(s)), \tag{6}$$

where $h_{\vec{x}}(s)$ is the difference from the uniform distribution such that $h_{\vec{x}}(0) = 0$. Substituting (6) into the expression for the expectation over all possible starting points, we arrive at

$$E_{\mathbf{P}_{\vec{x}_1}}[q_{\vec{x}_1}(t-s)] = \langle q(t-s) \rangle + \epsilon(s, t-s), \tag{7}$$

where

$$\epsilon(s, t-s) = \frac{1}{A} \int q_{\vec{x}}(t-s) h_{\vec{x}}(s) d\vec{x}$$

accounts for the difference of the initial distribution in (7) from the uniform. Now, substituting (7) into (5), we obtain

$$\langle q(s) \rangle (\langle q(t-s) \rangle + \epsilon(s, t-s)) = \langle q(t) \rangle. \tag{8}$$

Note that in the absence of the term $\epsilon(s, t-s)$ the solution of (8) with the proper initial condition would be $\langle q(t) \rangle = e^{-\lambda t}$ for some constant λ. Writing $\langle q(s) \rangle = e^{-\lambda s} + \delta(s)$ and taking the limit $t \to s + 0$ in (8) we can arrive at the following ordinary differential equation for $\delta(s)$:

$$\delta'(s) = -\lambda \delta(s) + z(s)\delta(s) + e^{-\lambda s} z(s), \tag{9}$$

where $z(s) = -\lambda h_{\vec{x}'}(s)$, with $\vec{x}'$ being a point at a distance r from the north pole. The equation (9) with the initial condition $\delta(0) = 0$ has the solution

$$\delta(s) = e^{-\lambda s + \int_0^s z(s')ds'} \int_0^s e^{-\int_0^{s'} z(u)du} z(s')ds'. \tag{10}$$

Thus, for $\langle q(t) \rangle$ we obtain

$$\langle q(t) \rangle = e^{-\lambda t} \left(1 + e^{\int_0^t z(s)ds} \int_0^t e^{-\int_0^s z(u)du} z(s)ds \right). \tag{11}$$

The latter expression involves the function $z(s)$ whose exact form depends on the particular model of random motion of the nodes. However, the fact that $z(s) = -\lambda h_{\vec{x}'}(s)$ combined with the observation that $|h_{\vec{x}'}(s)| < 1$ for all s allows us to justify the approximate expression for the ensemble average of the probability of no direct nearest neighbor transmission within time t valid for moderate times t such that for $\lambda t \ll 1$ we have

$$\langle p(t) \rangle \simeq 1 - e^{-\lambda t}, \tag{12}$$

where the parameter λ characterizes the nodes' mobility. In order to estimate the value of λ, we note that, from (11), that

$$\lambda = \frac{\partial \langle p(t) \rangle}{\partial t}\Big|_{t=0}.$$

To evaluate λ from this definition, we must calculate the number of nodes that enter a circle of radius r during a differential time interval assuming uniformly distributed nodes over a sphere of radius R which are moving at speed v. Given this interpretation the parameter λ can be shown to be equal to

$$\lambda(r) = \frac{2vr}{A} = \frac{vr}{2\pi R^2}, \tag{13}$$

Introducing a dimensionless parameter $x \equiv r/R$ we can rewrite the above expression as

$$\lambda(x) = \frac{v}{2\pi R} x, \tag{14}$$

Let us also introduce notation for the average length of time during which the identity of a nearest neighbor of a node remains unchanged. We will denote such an average by τ. Obviously, in the order of magnitude, τ is equal to the time it takes a node to travel an average distance between the nodes so that

$$\tau \sim \sqrt{\frac{A}{n}} \frac{1}{v} = \sqrt{\frac{cR^2}{n}} \frac{1}{v}, \tag{15}$$

where c is some constant depending on the details of the motion model.

C. Probability of capture

A typical information block on its way from the source node to the destination performs two hops: from the source to a relay node, and from the relay to the destination. We describe these two stages in turn below. For simplicity we assume, following [6], that in odd time slots the first stage is effected, and in even time slots – the second.

1) Source to relay: The source to relay transmission is effected to the nearest neighbor as described in [6]. There, it was shown that the capture probability approaches a finite number for very large number of nodes n. Let us denote this number η. The fact that η is substantially less than 1 reduces the delay limited capacity by effectively increasing τ – the average time between two successful transmissions in the first stage. In this paper, we do not go into the detail of transmission policy at this stage postponing it to the future work.

