Geographic Protocol Information and Capacity Deficit in Mobile Wireless Ad Hoc Networks

Alhussein A. Abouzeid, *Member, IEEE,* and Nabhendra Bisnik, *Member, IEEE*

Abstract

Overheads incurred by network protocols diminish the capacity available for relaying useful data in a dynamic communications network. Discovering lower bounds on the amount of protocol overhead incurred is important for the development of efficient network protocols, and for characterizing the *effective* capacity available for network users. This paper presents an information-theoretic framework for characterizing the minimum protocol overheads incurred for maintaining location information in a network with mobile nodes. Specifically, the minimum overhead problem is formulated as a rate-distortion problem. The formulation may be applied to networks with arbitrary traffic arrival and location service schemes. Lower bounds are derived for the minimum overheads incurred for maintaining the location of the nodes and consistent neighborhood information in terms of node mobility and packet arrival processes. This leads to a characterization of the deficit caused by the protocol overheads on the overall transport capacity.

Index Terms

Capacity, Geographic Routing, Overheads, Protocol Information, Rate-Distortion Theory, Routing, Wireless Ad-Hoc Networks, Wireless Networks

I. Introduction

A. Motivation

Communication networks have progressively become larger and more dynamic. Present society expects to be 'connected' at all times, even (and especially) when moving. Some task-specific networks, such as military networks, are even more dynamic and have to operate under more stringent environments, diverse traffic rates, and highly dynamic and in many cases unpredictable conditions.

Network protocols maintain various types of information about the network, depending on the application. For example:

- In a vehicular network, a central station may need to know the locations and speeds of various vehicles in order to detect impending traffic hazards and send warning signals to drivers/vehicles.
- In various Internet applications, servers want to send information to users depending on their current location, and users may be on the move.
- In a cognitive radio network, a secondary user may need to know the locations of all the primary users as well as the location of the intended receiver before it decides whether to transmit or not on the primary user spectrum.
- In a military wireless network (or a rescue and recovery mission), a commander may need to know the locations of all the (mobile) soldiers in a dynamic chaotic environment.

Protocols exchange information in order to maintain up to date information about the network, which we call in this paper *state information*. As networks continue to become larger and more dynamic, the state information changes rapidly over time and hence larger control overhead (i.e. protocol information) may be required to keep track of the network state. This reduces the *effective capacity* of the network; the total transport capacity less the protocol information (please see Section VI for more precise definition). This limits the scalability of networks. For example, according to a recent study, protocol overheads in a state-of-the-art U.S. Air Force network have exceeded 80% of the total carried traffic[1]. Thus, protocol information is a

A. A. Abouzeid (corresponding author) is with ECSE Department at Rensselaer Polytechnic Institute, Troy, NY 12180. email: abouzeid@ecse.rpi.edu Fax: +1 518 276 4403.

N. Bisnik is with Intel, Portland, Oregon. This work was conducted while he was a Ph.D. candidate with the ECSE Department at Rensselaer Polytechnic Institute, Troy, NY. email: nabhen@gmail.com

This work was funded in part by the National Science Foundation under grants CNS-322956 and CNS-546402.

[1]This information was communicated by the Air Force during a 2006 DARPA-organized informational meeting.

fundamental physical quantity in any dynamic network. A critical step towards developing scalable protocols is to recognize the limits on the control overhead that has to be incurred in order to guarantee certain performance constraints. Knowledge of such limits will not only provide a yardstick to evaluate performance of existing protocols but also inspire the development of protocols that actually achieve the limit.

In this paper, building on the seminal work by Gallager [1], we present an information-theoretic framework for characterizing the minimum overhead incurred by network protocols that maintain *state information related to the locations of the nodes*. We refer to this class of network protocols as *position-based network protocols*. One prominent example is geographic- or position- based routing protocols for mobile ad hoc networks, but there are many other cases as mentioned earlier. Information theory answered several fundamental questions in communication theory, such as finding the ultimate data compression and the ultimate transmission rate of communication [26]. Thus, it is reasonable to expect (e.g. see [6]) that information theory would be a suitable tool for developing lower bounds on the amount of overhead incurred by routing protocols for disseminating and gathering state information in a mobile ad hoc network. A detailed discussion of the connection with [1] is included in Section II.

In many network protocols, *the performance of the protocols, e.g. the packet delivery ratio, is a function of the extent of discrepancy between the actual and perceived state information*. The following example of a network protocol helps illustrate this point. In geographic routing each node maintains its location information at one or more 'location servers' (e.g. see [2], [3]). When a source wants to forward a packet to a destination, it queries an appropriate location server for the location of the destination. The location server replies to the source node with the available location information. Thereafter, the source and intermediate nodes forward the packet according to the location of the destination. It is pointed out in [4] that the fraction of packets delivered by geographical routing varies inversely with the average error in location information stored at the location servers. Thus, maintaining packet delivery ratio above a given threshold corresponds to maintaining location errors below a certain threshold.

We note that throughout this paper, without loss of generality, we make specific references to the case of geographic routing only for illustration purposes, but the approach and results are applicable to position-based network protocols in general.

B. Outline and Contributions

In wireless networks employing position-based protocols, nodes transmit messages periodically, typically called 'beacon' or 'HELLO' messages, in order to announce their existence to their neighbors. The objective is for nodes to maintain an up-to-date state of the list of neighbors. In addition, nodes send position updates to one or more location servers.

Two commonly incurred types of position-based protocol overheads are considered in this paper:

1) *Location update overhead*: The overhead incurred in updating the location information such that the location errors in the reply to location queries is less than ϵ, and,
2) *Beacon overhead*: The overhead incurred in beacon transmission such that the probability that a node has consistent neighborhood information when it is needed (e.g. at the instant a packet is to be forwarded) is greater than $1 - \delta$.

We formulate the problems of finding the minimum values of the above-mentioned overheads as rate-distortion problems. For location update overheads, the distortion measure used is squared error in the location information (*squared error distortion measure*). For beacon overheads, the distortion measure is the probability that a perceived neighbor is not an actual neighbor (*Hamming distortion measure*). Using a rate-distortion formulation, we present lower bounds on the minimum protocol information (i.e. overhead) incurred in terms of node mobility, packet arrival process and reliability criteria ϵ and δ.

We then unite the results obtained here on the minimum position based protocol overheads with the results on the transport capacity of stationary multi-hop wireless networks evaluated in [5] in order to characterize the *effective* capacity available for users. It is observed that when the node mobility is high and the average packet inter-arrival time is sufficiently small, the complete transport capacity of an ad hoc network may be consumed by protocol overheads. We derive an upper bound on the critical network size above which all the transport capacity of the network would be consumed by the protocol overheads and no useful communication would be possible (Theorem 5). This implies that protocols aimed at tracking position state information are not scalable beyond the specified network dynamism (i.e. number of nodes, area, mobility and traffic patterns).

The main contributions of this paper may be summarized as follows:

1) We present a new information-theoretic formulation for evaluating the minimum overheads incurred by position-based network protocols in maintaining position-related information. The formulation is general so that it may be applied to any node distribution and packet arrival process, and may be extended to any location service scheme and mobility model.
2) For Brownian mobility model and various *packet* (or location information *request*) inter-arrival time distributions, we evaluate lower bounds on the minimum rate at which a node must transmit its location information and beacons such that the errors in location and neighborhood information are bounded. Combining both overheads, we find a lower bound on the capacity deficit caused by protocol overheads in mobile wireless ad hoc networks.
3) We characterize the effective transport capacity of an ad hoc network after taking into account the minimum protocol overheads that must be incurred to keep state distortion bounded.
4) For a given packet/query arrival process, standard deviation of Brownian motion and reliability parameters (ϵ, δ), we evaluate the upper bound on the number of nodes the ad hoc network can support such that the complete transport capacity of the network is not used up by protocol overheads.

The rest of the paper is organized as follows. A brief overview of related work is presented in Section II. The network model is presented in Section III. The rate-distortion formulation and evaluation of a lower bound on the minimum position update and beacon overheads are presented in Sections IV and V respectively. A discussion of the capacity deficit caused by routing overheads is presented in Section VI. The application of the formulation to other scenarios and some possible extensions of the network model is discussed in Section VII. We present conclusions and directions for future research in Section VIII.

II. Related Work

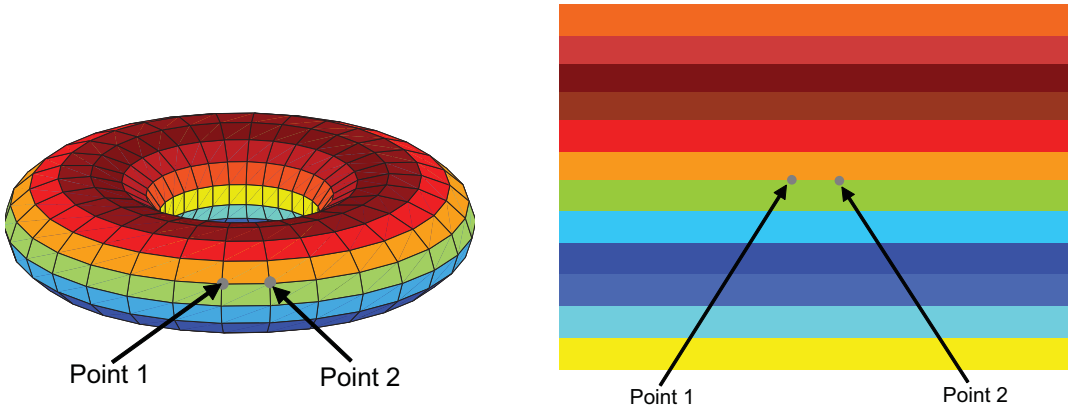
So far, we believe information theory has not significantly influenced the design and understanding of communication network protocols (an opinion we share with the authors of [6]). One of the earliest (and most significant) attempts in using information theory to enhance the understanding of communication network protocols was made in [1]. Gallager [1] used an information-theoretic approach in order to characterize a lower bound on the amount of protocol information required to keep track of the sender, receiver and timing of messages for a simple (stationary) network model. It is found that although the introduction of message delay decreases the protocol information, small average message length and high message arrival rate may lead to prohibitively high protocol overhead.

A few relatively recent papers have used information theory to understand the effects of node mobility on protocol overheads in wireless networks. Most of this literature appears in the context of routing protocols and thus we review the most related results below.

An analytical framework, based on entropy of node location, for characterizing delay and overhead associated with paging and routing a call to a mobile station in a cellular environment is provided in [7]. The complexity of tracking a mobile user in a cellular environment is studied using a information-theoretic approach and a position update and paging scheme is proposed in [8]. An entropy based modeling framework for evaluating and supporting route stability in mobile ad hoc networks is proposed in [9]. In [10], the authors propose the entropy of link change as the metric for mobility models against which performance of wireless network protocols could be evaluated.

The overhead incurred by routing protocols and their scalability properties has been studied in [11]–[21]. The initial studies [11], [12], [14], [15] are mainly simulation based. These studies point out that none of the routing protocols performs well across all scenarios. Instead each protocol performs well in some scenarios and bad in others.

Although simulation-based studies provide useful information about the performance of routing protocols, the observations may not be generalized to all scenarios. Therefore it is important to have analytical results for the performance of routing protocols in order to be able to develop a deep and general understanding of the trade-offs involved. In [13], [17], the authors present an analytical framework for characterizing the routing overhead for ad hoc routing protocols. Asymptotic results are provided for the overhead of proactive and reactive protocols in terms of network and routing protocol parameters such as packet arrival rate, hello packet transmission rate, hello packet size, size of route request packet, topology broadcast rate etc. The overheads of specific routing protocols were modeled in [16]. The impact of the routing-layer traffic patterns (in terms



(a) Points on a torus of central and lateral radii of R_c and R_l, respectively.

(b) Location of the same points when torus is opened up and transformed to a rectangle. The width and height of the rectangle are $\pi(R_l + R_c)$ and $\pi(R_l - R_c)$, respectively.

Fig. 1. Points can be mapped from a torus to a plane and vice-versa.

of number of hops between source-destination pairs) on the scalability of reactive routing protocols in ad hoc networks with unreliable but stationary nodes is studied in [20], [21]. It is found that the reactive routing protocols may scale infinitely (i.e. the routing overhead does not tend to infinity) with respect to network size for some traffic patterns. The analysis is extended to and similar results are obtained for cluster-based routing algorithms in [18].