2) Relay to destination: In the second stage our transmission policy is to transmit to the destination once it is at the distance r from the relay. It is clear that choosing higher value of r increases the probability of coming into the range within limited time and, at the same time, decreases the probability of capture. So one can hope to be able to choose the optimal value of r given the parameters of the system. The probability of coming in the range was studied in the previous section. Here, we approximately compute the probability of capture. We fix the parameter α describing the decay of the signal in space to $\alpha = 4$.

First, note that if any other transmitter is at a distance no more than $r' = r\beta^{1/4}$ from our destination then the capture is impossible. Let us denote by $\mathcal{A}$ the event that none of the other transmitting relays are within distance r' or less from our destination. The probability of $\mathcal{A}$ can be computed as

$$P(\mathcal{A}) = \left(1 - \frac{\sqrt{\beta}r^2}{4R^2}\right)^q, \tag{16}$$

where q is the number of simultaneously transmitting relays which is on average equal to

$$q = \langle p(d) \rangle \theta\eta n,$$

where θ is the sender density in the first stage. Using the fact that $r \ll R$ (as a consequence of large n) we can write an approximate expression for (16):

$$P(\mathcal{A}) = e^{-(1/4)\sqrt{\beta}n\theta\eta\langle p(d)\rangle x^2}, \tag{17}$$

where $x \equiv r/R$.

If $\mathcal{A}$ is true, i.e. none of other transmitting nodes is within distance r' from our destination, the capture may still be impossible due to total interference power from all the other transmitting nodes. Since the number of such nodes is large (proportional to n), we can approximate the distribution of the interfering power by the normal one with the mean equal to $q\overline{P}_1$ and standard deviation equal to $\sqrt{q}Std(P_1)$, where $\overline{P}_1$ and $Std(P_1)$ are the mean and standard deviation of the interfering power of *one* other node, respectively.

If the transmitting power of each node is equal to 1, we can calculate $\overline{P}_1$ and $Std(P_1)$ approximately as

$$\overline{P}_1 = \frac{1}{4}\frac{1}{R^2 r'^2},$$

and

$$Std(P_1) = \frac{1}{2\sqrt{3}}\frac{1}{Rr'^3}.$$

So, given that $\mathcal{A}$ is true, the probability that the transmission is successful, can be computed as

$$\Phi\left(\frac{\frac{1}{\beta r^4} - \frac{q}{4}\frac{1}{R^2 r'^2}}{\frac{\sqrt{q}}{2\sqrt{3}}\frac{1}{Rr'^3}}\right), \tag{18}$$

with $\Phi(.)$ denoting the standard normal cdf.

Finally, combining (17) and (18), we obtain an approximate (valid for large n) equation for the probability of capture $\gamma(x)$:

$$\gamma(x) = e^{-(1/4)\sqrt{\beta}n\theta\eta\langle p(d)\rangle x^2}\,\Phi\left(\frac{2\sqrt{3}(1 - \frac{\sqrt{\beta}\langle p(d)\rangle\theta\eta n}{4}x^2)}{\beta^{1/4}\sqrt{\langle p(d)\rangle\theta\eta n}\,x}\right). \tag{19}$$

It is convenient to introduce the dimensionless parameter $w \equiv \frac{vd}{2\pi R}$ which measures the delay d in "natural" units counting how many times a node could traverse the spherical region if it moved in a straight line. Using this new parameter and remembering that, in the first approximation, the quantity $\langle p(d)\rangle$ is equal to λd, we can rewrite (19) as

$$\gamma(x) = e^{-(1/4)\sqrt{\beta}\theta\eta n w x^3}\,\Phi\left(\frac{2\sqrt{3}(1 - \frac{\sqrt{\beta}\theta\eta n w}{4}x^3)}{\beta^{1/4}\sqrt{\theta\eta n w}\,x^{3/2}}\right). \tag{20}$$

IV. Asymptotically Optimal Relaying Strategy

A. Strategy description

We now describe our relaying strategy. The goal is to get close to the maximum capacity. We achieve it by using a different approach from that in [6]. Specifically, we spread the packet traffic between many nodes. Namely, after adding redundant information, we split the resulting packet into blocks and send the latter via different routes (relay nodes). We employ the coding scheme used in [10] in which Y extra bits are added to the packet of X information bits as overhead thus resulting in $B = X + Y$ bits that are treated as one new network-layer packet. The additional bits are calculated as a function of the original X bits so that the original bits can be correctly reconstructed from any subset of the B bits of size no less than X. The quantity

$$z = \frac{B}{X} \tag{21}$$

is the overhead factor. Our strategy consists of splitting the resulting B-bit packet into m equal size blocks and sending them via different (consecutive) relay nodes. Similarly, the blocks are communicated from the relays to the destination.

B. Approximate capacity calculation

The key question we have to answer is, given the maximum allowable delay d, how do we select the overhead ratio in order to achieve the required probability (close to 1) that a correct message is received in the required time. If we are able to find the minimum sufficient overhead ratio $z_{\min}$ then obviously the capacity such a strategy can achieve will be equal to

$$C_d = \frac{C_\infty}{z_{\min}}. \tag{22}$$

In this paper we will limit ourselves to the case when, for each B-bit packet, exactly one block is sent to the destination via each relay node. Observe that, in establishing the upper bound, we assumed the perfect knowledge of the future trajectories of the nodes. Thus we could get the information to the corresponding destination in time 100% of the time. Here, due to the lack of such knowledge, all we can aim at is to get the packets to the destination in time with some fixed (although arbitrary) average probability. Note though that if 100% on-time delivery were needed that could be achieved by combining our relaying strategy with occasional multi-hop transmissions. The asymptotic result that we report below would still hold.

Let us fix the desired level of service Q, the average probability that a packet will reach the destination within time d. Our task is to determine the minimum overhead ratio such that the desired level of service Q can be achieved. Using the results of [10], we can write an approximate expression for the probability of successful reconstruction of a packet at the destination within time t as

$$P_m(t) = \frac{1}{2} + \frac{1}{2} \cdot erf\left(\frac{\sum_{i=1}^m \gamma p_i(t_i) - \lceil m/z \rceil + 1/2}{\sqrt{2\sum_{i=1}^m \gamma p_i(t_i)(1-\gamma p_i(t_i))}} \right), \tag{23}$$

where m is the number of blocks the packet is split into, $p_i(t_i)$ is the probability the i-th block reaches the destination within time t_i[1], and z is the employed overhead ratio.

We demand that the ensemble average of $P_m(t)$ be equal to Q so that

$$\frac{1}{2} + \frac{1}{2} \cdot \left\langle erf\left(\frac{\sum_{i=1}^m \gamma p_i(t_i) - \lceil m/z \rceil + 1/2}{\sqrt{2\sum_{i=1}^m \gamma p_i(t_i)(1-\gamma p_i(t_i))}} \right) \right\rangle = Q. \tag{24}$$

The LHS of (24) seems to be hard to evaluate, so we simplify it by first noting that $\sum_{i=1}^m \gamma p_i(t_i)(1-\gamma p_i(t_i)) \le m/4$ and hence, if we demand that

$$\frac{1}{2} + \frac{1}{2} \cdot \left\langle erf\left(\frac{\sum_{i=1}^m \gamma p_i(t_i) - \lceil m/z \rceil + 1/2}{\sqrt{m/2}} \right) \right\rangle = Q \tag{25}$$

then the resulting level of service will be no less than Q. Next, we assume that the probabilities $p_i(t_i)$, $i = 1, \ldots, m$, are independent and identically distributed[2]. Under this assumption, the sum $\sum_{i=1}^m p_i(t_i)$ is approximately normally distributed with mean $\sum_{i=1}^m \langle p_i(t_i) \rangle$ and standard deviation not exceeding $\sqrt{m}/2$ (since the standard deviation of $p_i(t_i)$ is no more than 1/2 for all i). Therefore, the argument of $erf(\cdot)$ in (25) is approximately normally distributed with the mean of

$$\mu = \frac{\sum_{i=1}^m \gamma \langle p_i(t_i) \rangle - \lceil m/z \rceil + 1/2}{\sqrt{m/2}}$$

and standard deviation $s \le 1/\sqrt{2}$. This allows us to find the value of μ such that (25) holds from

$$\frac{1}{\sqrt{\pi}} \int_0^\infty \left(\frac{1}{2} + \frac{1}{2} erf(x) \right) e^{-(x-\mu)^2} dx = Q. \tag{26}$$

Table I shows the numerical values of μ for some fixed levels of service Q.