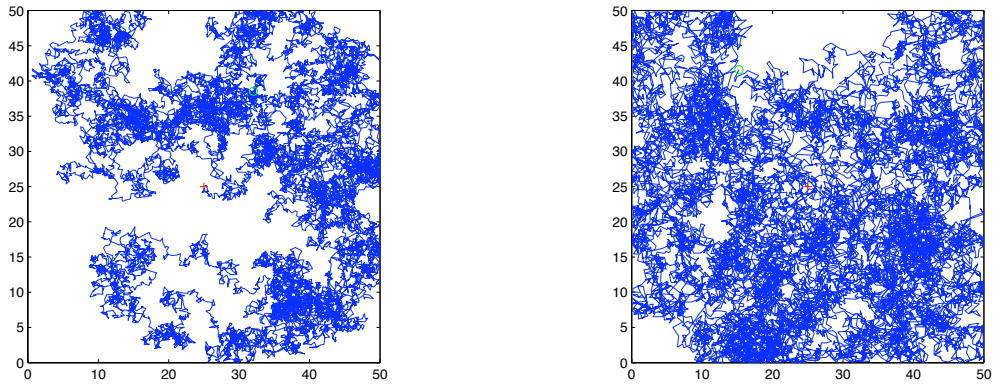
Our work is along the lines of [19] where the authors use an information-theoretic approach to characterize the minimum routing overhead and memory requirements of topology-based (proactive) hierarchical routing protocols for ad hoc networks. The entropy of ad hoc network topologies as well as the entropy rate are used in [19] to find the above mentioned bounds. However, here, the class of protocols considered is position-based protocols and the distortion constraints are also taken into account. This leads to a new problem formulation as rate distortion which was not considered in earlier work and new results on the effect on transport capacity.

It is worth noting that Berger [22] studies discrete-time and continuous-time sampling of Wiener processes. Our work has some intersection with Berger's discrete-time case. However, Berger's results do not directly apply here because: a) we study the movement of nodes under 1-D and 2-D cases, while Bergers paper only applies to the 1-D case, and (b) more importantly, in this work, the source under study is a Wiener process sampled at random time instants, while Berger's paper studies the Wiener process sampled with a fixed time interval, which is a special case of our work (please see Section IV-D). In fact, our result for deterministic arrival case (26) agrees with Berger's (the first equation on page 136 in [22]).

Finally, we note that in this paper we only consider the scenario where the networking protocol requests the position information of a destination as soon as '*a request*' (or, for routing protocols, '*a packet*' to be forwarded) is generated at the source. Thus the potential capacity improvement due to node mobility (achieved at the cost of delay associated with waiting for the destination to move to a nearby location) pointed out in [23], [24] is not applicable to our work.

III. Network Model

The network consists of n mobile nodes that perform Brownian motion. We consider two kinds of network deployments: (i) *One-dimensional case:* nodes located along a circle of perimeter L, and (ii) *Two-dimensional case:* nodes located over a torus of surface area A. The central and lateral radii of the torus are denoted by R_c and R_l respectively. The closed curve and surface are chosen for the study, instead of a finite line or a rectangle, in order to avoid the complexity of modeling the behavior of Brownian motion at boundary points. For space considerations we focus on the two-dimensional case and only list how the result wold change for one-dimensional case.



(a) $\sigma^2 = 30m^2/sec.$ (b) $\sigma^2 = 80m^2/sec.$

Fig. 2. Brownian motion sample trace for a square area of 50m x 50m, initial position at the center of the square, simulation time of 100 sec step time 0.01 sec. Simulation is performed over an infinite plan then mapped to the square area (modulo the side length).

Let t denote time. For the one-dimensional case, nodes perform one-dimensional Brownian motion denoted $B(t)$ with variance parameter σ^2, and hence variance $\sigma^2 t$. For brevity throughout the rest of the paper, we omit the word 'parameter.' For the two-dimensional case, nodes perform a two-dimensional Brownian motion in the complex plane $B(t)$ such that $B(t) = B_{i1}(t) + jB_{i2}(t)$ where $B_{i1}(t)$, $B_{i2}(t)$ are two independent one-dimensional Brownian processes with variance (parameter) σ^2. Notice that it follows from first principles that the variance of the two-dimensional $B(t)$ is $2\sigma^2$.

We assume that $L \gg \sigma^2$ and $R_c, R_l \gg \sigma^2$ for the two-dimensional case. The large dimensions ensure that the nodes do not wrap around the curve or surface during small intervals of time. So if we look at the motion of a node during a small interval of time, then the motion is similar to Brownian motion on an infinite line or plane with the initial node position as the origin. Thus, in the rest of the paper, we treat the motion of nodes during the time scale corresponding to packet inter-arrival times as motion on a plane or straight line. Over time the nodes do not drift apart from each other, as they would on a infinite line or plane, but just keep moving around on the circle or the torus.

Conversely, this may be viewed as if we are observing the Brownian motion of the nodes on an infinite line or plane and mapping their positions back on the circle or torus, respectively. For example, consider a node that performs Brownian motion along the x-axis and whose initial position is the origin. At time t, suppose the node is located at $X(t)$. Then it may be mapped to a point $\mod_L(X(t))$ away from the initial position of the node on the circle, with distance measured in counter-clockwise direction. Similar mapping is possible in the case of torus by considering an infinite plane. Such mapping of points on a torus to points on a plane and vice-versa is shown in Figure 1. Thus instead of keeping track of the positions of nodes on the circle or torus, we use the coordinates of the nodes on x-axis and infinite plane. This scheme works since we are only interested in the change in positions of nodes during packet inter-arrival periods.

The coordinates of nodes are denoted by $X_i(t)$. Hence $X_i(t) = \{X_{i1}(t)\}$ and $X_i(t) = \{X_{i1}(t), X_{i2}(t)\}$ for one and two-dimensional cases respectively. The available location information of node i (at the location server) at time t is denoted by $\hat{X}_i(t)$, hence, $\hat{X}_i(t) = \{\hat{X}_{i1}(t)\}$ and $\hat{X}_i(t) = \{\hat{X}_{i1}(t), \hat{X}_{i2}(t)\}$ for one and two-dimensional cases respectively.

The time instants at which the location information for a node is needed is governed by two processes. These processes are termed 'arrival processes', referring to the arrival or 'requests' for information. In the context of a network protocol used for routing, such instants will be typically the instants of arrival of packets forwarded to a destination. Thus, we interchangeably refer to these processes as packet/query arrival processes. The jth packet destined to (or query about) destination i is generated at a node (source of the jth packet) in the network at time $T_i(j)$, $\forall\ j \geq 1$. Define $T_i(0) \triangleq 0\ \forall\ 1 \leq i \leq n$. For all $j \geq 1$, let $S_j \triangleq T_i(j) - T_i(j-1)$ denote the packet inter-arrival time which is assumed to be independently and identically distributed (i.i.d.), and let S denote any of the S_j. Let the pdf of S be denoted by $f_S(t)$, such that $f_S(t) = 0\ \forall t < 0$ and $E[S]$ exists. Similarly let $\tau_i(k)$ denote the time at which the kth packet is forwarded by node i, with $\tau_i(0) \triangleq 0$. The forwarded packets

include both the packets generated by node i and the packets for which the node acts as an intermediate relaying node. The inter-arrival time of the forwarded packets, $\tau_i(k+1) - \tau_i(k)\ \forall\ k > 0$, is assumed i.i.d with pdf denoted by $f_\tau(t)$.

The i.i.d. assumption about the inter-generation times of packets or the inter-forwarding times may not hold for particular networks, and are adopted for mathematical tractability. However, we expect that for the dependent times case, the overheads can be reduced since correlations will reduce the uncertainty of the locations, and thus the results in this paper are worst case w.r.t. this particular aspect of the network traffic.

We assume a network application or protocol where nodes need to update a location server so that up to date information is maintained about the nodes in the network.

It is assumed that the position of a destination does not change significantly while the location server is being queried by the source and the response is being forwarded through the network. In other words, the time scale of forwarding a packet is much smaller than that required for a significant change in position. This delay is not necessarily critical. For typical networks, the delay is in the order of milliseconds and for average walking or driving speeds this does not amount to significant location change and thus can be neglected. However, there could be some applications where the delay is in seconds or even minutes or more and speeds are in hundreds of km/h (e.g. interplanetary/airborne networks or, in general, delay tolerant networks (DTNs)) and in such cases the model needs to be modified to explicitly model this delay.

Finally, the network is assumed to be always connected, in the graph-theoretic sense, such that nodes can communicate with the desired location servers. It has been shown that this condition can be achieved by proper adjustment of the radius of communication to be above the critical connectivity radius CTR [25]. Specifically, it has been shown that for any mobility model that satisfies the two conditions: (a) obstacle free and (b) bounded area, which are the conditions considered in this paper, CTR is shown to be equal to $c\sqrt{\frac{\ln n}{\pi n}}$ for some constant $c \geq 1$.

IV. Location Update Overhead

In this section we evaluate a lower bound on the minimum rate at which a node must transmit its location information such that the average error in its location stored at the location server is less than ϵ whenever the server is queried. We first introduce the notation and rate-distortion formulation, followed by analysis for one-dimensional and two-dimensional networks. We also evaluate lower bounds for deterministic, uniform and exponential packet arrival processes.

A. Notation and Rate-Distortion Formulation

Definition 1: $D_i(t)$ is the squared-error in the location information of destination i available at its location server at time t, i.e.,

$$D_i(t) = |X_i(t) - \hat{X}_i(t)|^2 \tag{1}$$

where $|X_i(t) - \hat{X}_i(t)| = \sqrt{\sum_{j=1}^{m}\left(X_{ij}(t) - \hat{X}_{ij}(t)\right)^2}$, where m denotes the dimension of the network (e.g. $m = 1(2)$ for one (two) dimensional networks.

Definition 2: $X_i^N = \{X_i(T_1), X_i(T_2), \ldots, X_i(T_N)\}$ is the vector of locations of destination i at time instances T_j, $1 \leq j \leq N$. Similarly $\hat{X}_i^N = \{\hat{X}_i(T_1), \hat{X}_i(T_2), \ldots, \hat{X}_i(T_N)\}$ is the vector of location information at the location server of destination i at time instances T_j, $1 \leq j \leq N$.

Definition 3: $\mathcal{X}_i^N$ and $\hat{\mathcal{X}}_i^N$ are sets of all possible vectors X_i^N and $\hat{X}_i^N$, respectively.

Definition 4: $P_N[x_i^N; \hat{x}_i^N]$ denotes the probability density function at $x_i^N \in \mathcal{X}_i^N$ and $\hat{x}_i^N \in \hat{\mathcal{X}}_i^N$.

Definition 5: $\overline{D}_{iN}$ is defined as

$$\overline{D}_{iN} \triangleq \frac{1}{N}\sum_{j=1}^{N} E[D_i(T_j)] \tag{2}$$

where $E[.]$ denotes the expectation operation.

Definition 6: $\mathcal{P}_N(\epsilon^2)$ is defined as the family of probability distribution functions $P_N[x_i^N; \hat{x}_i^N]$ for which $\overline{D}_{iN} \leq \epsilon^2$.

Definition 7: $R_N(\epsilon^2)$ is defined as the Nth-order rate-distortion function – the minimum rate at which a destination must transmit the location information such that $\overline{D}_{iN} \leq \epsilon^2$. According to [26], $R_N(\epsilon^2)$ is given by

$$R_N(\epsilon^2) = \inf_{P_N \in \mathcal{P}_N(\epsilon^2)} \frac{1}{N} I_{P_N}(X_i^N; \hat{X}_i^N) \tag{3}$$

where $I_{P_N}(X_i^N; \hat{X_i^N})$ is the mutual information between X_i^N and $\hat{X}_i^N$.

The minimum rate at which a destination must update its location information such that the error in location is bounded by ϵ^2 , represented by $R(\epsilon^2)$, is given by

$$R(\epsilon^2) = \liminf_{N \to \infty} R_N(\epsilon^2). \tag{4}$$

B. Location Overhead for One-Dimensional Networks

Lemma 1: The mutual information between X_i^N and $\hat{X}_i^N$ satisfies the following relationship

$$\inf_{P_N \in \mathcal{P}_N} I_{P_N}(X_i^N; \hat{X}_i^N) \geq N R_1(\epsilon^2). \tag{5}$$

The proof follows from the proof of [1, Theorem 3] (see [27] for details).