TABLE I

$\mu(Q)$ FOR SOME LEVELS OF SERVICE Q.

Q	0.90	0.95	0.98	0.99	0.999
μ	1.28	1.65	2.05	2.33	3.10

We can now find the value of $z_{\min}$ by solving the following optimization problem and equating the objective value to $\mu(Q)$:

$$\max_m \frac{\sum_{i=1}^m \gamma \langle p_i(t_i) \rangle - \lceil m/z \rceil + 1/2}{\sqrt{m/2}} = \mu(Q). \tag{27}$$

The above equation states that we want to reach the desired level of service Q for at least one (optimal) value of the number of blocks m which we denote m^*. The maximization over m will lead to the smallest value of $z_{\min}$ as illustrated in Fig.1. For further convenience, let us denote

$$g(m,r) \equiv \frac{\sum_{i=1}^m \gamma \langle p_i(t_i) \rangle - \lceil m/z \rceil + 1/2}{\sqrt{m/2}}$$

so that (27) reads

$$\max_m g(m,z) = \mu(Q). \tag{28}$$

We can now substitute the approximate expression (12) for $\langle p_i(t_i) \rangle$ into (27). Strictly speaking, since the time periods during which the identity of the nearest neighbor of a node is unchanged are random variables, we would have to take the averages over the corresponding distributions for calculating the quantities $\langle p_i(t_i) \rangle$ for all values of the index i. However we can use the fact (confirmed by the results) that the typical number m^* of the blocks into which a packet is split is large for large n and, therefore, we can get a good approximation by replacing the random variables by their means. In this way, we obtain

$$g(m,z) = \frac{\gamma(m - e^{-\lambda d} \sum_{i=1}^m e^{(i-1)\lambda\tau}) - \lceil m/z \rceil + 1/2}{\sqrt{m/2}} \tag{29}$$

Obtaining a closed form solution of (28) still seems to be a difficult task. Therefore, we use yet another approximation. For a large total number of nodes n, the optimal value of m

[1] t_i is the amount of time it has left until the 'deadline' t. Since the transmission from the source to the i-th relay for $i > 1$ happens later than time 0, $t_i < t$ unless $i = 1$.

[2] This assumption is a reasonable approximation provided $\tau m \ll d$.

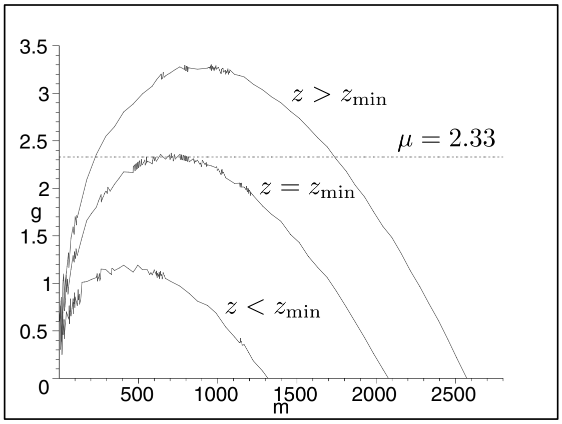

Fig. 1. $g(m, z)$ (as given by (29)) as a function of m for different values of r and $Q = 0.99$. We can see that for $z < z_{\min}$ the required level of service ($\mu = 2.33$) is not achieved for any m, while for $z > z_{\min}$ it is achieved or exceeded for a whole range of values of m.

is going to also be large. Thus we: (1) replace the sum by an integral; (2) replace $\lceil m/z \rceil$ in (29) by m/z; (3) drop terms which are small compared to m. As a result of the above approximations, (29) becomes

$$g(m, z) = \gamma\sqrt{\frac{2}{m}}\left(m\left(1 - \frac{1}{\gamma z}\right) - e^{-\lambda d}\frac{1}{\lambda\tau}(e^{m\lambda\tau} - 1)\right). \tag{30}$$

Differentiating (30) with respect to m, and Taylor expanding the result to first order in the small parameter $\lambda\tau$, we obtain the value m^* maximizing $g(m, z)$ for a fixed z.

$$m^* = \frac{2}{3}\frac{1}{\lambda\tau}\left(e^{\lambda d}\left(1 - \frac{1}{\gamma z}\right) - 1\right). \tag{31}$$

Now substituting (31) back into (30) we obtain

$$g(m^*, z) = \frac{4\gamma}{3\sqrt{3}}e^{-\lambda d}\sqrt{\frac{1}{\lambda\tau}\left(e^{\lambda d}\left(1 - \frac{1}{\gamma z}\right) - 1\right)^3}. \tag{32}$$

Finally, solving (28) for z using (32) we arrive at the following expression for $z_{\min}$:

$$z_{\min} = \frac{1}{\gamma}\left(1 - e^{-\lambda d} - \left(\left(\frac{3\sqrt{3}\mu}{4\gamma}\right)^2 e^{-\lambda d}\lambda\tau\right)^{\frac{1}{3}}\right)^{-1}, \tag{33}$$

provided the expression in the outmost brackets is positive. Otherwise, the required level of service Q cannot be achieved.

Thus, the capacity under the constraint that the end-to-end delay not exceed d with probability no less than Q is approximately equal to

$$C_d = \gamma\left(1 - e^{-\lambda d} - \left(\left(\frac{3\sqrt{3}\mu(Q)}{4\gamma}\right)^2 e^{-\lambda d}\lambda\tau\right)^{\frac{1}{3}}\right)C_\infty, \tag{34}$$

if the expression in the brackets is positive. Otherwise, the required level of service Q cannot be achieved, and we can consider the corresponding capacity to be equal to 0.

C. Optimal transmission range

Analyzing eq. (32), we can see that, for a given level of service Q, there exists a critical delay d_{cr} such that:

- For $d < d_{\text{cr}}$, the transmission strategy discussed above is unable to achieve the required level of service.
- For $d > d_{\text{cr}}$, the capacity increases with d as illustrated in Fig. 2.

We can estimate the value of d_{cr} from (34) by equating C_d to 0 as

$$d_{\text{cr}} = \left(\frac{3\sqrt{3}\mu(Q)}{4\gamma}\right)^{\frac{2}{3}}\left(\frac{\tau}{\lambda^2}\right)^{\frac{1}{3}}. \tag{35}$$

Expanding (34) to the first order in λd, we see that C_d as a function of delay d behaves approximately as

$$\frac{C_d}{C_\infty} = \begin{cases} 0 & \text{if } d < d_{\text{cr}} \\ \gamma\lambda(d - d_{\text{cr}}) & \text{if } d > d_{\text{cr}} \end{cases}, \tag{36}$$

provided $\lambda d \ll 1$. It is interesting to compare (36) with the corresponding approximate expression for the upper bound $C_d^{(u)}$:

$$C_d^{(u)} = \gamma\lambda d,$$

which can be obtained from (36) by setting $d_{\text{cr}} = 0$.

So far we have not chosen the transmission range r. Recall that the quantities λ and γ in (35) and (36) depend on r (or its dimensionless version x) as shown in (14) and (20), respectively. So we can choose the value of x such that the capacity is maximized for any fixed value of the delay d.

By inspection of (36) we can see that the capacity for values of the delay greater than d_{cr} depends linearly on the combination $\gamma(x)\lambda(x)$. At the same time, the value of d_{cr} from (35) is inversely proportional to a positive power of the same combination. Therefore, one can maximize the capacity by choosing the value of x so that $\gamma(x)\lambda(x)$ is maximized. In order to achieve this goal, note that, as x increases, the first factor in the expression (20) for $\gamma(x)$ drops much faster than the second one which stays at nearly 1 until the argument of $\Phi(\cdot)$ becomes roughly less than 2 ($\Phi(2) = 0.977$, $\Phi(1.5) = 0.933$). Therefore we can try to maximize $\gamma(x)\lambda(x)$ by setting the second factor in $\gamma(x)$ to 1 and later verifying that the resulting argument of $\Phi(\cdot)$ is large enough to justify the approximation. Thus we need to maximize the function

$$\frac{v}{2\pi R}\, x\, e^{-(1/4)\sqrt{\beta}\theta\eta n w x^3}$$

with respect to x which yields the optimal x equal to

$$x_{\text{opt}} = \sqrt[3]{\frac{4}{3\sqrt{\beta}\theta\eta n w}}, \tag{37}$$

so that the argument of $\Phi(\cdot)$ in the expression for $\gamma(x)$ is equal to 2 and our approximation is well justified.