Theorem 1: The minimum rate at which a node must update its location such that its location distortion at the time its location is queried in the network is bounded from above by ϵ is given by

$$R(\epsilon^2) \geq h(X_{i1}(T_1)) - \frac{1}{2} \log 2\pi e \epsilon^2 \tag{6}$$

where $h(X_{i1}(T_1))$ is the differential entropy of the location of destination i at the time of the first query.

Proof:

From Lemma 1 and the definition of rate distortion function (3) and (4) it follows that

$$R(\epsilon^2) \geq R_1(\epsilon^2) = \inf_{P_1 \in \mathcal{P}_1} I_{P_1}(X_{i1}(T_1); \hat{X}_{i1}(T_1)). \tag{7}$$

Now consider $I_{P_1}(X_{i1}(T_1); \hat{X}_{i1}(T_1))$,

$$\begin{aligned}
I_{P_1}(X_{i1}(T_1); \hat{X}_{i1}(T_1)) &= h(X_{i1}(T_1)) - h(X_{i1}(T_1)|\hat{X}_{i1}(T_1)) \\
&= h(X_{i1}(T_1)) - h(X_{i1}(T_1) - \hat{X}_{i1}(T_1)|\hat{X}_{i1}(T_1)) && (8)\\
&\geq h(X_{i1}(T_1)) - h(X_{i1}(T_1) - \hat{X}_{i1}(T_1)) && (9)\\
&\geq h(X_{i1}(T_1)) - h(\mathcal{N}(0, E[(X_{i1}(T_1) - \hat{X}_{i1}(T_1))^2])) && (10)\\
&\geq h(X_{i1}(T_1)) - \frac{1}{2} \log \left(2\pi e \epsilon^2\right). && (11)
\end{aligned}$$

Here (9) follows from (8) since conditioning does not increase entropy, (10) follows from (9) since for a fixed variance, the normal distribution has the highest differential entropy, and (11) follows from (10) since for $P_1 \in \mathcal{P}_1$, $E[(X_{i1}(T_1) - \hat{X}_{i1}(T_1))^2] \leq \epsilon^2$. Thus,

$$R_1(\epsilon^2) \geq h(X_{i1}(T_1)) - \frac{1}{2} \log \left(2\pi e \epsilon^2\right) \tag{12}$$

and (6) follows directly from it. ■

It should be noted that in some cases $h(X_{i1}(T_1))$ may be less than $\frac{1}{2} \log \left(2\pi e \epsilon^2\right)$. Such situations may occur when σ^2 and/or $E[S]$ is small. Under these circumstances, the change in position of node between two packet generation instances may be comparable to the fidelity criterion (ϵ^2) and hence small number of bits, if any, may be required to represent the change in position of a node between packet generation instances. However, over the course of time the position of a node may change appreciably which may require it to update its location information. When $h(X_{i1}(T_1)) < \frac{1}{2} \log \left(2\pi e \epsilon^2\right)$, the right hand side of inequality (6) is meaningless. A more appropriate inequality is thus

$$R(\epsilon^2) \geq \max \left(h(X_{i1}(T_1)) - \frac{1}{2} \log \left(2\pi e \epsilon^2\right), 0 \right). \tag{13}$$

We discuss the case where $h(X_{i1}(T_1)) < \frac{1}{2}\log\left(2\pi e\epsilon^2\right)$ in detail in Section VII.

Theorem 1 implies that the minimum update rate largely depends on $h(X_{i1}(T_1))$, which in turn depends on two factors: (i) the mobility pattern of the destination node and (ii) the packet inter-arrival process. Let $f_{X_1}(x)$ denote the pdf of $X_{i1}(T_1)$ (without loss of generality, $X_{i1}(T_0) = 0$). For Brownian motion with variance σ^2 and packet inter-arrival time distribution $f_S(t)$, $f_{X_1}(x)$ is given by

$$f_{X_1}(x) = \int_{\tau=0}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2\tau}} e^{-\frac{x^2}{2\sigma^2\tau}} f_S(\tau) d\tau \tag{14}$$

and $h(X_{i1}(T_1))$ is given by

$$h(X_{i1}(T_1)) = -\int_{x=-\infty}^{\infty} f_{X_1}(x) \log\left(f_{X_1}(x)\right) dx. \tag{15}$$

A lower bound on the minimum average overhead incurred by location update information in bits/second, denoted by $U(\epsilon^2)$, is given by

$$U(\epsilon^2) \geq \frac{1}{E[S]} \max\left(h(X_{i1}(T_1)) - \frac{1}{2}\log\left(2\pi e\epsilon^2\right), 0\right). \tag{16}$$

C. Location Overhead for Two-Dimensional Networks

In this section we present the update rate analysis for two-dimensional networks, which is based on the analysis for one-dimensional case. We also evaluate the lower bound for various packet arrival processes and discuss the effect of arrival processes on the minimum update rate.

Brownian motion in two-dimensional space may be decomposed into two independent one-dimensional Brownian motions along x and y coordinates each with a variance $\sigma^2/2$. Thus if $X_i(t) = \{X_{i1}(t), X_{i2}(t)\}$ denote the coordinates of destination i at time t then the distribution of $X_{i1}(t)$ is independent of the distribution of $X_{i2}(t)$. The following Lemma expresses $R\left(\epsilon^2\right)$ in terms of components corresponding to the two coordinates.

Lemma 2: For two-dimensional networks, the rate distortion $R\left(\epsilon^2\right)$ function may be written as

$$R(\epsilon^2) = \min_{0\leq\gamma\leq\epsilon} \left(R^{(1)}(\gamma^2) + R^{(2)}(\epsilon^2 - \gamma^2)\right) \tag{17}$$

where

$$R^{(1)}(\gamma^2) = \lim_{N\to\infty} \inf_{P_N\in\mathcal{P}_N(\gamma^2)} \frac{1}{N} I_{P_N}(X_{i1}^N; \hat{X}_{i1}^N) \tag{18}$$

and

$$R^{(2)}(\epsilon^2 - \gamma^2) = \lim_{N\to\infty} \inf_{P_N\in\mathcal{P}_N(\epsilon^2-\gamma^2)} \frac{1}{N} I_{P_N}(X_{i2}^N; \hat{X}_{i2}^N). \tag{19}$$

Proof: Recall the rate distortion function is

$$R(\epsilon^2) = \lim_{N\to\infty} \inf_{P_N\in\mathcal{P}_N(\epsilon^2)} \frac{1}{N} I_{P_N}(X_i^N; \hat{X}_i^N).$$

Now consider $I_{P_N}(X_i^N; \hat{X}_i^N)$. For two-dimensional networks, this may be written as

$$\begin{aligned}
I_{P_N}(X_i^N; \hat{X}_i^N) &= I_{P_N}(X_{i1}^N, X_{i2}^N; \hat{X}_i^N) \\
&= I_{P_N}(X_{i1}^N; \hat{X}_i^N) + I_{P_N}(X_{i2}^N; \hat{X}_i^N | X_{i1}^N) \\
&= I_{P_N}(X_{i1}^N; \hat{X}_i^N) + I_{P_N}(X_{i2}^N; \hat{X}_i^N) \\
&= I_{P_N}(X_{i1}^N; \hat{X}_{i1}^N) + I_{P_N}(X_{i1}^N; \hat{X}_{i2}^N | \hat{X}_{i1}^N) + I_{P_N}(X_{i2}^N; \hat{X}_{i2}^N) + I_{P_N}(X_{i2}^N; \hat{X}_{i1}^N | \hat{X}_{i2}^N) \\
&= I_{P_N}(X_{i1}^N; \hat{X}_{i1}^N) + I_{P_N}(X_{i2}^N; \hat{X}_{i2}^N)
\end{aligned} \tag{20}$$

where $X_{ik}^N = \{X_{ik}(T_i)\}_{i=1}^N$ and $\hat{X}_{ik}^N = \{\hat{X}_{ik}(T_i)\}_{i=1}^N, k = \{1, 2\}$.

$\overline{D}_{iN} \leq \epsilon^2$ implies

$$\frac{1}{N}\sum_{j=1}^{N} E\left[(X_{i1}(T_j) - \hat{X}_{i1}(T_j))^2\right] + \frac{1}{N}\sum_{j=1}^{N} E\left[(X_{i2}(T_j) - \hat{X}_{i2}(T_j))^2\right] \leq \epsilon^2.$$

The distortion constraint is satisfied if $\frac{1}{N}\sum_{j=1}^{N} E\left[(X_{i1}(T_j) - \hat{X}_{i1}(T_j))^2\right] \leq \gamma^2$ and $\frac{1}{N}\sum_{j=1}^{N} E\left[(X_{i2}(T_j) - \hat{X}_{i2}(T_j))^2\right] \leq \epsilon^2 - \gamma^2$. Combining this and (20), we get

$$R(\epsilon^2) = \min_{0\leq\gamma\leq\epsilon} \lim_{N\to\infty} \left(\inf_{P_N\in\mathcal{P}_N(\gamma^2)} \frac{1}{N} I_{P_N}(X_{i1}^N; \hat{X}_{i1}^N) + \inf_{P_N\in\mathcal{P}_N(\epsilon^2-\gamma^2)} \frac{1}{N} I_{P_N}(X_{i2}^N; \hat{X}_{i2}^N) \right) \tag{21}$$

which leads to (17). ■

From Theorem 1, it follows that

$$R^{(1)}(\gamma^2) \geq h(X_{i1}(T_1)) - \frac{1}{2}\log\left(2\pi e\gamma^2\right) \tag{22}$$

and

$$R^{(2)}(\epsilon^2 - \gamma^2) \geq h(X_{i2}(T_1)) - \frac{1}{2}\log\left(2\pi e(\epsilon^2 - \gamma^2)\right). \tag{23}$$

From (22) and (23), it is clear that the right hand side of (17) is minimized for $\gamma^2 = \epsilon^2/2$ for $X_{i1}(T_1)$ and $X_{i2}(T_1)$ i.i.d. with the same variance. This leads to the following theorem.

Theorem 2: In order to ensure that the average error in location information used for forwarding packets is less than ϵ, the lower bound on the location update rate (in bits per packet) for a two-dimensional network is given by

$$R(\epsilon^2) \geq \max\left(h(X_{i1}(T_1)) + h(X_{i2}(T_1)) - \log\left(\pi e\epsilon^2\right), 0\right) \tag{24}$$

and the overhead incurred in bits/second ($U(\epsilon)$) is given by

$$U(\epsilon^2) \geq \frac{1}{E[S]} \max\left(h(X_{i1}(T_1)) + h(X_{i2}(T_1)) - \log\left(\pi e\epsilon^2\right), 0\right). \tag{25}$$

D. Comparison of Location Update Rates for Various Arrival Processes

We evaluate the bound on location update rates for various arrival processes, since the results for some of these cases are available as closed form, yielding additional insights, and some have extremal nature. We primarily consider the case of two-dimensional networks, while we only comment about how the result would change in one-dimensional case. The results for one-dimensional case are available in [27].

Suppose that requests for location information of some node is generated in the network after every T (i.e. deterministic arrival rate), that is $t = kT$, $k = 1, 2, \ldots \infty$, thus,

$$h(X_{i1}(T_1)) = h(X_{i2}(T_1)) = \frac{1}{2}\log\left(\pi e\sigma^2 T\right)$$

and the lower bound on the location update rate is given by

$$R(\epsilon^2) \geq \max\left(\log\left(\frac{\sigma^2 T}{\epsilon^2}\right), 0\right), \tag{26}$$

$$U(\epsilon^2) \geq \frac{1}{T}\left(\log\left(\frac{\sigma^2 T}{\epsilon^2}\right), 0\right). \tag{27}$$

For the deterministic inter-arrival time the lower bound is similar to the minimum number of bits required to represent a Gaussian random variable with variance $\sigma^2 T$ subject to the constraint that the expected squared-error is less than ϵ^2. This is because the change in position of a node during a packet inter-arrival interval is indeed a Gaussian random variable with variance $\sigma^2 T$.

Notice that $R(\epsilon^2)$ increases with both σ^2 and T. This is because larger σ^2 and T would imply larger uncertainty in the change in position of the destination during the packet arrival duration and hence more bits are required to represent the destination's position for each new request. $U(\epsilon^2)$ also increases with σ^2 for the same reason. However, $U(\epsilon^2)$ decreases with

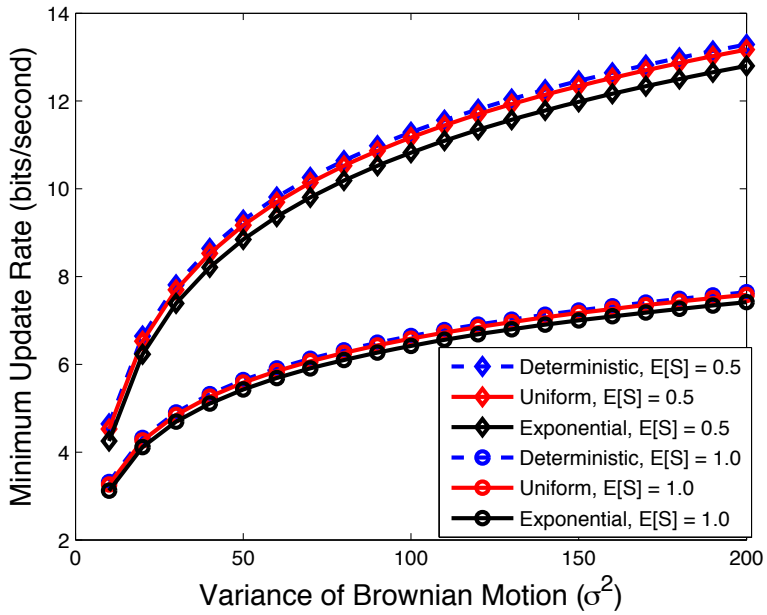


Fig. 3. Update rate vs. variance of Brownian motion; $\epsilon = 1.0$.

increase in T. This is because the increase in update rate due to higher uncertainty associated with high T is over-compensated by the fact that larger T implies that updates have to be made less often.

The behavior of $R(\epsilon^2)$ and $U(\epsilon^2)$ is similar to that observed in the one-dimensional case. In fact the minimum rate is simply the double of what is observed in the one-dimensional case. This is not surprising since the change in x-coordinate of the destination node is independent of the change in the y-coordinate of the destination. The change has variance $\sigma^2/2$ rather than σ^2 as in the case of one-dimensional motion thus one might expect the minimum rate in 2-D to be less than 2 times the minimum rate in 1-D case. However this decrease in entropy due to decreased variance is compensated by the fact that allowable squared error in both the coordinates is also decreased, leading to a factor of 2.

For the case of *uniform inter-arrival time distribution*, among all continuous time distributions with a given finite support, the uniform distribution has the highest entropy. Thus uniform distribution maximizes the uncertainty of packet arrival instances and thus would lead to maximum position update rate among all distributions with the same finite support. Finally, the the *exponential distribution* is widely used to model arrival processes, and also it has the highest entropy among all the continuous time distribution with base $[0, \infty)$ for a given mean.

Figures 3 and 4 show the plot of the lower bound on $U(\epsilon^2)$ against σ^2 and $E[S]$. It is observed that for high σ^2 and low $E[S]$, the rate at which a source must update its location servers becomes very high. Also it is observed that the rate required for deterministic packet arrival is higher than that required for uniform and exponential arrival processes. In fact the update rate for deterministic packet arrival process is higher than any other packet arrival process with the same mean inter-arrival time. Consider a packet arrival process with pdf $f_S(t)$ and mean $E[S]$, then $\mathrm{Var}\left(X_{i1}(T_1)\right)$ and $\mathrm{Var}\left(X_{i2}(T_1)\right)$ are given by

$$\int_0^\infty \int_{-\infty}^\infty x^2 \frac{1}{\sqrt{\pi\sigma^2 t}} e^{-\frac{x^2}{\sigma^2 t}} \mathrm{d}x f_S(t)\mathrm{d}t = \int_0^\infty \frac{\sigma^2 t}{2} f_S(t)\mathrm{d}t = \frac{\sigma^2 E[S]}{2}. \tag{28}$$

This implies that notwithstanding the packet arrival process, the variance of the change in location between two packet arrival instances depends only on σ and $E[S]$. For the deterministic packet arrival process, $X_{i1}(T_1)$ and $X_{i2}(T_1)$ are Gaussian random variables. Since among random variables with the same variance Gaussian random variable has the highest entropy, deterministic packet arrival leads to the highest update rate.

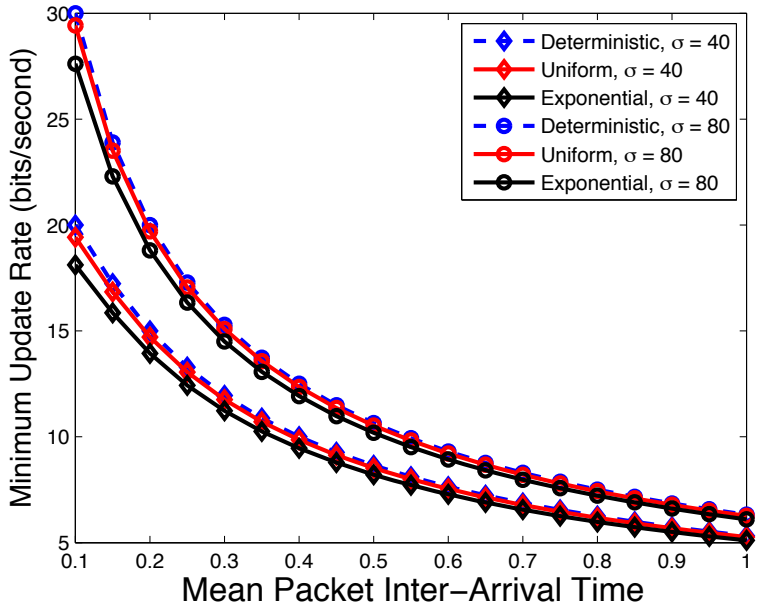


Fig. 4. Update rate vs. mean packet inter-arrival time; $\epsilon = 1.0$.

V. Beacon Overhead

In this section we evaluate a lower bound on the minimum rate at which the nodes need to transmit beacons so that neighbors maintain a consistent neighborhood information. We then bound overhead in terms of bits per second. We first consider Brownian motion in one dimension and later extend the results to two-dimensional case.

The question addressed here is to find the minimum rate at which a node j must transmit beacons such that a neighboring node i will know whether j is a neighbor of i or not when it has a packet to send.

A. Notation and Minimum Beacon Rate Formulation

All the notations defined in subsections III and IV-A are used throughout this section. In this subsection we define some additional notations for the analysis.

Definition 8: $N_i(t)$ is the set of nodes that belong to the neighborhood of node i. That is

$$N_i(t) = \{j : |X_i(t) - X_j(t)| \leq r, 1 \leq j \leq n, j \neq i\}. \tag{29}$$

Definition 9: $\hat{N}_i(t)$ is the set of nodes that the node i *perceives* to be its neighbors.

The set $\hat{N}_i(t)$ is constructed by node i based on the beacons it receives. A node that belongs to $\hat{N}_i(t)$ may be excluded from $\hat{N}_i(t+\tau)$ if sufficient beacons are not received from the node during time interval $[t, t+\tau]$. Similarly, a node not belonging to $\hat{N}_i(t)$ may be included in $\hat{N}_i(t+\tau)$ if sufficient beacons are received from the node during time interval $[t, t+\tau]$. The deviation of $\hat{N}_i(t)$ from $N_i(t)$ depends on the rate at which the nodes transmit beacons.

Definition 10: $Z_{ij}(t)$ and $\hat{Z}_{ij}(t)$ $(1 \leq i, j \leq n,\ i \neq j)$ are indicator random variables, defined in the following manner

$$Z_{ij}(t) = \begin{cases} 1, & \text{if } j \in N_i(t) \\ 0, & \text{otherwise} \end{cases} \tag{30}$$

and

$$\hat{Z}_{ij}(t) = \begin{cases} 1, & \text{if } j \in \hat{N}_i(t) \\ 0, & \text{otherwise} \end{cases} \tag{31}$$

In other words, $Z_{ij}(t)$ equals 1 if node j belongs to the neighborhood of node i at time t. Note that $Z_{ij}(t)$ is a symmetric relation, i.e. $Z_{ij}(t) = Z_{ji}(t)$. On the other hand, $\hat{Z}_{ij}(t)$ is 1 if node node i perceives node j to be its neighbor at time t. Unlike

$Z_{ij}(t)$, $\hat{Z}_{ij}(t)$ is not necessarily symmetric. That is $\hat{Z}_{ij}(t)$ may be 0 although $\hat{Z}_{ji}(t)$ is 1. This may happen because the beacon transmission rate of i is high enough to allow j to maintain a consistent neighborhood set while the beacon transmission rate of node j is not high enough to allow node i to maintain a consistent neighborhood set. The deviation of the perceived neighborhood, $\hat{N}_i(t)$, from the actual neighborhood, $N_i(t)$, is reflected by the deviation of $\hat{Z}_{ij}(t)$ from $Z_{ij}(t)$.

Definition 11: The difference of $\hat{Z}_{ij}(t)$ and $Z_{ij}(t)$ is defined as $E_{ij}(t)$, i.e.

$$E_{ij}(t) = Z_{ij}(t) - \hat{Z}_{ij}(t). \tag{32}$$

$E_{ij}(t) = 0$ implies that node i has accurate information about whether j belongs to its neighborhood or not. It is not necessary that $E_{ij}(t) = 0$ for all t, however it is desirable that $E_{ij}(t) = 0$ with high probability at all time instances when node i has a packet to forward. This is because correct neighborhood information is highly critical for node i to make correct forwarding decisions.

We can now state the minimum beacon rate problem in the following manner.

Minimum beacon rate problem: To find the minimum rate at which node j must transmit beacons such that

$$P[E_{ij}(\tau_i(k)) = 0] \geq 1 - \delta \ \ \forall \ 1 \leq i \neq j \leq n, 1 \leq k < \infty. \tag{33}$$

In order to formulate the above minimum beacon rate problem as a rate distortion problem we present two more definitions.

Definition 12: Let the vectors Z_{ij}^N and $\hat{Z}_{ij}^N$ be defined in the following manner

$$Z_{ij}^N \triangleq \{Z_{ij}(\tau_i(1)), Z_{ij}(\tau_i(2)), \ldots, Z_{ij}(\tau_i(N))\}. \tag{34}$$

$$\hat{Z}_{ij}^N \triangleq \{\hat{Z}_{ij}(\tau_i(1)), \hat{Z}_{ij}(\tau_i(2)), \ldots, \hat{Z}_{ij}(\tau_i(N))\}. \tag{35}$$

Definition 13: Let $\mathcal{P}_N^{(b)}(\delta)$ denote the family of joint probability distribution function of Z_{ij}^N and $\hat{Z}_{ij}^N$ such that $P[E_{ij}(\tau_i(k)) = 0] \leq 1 - \delta \ \forall \ 1 \leq k \leq N$.

The superscript in $\mathcal{P}_N^{(b)}(\delta)$ is used in order to distinguish the notation from the one used in the previous section. This superscript will be used for similar purpose in the rest of this section.

Thus the minimum beacon rate, $R^{(b)}(\delta)$, may be expressed in the following manner

$$R^{(b)}(\delta) = \liminf_{N \to \infty} R_N^{(b)}(\delta) \tag{36}$$

where

$$R_N^{(b)}(\delta) = \inf_{P_N \in \mathcal{P}_N^{(b)}(\delta)} \frac{1}{N} I_{P_N}(Z_{ij}^N; \hat{Z}_{ij}^N) \tag{37}$$

and $I_{P_N}(Z_{ij}^N; \hat{Z}_{ij}^N)$ is the mutual information between Z_{ij}^N and $\hat{Z}_{ij}^N$. In the next subsection we evaluate a lower bound on $R^{(b)}(\delta)$.

B. Beacon Rate Analysis for One-Dimension Networks

We first present the lower bound on the beacon rate of a node for the one-dimensional case and then extend the result to the two-dimensional case.

Lemma 3: The minimum beacon rate of node j, $R^{(b)}(\delta)$ is greater than or equal to $R_1^{(b)}(\delta)$, that is

$$R^{(b)}(\delta) \geq R_1^{(b)}(\delta). \tag{38}$$

The proof of the above Lemma is similar to that of Lemma 1.