Substituting (37), (14) and (20) into (35), and solving the resulting equation for d_{cr}, we obtain:

$$d_{\text{cr}} = \left(\frac{3\sqrt{3}\mu}{4}\right)^{\frac{6}{7}}(3\pi^2 e\sqrt{\beta}\theta\eta)^{\frac{2}{7}} c^{\frac{3}{14}} n^{\frac{1}{14}}\frac{R}{v}, \tag{38}$$

or, in the "natural" units:

$$w_{\rm cr} = \left(\frac{3\sqrt{3}\mu}{4}\right)^{\frac{6}{7}} \frac{(3\pi^2 e\sqrt{\beta}\theta\eta)^{\frac{2}{7}}}{2\pi} c^{\frac{3}{14}}\, n^{\frac{1}{14}}. \qquad (39)$$

We also get:

$$\gamma(x_{\rm opt})\lambda(x_{\rm opt})d = \left(\frac{v^2 d^2}{3\pi^2 e\sqrt{\beta}\theta\eta n R^2}\right)^{\frac{1}{3}}. \qquad (40)$$

Thus the behavior of the capacity as a function of d for moderate values of d is such as shown in Fig. 2.

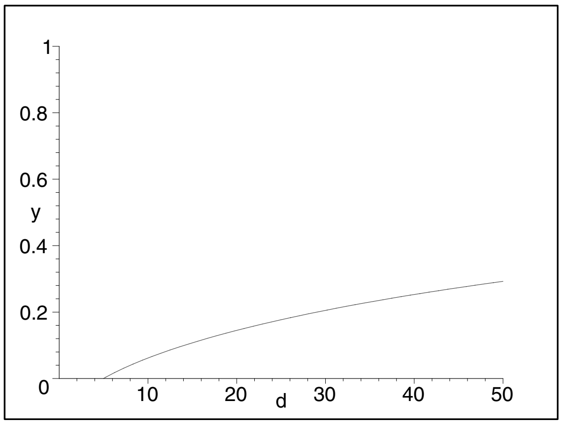

Fig. 2. The ratio $y = C_d/C_\infty$ as a function of delay d.

Several observations are now in order.

- The dependence of the critical delay $d_{\rm cr}$ on the number of nodes n turns out to be very slow ($n^{1/14}$) which is a desirable feature.
- Beyond $d_{\rm cr}$, the dependence of the capacity on d is as $d^{2/3}$ so significant capacity is reached relatively quickly.
- As can be seen from (37), the optimal transmission range decreases as $n^{-1/3}$, i.e. for large values of n it can be significantly larger than the typical internode distance that scales as $n^{-1/2}$. It is also interesting to note that the optimal transmission range decreases with d as $d^{-1/3}$.

From the above expressions, we can easily find that, for the optimal value of the transmission range,

$$\lambda\tau = \frac{1}{\pi}\sqrt{c}\left(\frac{1}{6\sqrt{\beta}\theta\eta w}\right)^{\frac{1}{3}} \frac{1}{n^{5/6}},$$

and the upper bound on the delay limited capacity $C_d^{(u)}$ can be approximated as $C_d^{(u)} = (1 - e^{-\lambda d})\gamma C_\infty$. So if we fix $C_d^{(u)}$ and increase the number of nodes n we see that the ratio $C_d/C_d^{(u)}$ approahes[3] 1. Thus we have proved the following theorem.

Theorem 2: There exists a transmission strategy that asymptotically achieves the upper bound $C_d^{(u)}$ on the delay-limited capacity of an ad hoc network.

[3]We should note that this ratio approaches 1 rather slowly as shown in Fig.3.

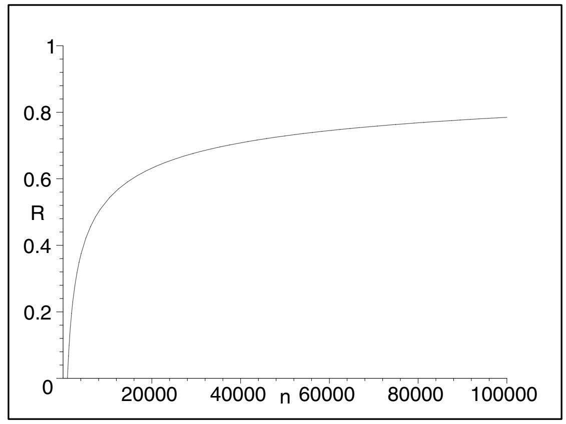

Fig. 3. The ratio $R = C_d/C_d^{(u)}$ as a function of number of nodes n. Note that the ratio approaches 1 rather slowly. This example uses $Q = 0.99$ and $\lambda d = 0.2$.