Following the result of Lemma 3, we need to find a lower bound on $R_1^{(b)}(\delta)$ in order to bound $R^{(b)}(\delta)$. We first find a lower bound on $R_1^{(b)}$ for the case when $Z_{ij}(0) = 1$ and then find the bound for the case when $Z_{ij}(0) = 0$. The bounds for the these two cases provide answers to the following two questions: (i) If i and j are neighbors at time $t = 0$, then at what rate must j transmit beacons such that node i knows whether j is neighbor of i or not when it has a packet to send at $t = \tau_i(1)$; (ii) If i and j are not neighbors at $t = 0$, then at what rate must j transmit beacons such that i knows whether j is a neighbor of

i or not when it has a packet to send at $t = \tau_i(1)$. Since the constraint $P[E_{ij}(\tau_i(1)) = 0] \geq 1 - \delta$ has to be satisfied for all $1 \leq i \leq n$, whether i is a neighbor of j or not at $t = 0$, the lower bound is the maximum of the two cases.

Lemma 4: If $Z_{ij}(0) = 1$, then $R_1^{(b)}(\delta)$ is bounded by

$$R_1^{(b)}(\delta) \geq \sup_{X_j(0) \in L_i} H(Z_{ij}(\tau_i(1))) - \mathcal{H}\left(\frac{\delta}{2}, 1-\delta, \frac{\delta}{2}\right) \tag{39}$$

where

$$\mathcal{H}(p_1, p_2, p_3) \triangleq -\sum_{i=1}^{3} p_i \log p_i \tag{40}$$

and, in particular,

$$\mathcal{H}\left(\frac{\delta}{2}, 1-\delta, \frac{\delta}{2}\right) = -\delta \log\left(\frac{\delta}{2}\right) - (1-\delta)\log(1-\delta) \tag{41}$$

and

$$L_i \triangleq [X_i(0) - r, X_i(0) + r] \tag{42}$$

i.e. L_i is the set of possible positions of j at time $t = 0$ such that $Z_{ij}(0) = 1$.

Proof: Recall that $R_1^{(b)}(\delta)$ is given by

$$R_1^{(b)}(\delta) = \inf_{P_1 \in \mathcal{P}_1(\delta)} I_{P_1}(Z_{ij}(\tau_1); \hat{Z}_{ij}(\tau_1)). \tag{43}$$

Now $I_{P_1}(Z_{ij}(\tau_1); \hat{Z}_{ij}(\tau_1))$ is given by

$$\begin{aligned} I_{P_1}(Z_{ij}(\tau_1); \hat{Z}_{ij}(\tau_1)) &= H(Z_{ij}(\tau_1)) - H(Z_{ij}(\tau_1)|\hat{Z}_{ij}(\tau_1)) && (44)\\ &= H(Z_{ij}(\tau_1)) - H(Z_{ij}(\tau_1) - \hat{Z}_{ij}(\tau_1)|\hat{Z}_{ij}(\tau_1)) && (45)\\ &\geq H(Z_{ij}(\tau_1)) - H(Z_{ij}(\tau_1) - \hat{Z}_{ij}(\tau_1)) && (46)\\ &= H(Z_{ij}(\tau_1)) - H(E_{ij}(\tau_1)). && (47) \end{aligned}$$

We know that the probability distribution of $E_{ij}(\tau_1)$ is given by

$$E_{ij}(\tau_1)) = \begin{cases} -1, & \text{w.p. } p_1 \\ 0, & \text{w.p. } p_2 \\ 1, & \text{w.p. } p_3 \end{cases} \tag{48}$$

where $p_2 \geq 1 - \delta$, $p_1 + p_3 \leq \delta$ and $p_1 + p_2 + p_3 = 1$ (since $P_1 \in \mathcal{P}_1(\delta)$). Under these constraints, for small δ (i.e. $\delta < \frac{2}{3}$), $H(E_{ij}(\tau_1))$ is maximized when $p_2 = 1 - \delta$ and $p_1 = p_3 = \delta/2$, when $H(E_{ij}(\tau_1)) = \mathcal{H}\left(\frac{\delta}{2}, 1-\delta, \frac{\delta}{2}\right)$. In the case that $\delta > \frac{2}{3}$, the distribution that maximizes the entropy is $p_i = \frac{1}{3} \, \forall \, i = 1, 2, 3$. Define

$$\mathcal{H}^*(\delta) = \begin{cases} \mathcal{H}\left(\frac{\delta}{2}, 1-\delta, \frac{\delta}{2}\right), & \delta \leq \frac{2}{3} \\ \log 3, & \delta > \frac{2}{3} \end{cases} \tag{49}$$

Thus

$$I_{P_1}(Z_{ij}(\tau_1); \hat{Z}_{ij}(\tau_1)) \geq H(Z_{ij}(\tau_1)) - \mathcal{H}^*(\delta). \tag{50}$$

Now $H(Z_{ij}(\tau_1))$ depends on the position of node j at $t = 0$, $X_j(0)$. We know that $j \in N_i(0)$ which implies that $X_j(0) \in L_i$. Depending upon $f_\tau(t)$, $H(Z_{ij}(\tau_1))$ is maximized for some $X_j(0) = x \in L_i$. In order to ensure that $P[E_{ij}(\tau_i(k)) = 0] \geq 1 - \delta$ $\forall \, i$, the beacon rate must take care of this worst case. Thus

$$I_{P_1}(Z_{ij}(\tau_1); \hat{Z}_{ij}(\tau_1)) \geq \sup_{X_j(0) \in L_i} H(Z_{ij}(\tau_i(1))) - \mathcal{H}^*(\delta). \tag{51}$$

From the above equation and (43) we get (39). ■

We now outline how to evaluate $H(Z_{ij}(\tau_i(1)))$. Without loss of generality we may assume that $X_i(0) = 0$. So if $Z_{ij}(0) = 1$,

then $X_j(0) = l$ where $-r \leq l \leq r$. From the point of reference of node i, node j performs Brownian motion with variance $2\sigma^2$. So

$$P[Z_{ij}(\tau_i(1)) = 1 | X_j(0) = l \in L_i, \tau_i(1) = \tau'] = \frac{1}{2}\text{erf}\left(\frac{r-l}{\sqrt{4\sigma^2\tau'}}\right) + \frac{1}{2}\text{erf}\left(\frac{r+l}{\sqrt{4\sigma^2\tau'}}\right). \tag{52}$$

Now using the fact that the inter-arrival time of packets to be served by i has distribution $f_\tau(t)$, we get

$$p(l) \triangleq P[Z_{ij}(\tau_i(1)) = 1 | X_j(0) = l \in L_i] = \frac{1}{2}\int_{t=0}^{\infty} \text{erf}\left(\frac{r-l}{\sqrt{4\sigma^2 t}}\right) f_\tau(t)dt + \frac{1}{2}\int_{t=0}^{\infty} \text{erf}\left(\frac{r+l}{\sqrt{4\sigma^2 t}}\right) f_\tau(t)dt. \tag{53}$$

We know that $H(Z_{ij}(\tau_i(1))) = \mathcal{H}(P[Z_{ij}(\tau_i(1)) = 1])$, where $\mathcal{H}(x) = -x\log(x) - (1-x)\log(1-x)$ $(0 \leq x \leq 1)$. We know that $\mathcal{H}(x)$ is maximum at $x = 0.5$, symmetric about $x = 0.5$ and is strictly increasing and decreasing in the interval $[0, 0.5)$ and $(0.5, 1]$ respectively.

Definition 14: Let $l^\star$ be defined as

$$l^\star \triangleq \arg\min_{-r \leq l \leq r} |P[Z_{ij}(\tau_i(1)) = 1 | X_j(0) = l] - 0.5|. \tag{54}$$

In other words, $l^\star$ is the value of $X_j(0)$ for which $H(Z_{ij}(\tau_i(1)))$ is maximized. This leads us to the following Corollary.

Corollary 1: The minimum beacon transmission rate of node j, denoted by $R^{(b_1)}(\delta)$, such that with probability at least $1-\delta$ the current neighbors of j know whether j belongs to their neighborhood at the time of forwarding next packet equals

$$R^{(b_1)}(\delta) \geq \mathcal{H}(p(l^\star)) - \mathcal{H}\left(\frac{\delta}{2}, 1-\delta, \frac{\delta}{2}\right) \text{ beacons/msg} \tag{55}$$

where $p(l^\star) = P[Z_{ij}(\tau_i(1)) = 1 || X_j(0) - X_i(0)| = l^\star]$ and $l^\star$ is given by (54).

We now turn our attention to the second question i.e. at what rate should j transmit such that with probability at least $1-\delta$ all the nodes that are not neighbors of j at $t = 0$ know whether j belongs to their neighborhood or not when they have a packet to forward?

It is easy to see that Lemma 4 holds even if $Z_{ij}(0) = 0$. We formally state a similar Lemma without proof for the case when $Z_{ij}(0) = 0$.

Lemma 5: If $Z_{ij}(0) = 0$, then $R_1^{(b)}(\delta)$ is bounded by

$$R_1^{(b)} \geq \max_{X_j(0) \in L_i'} H(Z_{ij}(\tau_i(1))) - \mathcal{H}^*(\delta) \tag{56}$$

where $\mathcal{H}\left(\frac{\delta}{2}, 1-\delta, \frac{\delta}{2}\right)$ is given by (41) and L_i' is the set of possible positions of node j at $t = 0$ such that $Z_{ij} = 0$, i.e.,

$$L_i' \triangleq \{x : |x - X_i(0)| \geq r\}. \tag{57}$$

Again, without loss of generality we assume that $X_i(0) = 0$. Let $X_j(0) = l$, such that $|l| \geq r$, i.e. j does not belong to the neighborhood of i at time $t = 0$. $P[Z_{ij}(\tau_i(1)) = 1 | X_j(0) = l, |l| \geq r, \tau_i(1) = \tau']$ is given by

$$P[Z_{ij}(\tau_i(1)) = 1 | X_j(0) = l, |l| \geq r, \tau_i(1) = \tau'] = \frac{1}{2}\text{erf}\left(\frac{|l|+r}{\sqrt{4\sigma^2\tau'}}\right) - \frac{1}{2}\text{erf}\left(\frac{|l|-r}{\sqrt{4\sigma^2\tau'}}\right). \tag{58}$$

Thus

$$p'(l) \triangleq P[Z_{ij}(\tau_i(1)) = 1 | X_j(0) = l, |l| \geq r] = \frac{1}{2}\int_0^{\infty} \text{erf}\left(\frac{|l|+r}{\sqrt{4\sigma^2 t}}\right) f_\tau(t)dt - \frac{1}{2}\int_0^{\infty} \text{erf}\left(\frac{|l|-r}{\sqrt{4\sigma^2 t}}\right) f_\tau(t)dt. \tag{59}$$

Note that that $p'(l) \leq 0.5 \, \forall \, |l| \geq r$ and its value is maximized for $|l| = r$. Since $\mathcal{H}(x)$ is an increasing function of x in the interval $[0, 0.5)$, $|l| = r$ maximizes the value of $H(Z_{ij}(\tau_i(1))) = \mathcal{H}(P[Z_{ij}(\tau_i(1)) = 1 | X_j(0) = l, |l| \geq r])$. This leads to the following Corollary.

Corollary 2: The minimum beacon transmission rate of node j, denoted by $R^{(b_2)}(\delta)$, such that with probability at least $1-\delta$ the nodes that are not current neighbors of j know whether j belongs to their neighborhood at time of forwarding the next packet equals

$$R^{(b_2)}(\delta) \geq \mathcal{H}(p'(r)) - \mathcal{H}^*(\delta) \text{ beacons/msg} \tag{60}$$

where

$$p'(r) = \frac{1}{2}\int_0^\infty \text{erf}\left(\frac{r}{\sqrt{\sigma^2 t}}\right) f_\tau(t)dt.$$

To summarize, Corollary 1 provides a lower bound on the beacon transmission rate such that each of the current neighbors are able to maintain consistent neighborhood information with high probability. Corollary 2 provides a lower bound on the beacon transmission rate of a node such that all nodes that are not neighbor of the node know with high probability if the node joins their neighborhood. The minimum beacon transmission rate is therefore the maximum of the two rates given by (55) and (60).

Finally, note that according to the definition of $l^\star$, $\mathcal{H}(p(l^\star)) \geq \mathcal{H}(p'(r))$. (This can be shown by noting that $p'(r) = p(r)$.) This implies that the lower bound on $R^{(b_2)}(\delta)$ in (60) is always less than the lower bound on $R^{(b_1)}(\delta)$ in (55). Thus the lower bound on the minimum beacon transmission rate is given by the lower bound on $R^{(b_1)}(\delta)$ in (55). The following theorem formally states this discussion.

Theorem 3: A lower bound on the minimum beacon transmission rate of a node such that the constraint in (33) is satisfied is given

$$\begin{aligned} R^{(b)}(\delta) &\geq \max\left(R^{(b_1)}(\delta), R^{(b_2)}(\delta)\right) && (61)\\ &\geq \mathcal{H}(p(l^\star)) - \mathcal{H}^*(\delta) \ \text{ beacons/msg} && (62)\end{aligned}$$

where $R^{(b_1)}(\delta)$, $R^{(b_2)}(\delta)$ and $l^\star$ are given by (55), (60) and (54) respectively.

When $\mathcal{H}(p(l^\star)) < \mathcal{H}^*(\delta)$, left hand side of (62) will be meaningless. Thus (62) may be expressed as

$$R^{(b)}(\delta) \geq \max\left(\mathcal{H}(p(l^\star)) - \mathcal{H}^*(\delta), 0\right) \ \text{ beacons/msg}. \tag{63}$$

The minimum overhead in bits/second, denoted by $U^{(b)}(\delta)$, is given by

$$U^{(b)}(\delta) \geq \frac{B}{E[\tau]}\max\left(\mathcal{H}(p(l^\star)) - \mathcal{H}^*(\delta), 0\right) \ \text{ bits/second} \tag{64}$$

where $E[\tau]$ is the expected packet inter-arrival time and B is the size of beacon packet in bits.

C. Beacon Rate Analysis for Two-Dimensional Networks

In this subsection we extend the minimum beacon rate analysis to two-dimensional networks. For a arbitrary node pair i and j, we choose an orthogonal coordinate system such that $X_{i1}(0) = X_{i2}(0) = 0$, $X_{j1}(0) = l$, and $X_{j2}(0) = 0$. That is, the origin of the coordinate system corresponds to the position of node i at $t = 0$ and the x-axis of the coordinate system corresponds to the line joining the position of nodes i and j at $t = 0$. It can be easily verified that a Brownian motion with variance σ^2 can be decomposed into two independent Brownian motions with variance $\sigma^2/2$ along each axis. Also note that Lemmas 3 and 4 hold for the two-dimensional case as well and may be proved in a similar manner. Thus the minimum beacon rate, $R^{(b)}(\delta)$, satisfies the following relationship

$$R^{(b)}(\delta) \geq H\left(Z_{ij}(\tau_i(1))\right) - \mathcal{H}^*(\delta). \tag{65}$$

Similar to the approach in the last section, we proceed by individually considering the cases $Z_{ij}(0) = 1$ and $Z_{ij}(0) = 0$. For the case when $Z_{ij}(0) = 1$, the probability that j is in the neighborhood of i when i has a packet to send ($p(l)$) is given by

$$\begin{aligned} p(l) \triangleq P\left[Z_{ij}(\tau) = 1 | X_{j1}(0) = l, X_{j2}(0) = 0, |l| \leq r\right] = \int_{x=-r}^{x=r} P\left[X_{j1}(\tau_i(1)) \in x + dx | X_{j1}(0) = l\right] \cdot \\ P\left[-\sqrt{r^2 - x^2} \leq X_{j2}(\tau_i(1)) \leq \sqrt{r^2 - x^2} | X_{j2}(0) = 0\right]. \end{aligned}$$

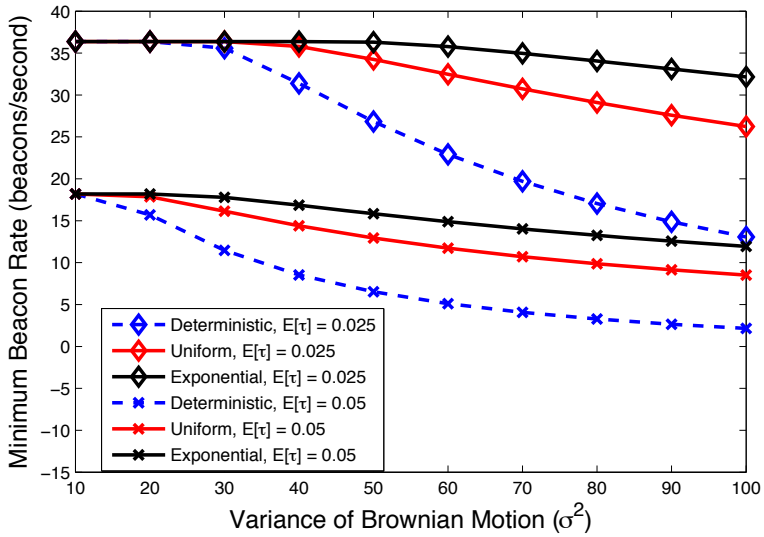


Fig. 5. Beacons per second versus variance of Brownian motion; $\delta = 0.01$, $r = 1.0$.

Relative to node i, node j performs Brownian motion with variance $2\sigma^2$. Thus

$$P\left[X_{j1}(\tau) \in x + dx | X_{j1}(0) = l\right] = \frac{1}{\sqrt{2\pi\sigma^2\tau}} \exp\left(-\frac{(l-x)^2}{2\sigma^2\tau}\right) dx$$

and

$$P\left[-\sqrt{r^2 - x^2} \leq X_{j2}(\tau) \leq \sqrt{r^2 - x^2} | X_{j2}(0) = 0\right] = \text{erf}\left(\frac{\sqrt{r^2 - x^2}}{\sqrt{2\sigma^2\tau}}\right).$$

Therefore,

$$p(l) = \int_0^\infty \int_{-r}^{r} \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left(-\frac{(l-x)^2}{2\sigma^2 t}\right) \text{erf}\left(\frac{\sqrt{r^2 - x^2}}{\sqrt{2\sigma^2 t}}\right) f_\tau(t) dx\, dt. \tag{66}$$

Thus in order to satisfy (33) at all neighbors that are neighbor at time 0, node j must transmit beacon at a rate higher than

$$\mathcal{H}\left(p(l^\star)\right) - \mathcal{H}^*(\delta) \tag{67}$$

where $l^\star$ is given by (54).

Now consider the case when j does not belong to the neighborhood of i at $t = 0$. It can be easily verified that the probability that $Z_{ij}(\tau_i(1)) = 1$ given that $Z_{ij}(0) = 0$ ($p'(l)$) is given by the same expression as $p(l)$ in equation (66). $p'(l)$ increases with decrease in $|l|$ and is maximized for $|l| = r$. For other values of $l > r$, $p'(l) < 0.5$. Thus similar to the one-dimensional networks, the beacon transmission rate is determined by the rate required to satisfy (33) at the initial neighbors. This leads the following theorem.

Theorem 4: The lower bound on the minimum beacon transmission rate of a node such that the constraint in equation (33) is satisfied is given

$$R^{(b)}(\delta) \geq \max\left(\mathcal{H}\left(p(l^\star)\right) - \mathcal{H}^*(\delta), 0\right) \text{ beacons/pkt} \tag{68}$$

where $p(l)$ and $l^\star$ are given by (66) and (54) respectively. The beacon transmission overhead in bits per second, $U^{(b)}(\delta)$, is given by

$$U^{(b)}(\delta) \geq \frac{B}{E[\tau]} \max\left(\mathcal{H}(p(l^\star)) - \mathcal{H}^*(\delta), 0\right). \tag{69}$$

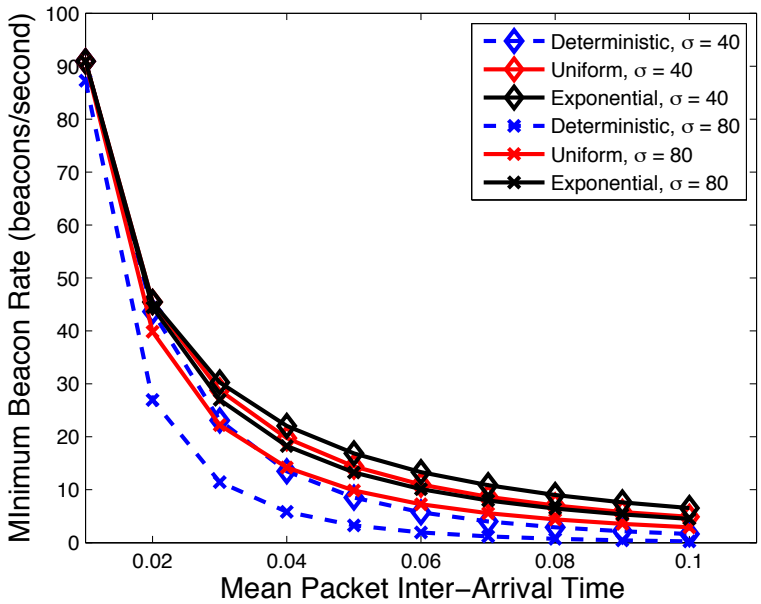


Fig. 6. Beacons per second versus the mean inter-arrival time of packets to be forwarded at a node; $\delta = 0.01$, $r = 1.0$.

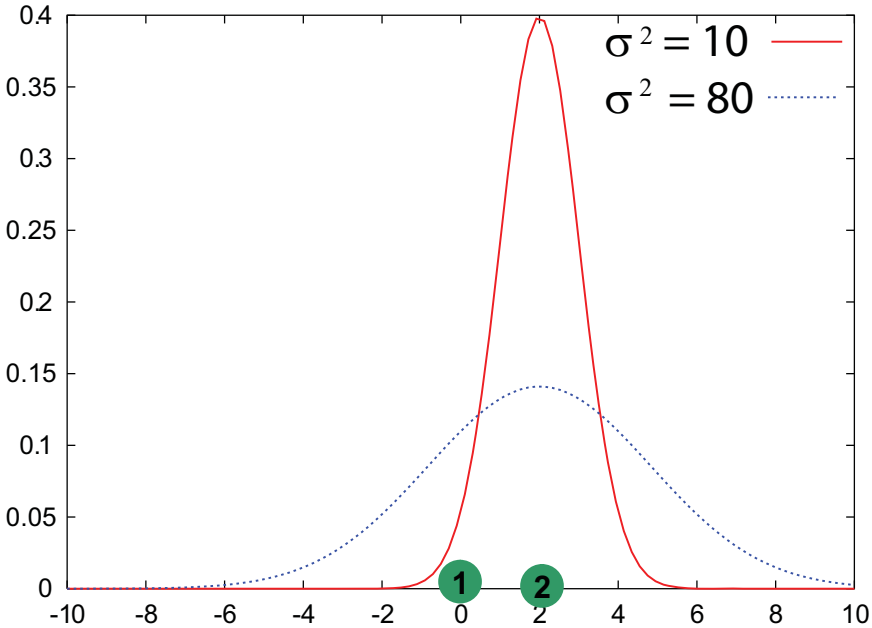


Fig. 7. The pdf of position of node 2 with respect to node 1 given their initial positions.

D. Comparison of Beacon Transmission Rates for Various Arrival Processes

Closed form expression for the integral in (66) cannot [2] be found. So we use numerical computations to evaluate $R^{(b)}(\delta)$ for deterministic, uniform and exponential packet arrival processes. Figures 5 and 6 show plots of minimum beacon rate in bits per second. Figure 5 shows the plot of minimum beacon transmission rate against variance of Brownian motion for different mean packet inter-arrival times. It is observed that for low variance the rate is almost constant, while as the the variance increases the rate starts decreasing. When the variance of Brownian motion is very small, the variance of the change in position of a node within a packet arrival epoch is also small. For this case, $l^\star = r$ and $p(l^\star) \approx 0.5$ which leads to high beacon rate. As the variance increases, the probability that two neighbors remain neighbors at the end of a packet arrival epoch is very small,

[2]Please see [27] for simplified expressions for the one-dimensional case for the case of deterministic, uniform and exponential arrival rates.