V. Conclusion

In this paper we have conducted a preliminary exploration of the problem of the influence of the end-to-end delay on the maximum capacity of a wireless ad hoc network confined to a certain area. Aiming at general results, we have made a number of simplifying assumptions. Thus, we adopted a "totally random" model of motion ignoring both the details of the corresponding distributions and possible patterns in the nodes' motion. We also limited ourselves to "one relay" class of strategies in the spirit of ref. [6]. Confining the analysis to the above class of strategies we found an expression for the upper bound of the delay limited capacity that involves the ensemble average of the probability of two nodes coming within certain range within the maximum delay time and the corresponding capture probability. In order to establish that upper bound we assumed the perfect knowledge of the future trajectories of the nodes. We then obtained a general expression for that ensemble average which depends on the transmission range from the relay to the destination. Next, we found an approximate expression of the average probability of success of that transmission as a function of the transmission range.

We then proceeded to construct a relaying strategy that would asymptotically achieve the upper bound. We used the *diversity coding* approach in combination with the "secondary" *diversity routing* of [6] in order to asymptotically achieve the upper bound for this class of strategies in the absence of any information about the nodes' motion. We also made use of a number of approximations that allowed us to end up with closed form expressions. Analyzing the dependence of the resulting capacity on the transmission range, we found an approximate expression for the optimal range that maximizes the capacity for any fixed value of the delay. We found that, for moderate delays, the dependence of the optimal capacity on the delay is characterized by the "critical delay" below which our relaying and transmission strategy does not lead to any appreciable capacity. For the values of the delay higher than the critical value, the capacity grows approximately as $d^{2/3}$.

The critical delay has a very slow dependence on the number of nodes n so that it practically is independent on n. It is interesting to note that the existence of that minimum delay is precisely the price we have to pay for the lack of the knowledge of the nodes' motion.

Finally, we have shown that our transmission and relaying strategy is asymptotically optimal in the sense that the ratio of the achieved capacity and the upper bound approaches 1 (albeit rather slowly) as the number of nodes n grows.

REFERENCES

[1] A. ALWAN ET AL., *Adaptive mobile multimedia networks*, IEEE Personal Communications Magazine, **2**, no. 3 (1996) 34.

[2] Z.J. HAAS AND M. PEARLMAN, *Panel report on ad hoc networks*, Mobile Computation and Communication Review, **2**, no. 2 (1998).

[3] A. EPHREMIDES, J.E. WIESELTHIER, AND D. BAKER, *A design concept for reliable mobile radio networks with frequency hopping signaling*, Proceedings of IEEE, **75**, no. 1 (1987) 56.

[4] J. JUBIN AND J.D. TORNOW, *The DARPA packet radio network protocols*, Proceedings of IEEE, **75** no. 1 (1987).

[5] P. GUPTA AND P.R. KUMAR, *The capacity of wireless networks*, IEEE Transactions on Information Theory, **46**, no. 2 (2000) 388.

[6] M. GROSSGLAUSER AND D. TSE, *Mobility increases the capacity of ad hoc wireless networks*, In *INFOCOM 2000 Proceedings*, **3** (2000) 1360.

[7] P. GUPTA AND P.R. KUMAR, *Towards an information theory of large networks: an achievable rate region*, In *Proceedings of 2001 IEEE International Symposium on Information Theory*, (2001) 159.

[8] S. TOUMPIS AND A. GOLDSMITH, *Capacity regions for wireless ad hoc networks*, working paper.

[9] E. AYANOGLU ET AL. *Diversity coding for transparent self-healing and fault-tolerant communications networks*, IEEE Transactions on Communications, **41** no. 11 (1993).

[10] A. TSIRIGOS AND Z. HAAS, *Multipath routing in the presence of frequent topological changes*, IEEE Communications Magazine, **39**, no. 11 (2001) 132.