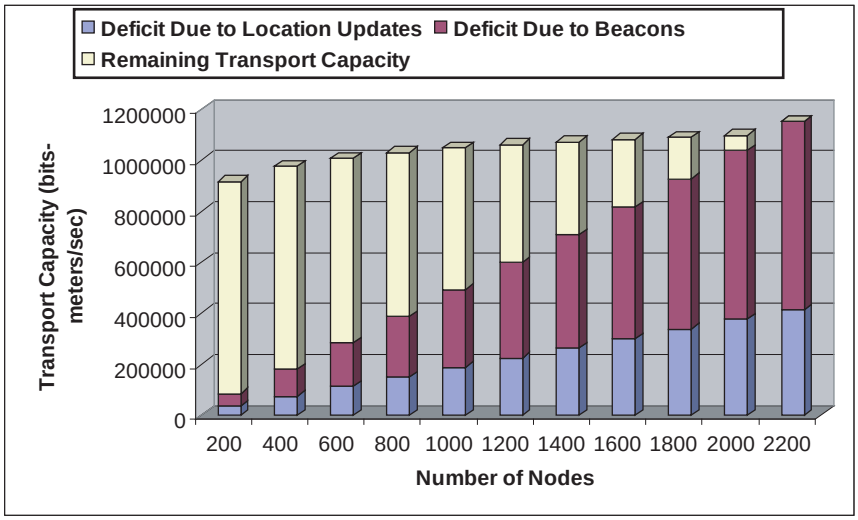


Fig. 8. Capacity deficit versus number of nodes for $\sigma = 80$, $\epsilon = 1.0$, $\delta = 0.01$, $A = 100$.

no matter what the initial position of nodes might be. That is, when σ^2 is high, $p(l) < 0.5 \ \forall \ l$, which leads to low beacon rate when variance is high. This is illustrated in Figure 7. Suppose the two nodes shown in the figure have communication range of two units. Initially nodes 1 and 2 are located at $X = 0$ and $X = 2$ respectively. With respect to node 1, the pdf of position of node 2 when the next packet is forwarded is shown for $\sigma^2 = 10$ and $\sigma^2 = 80$. It is observed that for $\sigma^2 = 10$ the probability that nodes 2 will be in neighborhood of node 1 at the next packet arrival instant is approximately 0.5. This corresponds to large uncertainty which leads to high beacon update rate. On the other hand when $\sigma^2 = 80$ the probability that nods 1 and 2 are neighbors at the next packet arrival instant is less than 0.5. It is therefore more likely that the two nodes will not be neighbors which reduces the uncertainty and hence leads to lower beacon update rate. Notice that this discussion depicts the case when $Z_{ij}(0) = 1$, which turns out to be the dominant case i.e. the one with the higher lower bound (refer to the explanation after (38)).

The trend shown in Figure 5 implies that when nodes are highly mobile they need to transmit beacons less frequently and the membership of nodes in a neighborhood may be more efficiently deciphered by the absence of beacons. Figure 6 shows that as the rate of packet arrival increases, the beacon overhead may become prohibitively high. Also, for a given packet arrival rate, it is observed that the rate for a deterministic packet arrival process is smaller than that for exponential and uniform arrivals. This is because the probability that a node leaves the neighborhood of a certain neighbor within a packet inter-arrival duration is the highest ($p(l)$ is close to 1) for deterministic arrival. For the uniform and exponential distributions the probability that packet inter-arrival time is less than the mean inter-arrival time is 0.5 and 0.63 respectively. Thus the probability that a node moves out of neighborhood during an inter-arrival duration is smaller than that for the deterministic arrival process.

VI. Capacity Deficit

A wireless ad hoc network is said to transport one bit-meter when a bit is transmitted over a distance of one meter [5]. The transport capacity of a network (in bit-meters per second) is defined as the supremum over the set of feasible rate vectors of the distance weighted sum of rates [28]. The transport capacity is expressed as $\lambda n \overline{L}$, where λ is the average arrival rate at the nodes, n is the number of nodes and $\overline{L}$ is the average distance traveled by the bits. It is shown in [5] that the transport capacity of an arbitrary wireless network is $\Theta\left(W\sqrt{nA}\right)$ where n, W and A are the number of nodes deployed, transmission rate of the nodes and area over which the network is deployed respectively. It is shown in [5] that for a particular interference model known as the *Protocol Model*, the upper bound on the transport capacity of an arbitrary wireless network is given by

$$\lambda n \overline{L} \leq \frac{\sqrt{8}}{\pi} \frac{1}{\Delta} W \sqrt{nA} \tag{70}$$

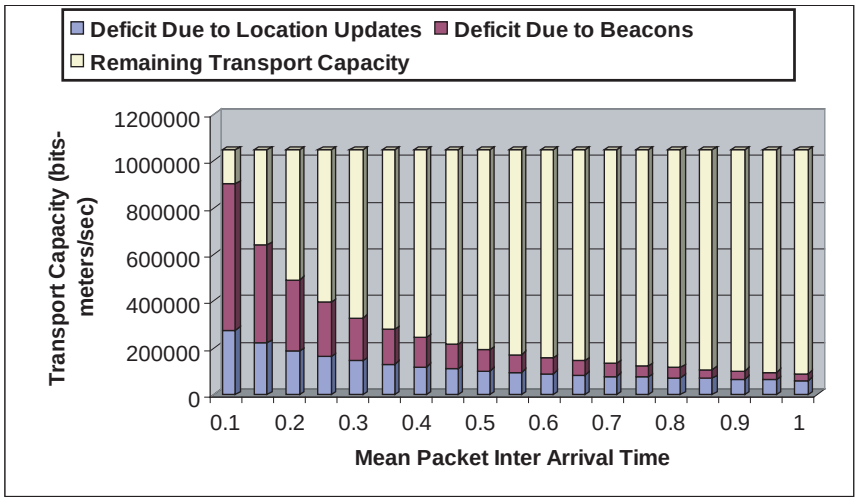


Fig. 9. Capacity deficit versus packet inter arrival time for $n = 1000$, $\sigma = 80$, $\epsilon = 1.0$, $\delta = 0.01$, $A = 100$.

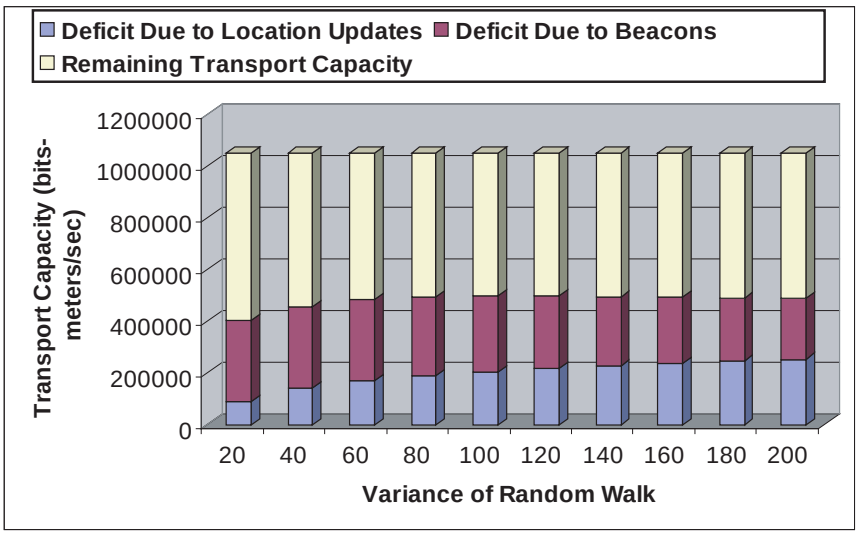


Fig. 10. Capacity deficit versus σ^2 for $n = 1000$, $\epsilon = 1.0$, $\delta = 0.01$, $A = 100$.

where $\Delta > 0$ is a parameter specified in the *Protocol Model* to prevent a neighboring node from transmitting on the same sub-channel at the same time.

Let η denote the expected distance between a node and its location server. Thus, on average, the location update information of a node travels at least η before reaching its location server. Thus the average overhead incurred by a node for updating its location information is at least $\eta U(\epsilon^2)$ and the overhead incurred by location update information on the network equals at least $n\eta U(\epsilon^2)$, where $U(\epsilon^2)$ is given by (25). A beacon transmitted by a node travels a distance equal to the communication radius. Thus the overhead incurred by the beacon packets on the network is at least $nrU^{(b)}(\delta)$, where $U^{(b)}(\delta)$ is given by (69). Thus the total transport *capacity deficit* due to the protocol overhead is at least $n\eta U(\epsilon^2) + nrU^{(b)}(\delta)$. This leads to the following theorem.

Theorem 5: For the Protocol Model, the upper bound on the *effective (or residual) transport capacity* available to an arbitrary network for transmitting data ($\lambda n\overline{L}$) is given by

$$\lambda n\overline{L} \leq \frac{\sqrt{8}}{\pi}\frac{1}{\Delta}W\sqrt{nA} - n\eta U(\epsilon^2) - nrU^{(b)}(\delta). \tag{71}$$

Theorem 5 has interesting implications. The raw transport capacity of a wireless network scales as $\sqrt{n}$ while the protocol

overhead scales as n. Therefore if the number of nodes deployed in a network increases beyond a certain threshold, denoted by $n^\star$, then no useful information may be transported in the network and the whole capacity is used up by the geographic protocol overheads. This leads to the following corollary.

Corollary 3: For position-based network protocols (such as geographic routing), the upper bound on the maximum number of nodes that may be deployed in a network while ensuring that it has non-zero residual transport capacity is given by

$$n^\star \leq \left(\frac{\frac{\sqrt{8}}{\pi}\frac{1}{\Delta}W\sqrt{A}}{\eta U(\epsilon^2) + rU^{(b)}(\delta)} \right)^2. \tag{72}$$

Proof: If the residual transport capacity is greater than zero then

$$\frac{\sqrt{8}}{\pi}\frac{1}{\Delta}W\sqrt{nA} - n\eta U(\epsilon^2) - nrU^{(b)}(\delta) \geq 0$$

which implies that

$$\sqrt{n}\left(\eta U(\epsilon^2) + rU^{(b)}(\delta)\right) - \frac{\sqrt{8}}{\pi}\frac{1}{\Delta}W\sqrt{A} \leq 0.$$

Rearranging the above equation yields (72). ■

Figures 8, 9, and 10 show the lower bound on the fraction of transport capacity that is used by the protocol overheads. Figures 8 and 9 illustrate that for large network size and packet arrival rate, complete transport capacity of the network may be occupied by the protocol overheads. Figure 10 shows that for a fixed network size and packet arrival rate, the lower bound in capacity deficit is not very sensitive to the variance of Brownian motion. This is because of the fact that although the minimum location update overhead increases with σ^2, the minimum beacon transmission rate decreases with increase σ^2.

VII. Discussions

In this section we briefly discuss some extensions of the rate-distortion formulation proposed in this paper. We first discuss the cases where our formulation and analysis yield trivial lower bound of zero. We then discuss how the formulation and analysis may be applied to other scenarios such as other location services, mobility models, etc. We then address some possible extensions to the model and analytical approach used in this paper.

A. Trivial Lower Bound Scenarios

The lower bound on the rate at which a node needs to update its location server, derived in Section IV, may be zero in certain cases. This lower bound is trivial and does not provide much useful information. In this subsection we provide a technique for obtaining more useful bounds for such cases. For clarity, we first consider the location update rate for one-dimensional networks and then explain how similar technique may be applied to two-dimensional networks. We also comment on cases where the results of Section V yield trivial lower bounds on minimum beacon overhead rate.

The lower bound on location overhead evaluated in IV-B is zero when $h(X_{i1}(T_1)) \leq \frac{1}{2}\log\left(2\pi e\epsilon^2\right)$. This occurs when the second moment of the change in position of a node during a packet inter-arrival period is less than the fidelity criterion ϵ^2. This may be the result of low mobility (small σ^2) or average packet inter-arrival time (small $E[S]$). Although average change in position of a node during a packet inter-arrival time may be small, the position change accumulates over time. Thus instead of updating the location server between every packet inter-arrival duration, the nodes may need to update the location information once in a few packet inter-arrival durations. Thus we modify the problem formulation for minimum location update overhead to take this into account. The distortion criterion considered in Section IV is that the expected squared error in location information of a node must be less than ϵ^2 at every time instant a packet destined to the node is generated in the network. We relax the distortion measure such that the squared error in location information is required to be less than ϵ^2 for every k^{th} generated for the node.

Let $f_S^k(t)$ denote the pdf of T_k, the time at which k^{th} packet destined to a node is generated in the network. T_k is simply the sum of k packet inter-arrival durations, each of which is independently and identically distributed according to $f_S(t)$. Thus

$f_S^k(t)$ is given by

$$f_S^k(t) = f_S^{k-1}(t) \star f_S^1(t) = \int_0^\infty f_S^{k-1}(t-\tau) f_S^1(\tau) d\tau, \tag{73}$$

$$f_S^1(t) = f_S(t) \tag{74}$$

where $\star$ is the convolution operator. The pdf of $X_{i1}(T_k)$, denoted by $f_{X_1}^k(x)$ is given by

$$f_{X_1}^k(x) = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2\tau}} e^{-\frac{x^2}{2\sigma^2\tau}} f_S^k(\tau) d\tau \tag{75}$$

and $H(X_{i1}(T_k))$ is given by

$$H\left(X_{i1}(T_k)\right) = -\int_{-\infty}^{\infty} f_{X_1}^k(x) \log\left(f_{X_1}^k(x)\right) dx. \tag{76}$$

Let $k^\star$ be defined as

$$k^\star \triangleq \min k | H\left(X_{i1}(T_k)\right) > \frac{1}{2}\log\left(2\pi e \epsilon^2\right). \tag{77}$$

In other words, $k^\star$ is the minimum number of packet inter-arrival durations after which the second moment of the change in position of a node is greater than ϵ^2. The value of $k^\star$ will depend on σ^2 and the distribution of packet inter-arrival duration. So if we consider the relaxed distortion measure where we are only concerned with the squared error in location information after every $k^\star$ packet generation instances, then following the same steps as in Subsection IV-B we get the following lower bounds

$$R\left(\epsilon^2\right) \geq \frac{1}{k^\star}\left(H\left(X_{i1}(T_{k^\star})\right) - \frac{1}{2}\log\left(2\pi e \epsilon^2\right)\right), \tag{78}$$

$$U\left(\epsilon^2\right) \geq \frac{1}{k^\star E[S]}\left(H\left(X_{i1}(T_{k^\star})\right) - \frac{1}{2}\log\left(2\pi e \epsilon^2\right)\right). \tag{79}$$

Note that the right hand side of the above constraints is always positive and hence are more meaning for the cases where results of Section IV yield trivial bounds. In fact the bounds evaluated in Section IV are special cases of (78) and (79) corresponding to $k^\star = 1$.

Similarly for two-dimensional network we can find more meaningful bounds by considering the change in location during k packet inter-arrival periods. An analogous $k^\star$ may be defined as

$$k^\star \triangleq \min_k k \in \left\{k | H\left(X_{i1}(T_k)\right) + H\left(X_{i2}(T_k)\right) > \log\left(\pi e \epsilon^2\right)\right\}. \tag{80}$$

For the relaxed distortion constraints we thus get the following bounds for the two-dimensional case

$$R\left(\epsilon^2\right) \geq \frac{1}{k^\star}\left(H\left(X_{i1}(T_{k^\star})\right) + H\left(X_{i2}(T_{k^\star})\right) - \log\left(\pi e \epsilon^2\right)\right), \tag{81}$$

$$U\left(\epsilon^2\right) \geq \frac{1}{k^\star E[S]}\left(H\left(X_{i1}(T_{k^\star})\right) + H\left(X_{i2}(T_{k^\star})\right) - \log\left(\pi e \epsilon^2\right)\right). \tag{82}$$

The lower bound on the minimum beacon overhead, evaluated in Section V, is zero when $\mathcal{H}(p(l^\star)) \leq \mathcal{H}\left(\frac{\delta}{2}, 1-\delta, \frac{\delta}{2}\right)$. This may happen if the $\sigma^2 E[\tau] >> 2r$. When $\sigma^2 E[\tau] >> 2r$, with probability greater than $1-\delta$ a node will leave the neighborhood of its neighbors within a packet inter-arrival period. According to our formulation the node does not need to send a beacon as the neighbors may simply assume that it has left their neighborhoods. As a result the bound may be loose in such circumstances because it is inevitable that the nodes will need to exchange beacons in order to maintain consistent neighborhood information. The idea of relaxing the distortion constraint by considering accuracy of neighborhood information over several packet arrival intervals, as done for the location update case, does not work in this case. This is because as we increase the number of packet arrival intervals the probability that two nodes remain neighbors over that time duration decreases. However if the fidelity criterion is very strict ($\delta \approx 0$) then the bounds will be meaningful over a wide range of σ and packet arrival rate distributions.

B. Application to Other Scenarios

In this section we comment on generality of the rate-distortion model and how it may be easily extended to other scenarios.

1) Per-Session Location Discovery: The network model used in this paper assumed that the location server of the destination is queried at the arrival of each new packet. In the context of geographic routing, another implementation is possible where a source node queries the location server only on the arrival of a new session. During the session, the source and destination may piggy-back their location information along with the data and ACK packets respectively. Of course such a scheme would work well only if the session involves flow of traffic in reverse direction (destination to source) and the time elapsed between arrival of packets in the session is not large.

The overhead incurred in updating the location server for such a protocol would be different from the results presented in Section IV. Also other than the overhead associated with updating the location server, another overhead associated with sending the location update information along with every ACK/data packet is also introduced.

The rate-distortion analysis of Section IV may be easily extended to analyze the overhead incurred by position-based network protocols with per-session location discovery. Let S_1 denote the time interval elapsed between arrival of two sessions destined for the same destination and let S_2 denote the inter-packet arrival time within a session. Let $f_{S_1}(t)$ and $f_{S_2}(t)$ denote the pdfs of S_1 and S_2 respectively. sing similar approach as Theorem 1, the overhead associated with updating the location server is greater than equal to

$$h(X_{i1}(T_1)) - \frac{1}{2}\log\left(2\pi e\epsilon^2\right) \tag{83}$$

where

$$f_{X_1}(x) = \int_{\tau=0}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2\tau}} e^{-\frac{x^2}{2\sigma^2\tau}} f_{S_1}(\tau) d\tau. \tag{84}$$

Similarly the overhead incurred in sending the location information along with ACK/data packets is greater than equal to

$$2\left(h(f'_{X_1}(x)) - \frac{1}{2}\log\left(2\pi e\epsilon^2\right)\right) \tag{85}$$

where

$$f'_{X_1}(x) = \int_{\tau=0}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2\tau}} e^{-\frac{x^2}{2\sigma^2\tau}} f_{S_2}(\tau) d\tau. \tag{86}$$

The factor of two appears in the above expression because both the source and destination need to send their location information piggybacked with each data and ACK packet. The overall overhead incurred in maintaining location information is given by

$$U(\epsilon^2) \geq \frac{1}{E[S_1]}\left(h(f_{X_1}(x)) - \frac{1}{2}\log\left(2\pi e\epsilon^2\right)\right) + \frac{2}{E[S_2]}\left(h(f'_{X_1}(x)) - \frac{1}{2}\log\left(2\pi e\epsilon^2\right)\right). \tag{87}$$

The total transport capacity deficit in bit-meters per second is greater than equal to

$$\frac{n\eta}{E[S_1]}\left(h(f_{X_1}(x)) - \frac{1}{2}\log\left(2\pi e\epsilon^2\right)\right) + \frac{2n\eta'}{E[S_2]}\left(h(f'_{X_1}(x)) - \frac{1}{2}\log\left(2\pi e\epsilon^2\right)\right) + nrU^{(b)}(\delta). \tag{88}$$

2) Other Location Services: The network model used in this paper accounts for a location service where the location of each destination is maintained by a single location server. Often location services use multiple location servers for maintaining location information of each destination. It should however be noted that the minimum rate at which a node must transmit its location information does not change and is given by (12). The multiple location servers affects the distance traveled by this location information generated by a node.

Consider a location service that assigns k location servers for each node. Each of the k location servers is updated by a node using k independent messages. Let η_i be the average distance between a node and its ith location server. The transport capacity deficit caused by a such a scheme would be greater than equal to

$$\left(\sum_{i=1}^{k}\eta_i\right) nU\left(\epsilon^2\right) + nrU^{(b)}(\delta). \tag{89}$$

In several location service schemes independent update messages are not issued. Instead the update messages may be routed along a multicast tree so that all the location servers are updated. In such a case, the distance traveled by the location update messages, η, would be equal to the length of the multicast tree. For example, consider XYLS (also known as column-row location service) [29]. In XYLS, each node maintains its location at every node that lies within the column containing the

node. This is done by sending an update message in the north-south direction. Every node that overhears the messages updates the location information. Thus in a square field of area A meter^2, a location update bit would travel $\sqrt{A}$ meters, i.e., $\eta = \sqrt{A}$. Thus the transport capacity deficit caused is greater than equal to

$$\sqrt{A}nU\left(\epsilon^2\right) + nrU^{(b)}(\delta). \tag{90}$$

3) Hierarchical Location Services and Distance Effects: In hierarchical location service schemes, there exists a hierarchy of location servers [30], [31]. The level of the server in the hierarchy depends on the distance from the node it serves. The location servers closest to a destination, and lowest in the hierarchy, have the most accurate information about the destination. As we move up the hierarchy, the location information becomes less accurate. That is farther a location server is from the destination it serves, less accurate would be the location information maintained by it. This scheme works well due to the *distance effect* - greater the distance from a node, slower it appears to move.

Let ϵ_i be the expected error in location information available with the location servers at level i. Let η_i be the average distance traveled by bits in order to update the servers at level i. Thus the transport capacity deficit for such a hierarchical location service scheme is greater than equal to

$$n\sum_{i=1}^{k} U\left(\epsilon_i^2\right)\eta_i + nrU^{(b)}(\delta). \tag{91}$$

We also note that, in this paper, we assumed the nodes move on a torus. In case that the area is not symmetric, distance effects will arise where some nodes are, on average, farther from the home region than in the case of a torus, all other parameters held constant. Thus, we expect the location update overheads in the non-torus cases would be higher than those derived in this paper, since updates have to travel longer distances.

4) Other Mobility Models: In this paper we use Brownian motion to model node mobility. Although Brownian motion may not be the most realistic mobility model, the existence of closed form expression for node location pdf makes analysis tractable. Also Brownian motion allows us to change the degree of uncertainty in node position by changing a single parameter σ.

Many other mobility models for ad hoc networks have been proposed [32], [33], some of which claim to be a better reflection of reality than the other. Our analysis may be easily extended to any other mobility model, as long as the expression for the steady state pdf of the node locations is known. The pdf of node location may be plugged into (14) in order to obtain $f_{X_1}(x)$. However it is a non-trivial exercise to evaluate the closed form expression of the node location for mobility models like random way-point model, Gauss-Markov model, etc., and, up to our knowledge, a steady state pdf expression is not known for the aforementioned mobility models.

5) Application to Specific Protocols Design: In this paper we have focused on providing a mathematical framework to derive bounds on protocol information and have not applied the results to the design of specific protocols, primarily because we want the analysis to be as general as possible. However, in [27] we have utilized the intuition provided in the paper to improve the design of a particular ad hoc networks protocol, specifically AODV, by manipulating the timing of location update instants (thus reshaping the packet arrival distribution) at the expense of increased packet delivery delay. The details of this protocol, called B-AODV (where the letter 'B' stands for 'Buffered'), as well as network simulations and performance measurements are included in [27].

C. Extensions to the Model and Analysis

In this subsection we discuss some of the possible extensions to the model and analysis that may be incorporated in the future.

1) Incorporating the Caching of Location Information: The network model does not take into account the caching of location information at the source and intermediate nodes. For example, the source node may cache the location information of the destination received from the location server and may add the cached information to the headers of future packets. The cached information may be used until it expires after some fixed time. Also intermediate nodes with fresher information regarding the location of the destination may update the packet header. Such a caching scheme has not been incorporated in our model.

Caching may save the overhead associated with periodically querying the location server, although the location update and beacon overheads may remain unchanged.

2) Exploring Closed Form and Tighter Bounds: Although the network model used in this paper is very simple, still it is not possible to obtain closed form analytical expressions for many quantities. For example, for the minimum beacon overhead case, all the results are expressed in terms of $\mathcal{H}\left(p\left(l^{\star}\right)\right)$. However numerical methods may be applied to evaluate the values of the expressions for given parameters. Finding tight closed form bounds for the expressions would also be the focus of future work.

VIII. Conclusion and Future Work

In this paper we studied the protocol overhead incurred by position-based network protocols so as to maintain a bounded distortion between the actual and perceived states. Specifically, the protocol overhead is categorized into *location update overhead* and *beacon overhead*. For both kind of overheads, the problem of finding minimum overhead incurred is formulated as rate-distortion problem. For location updates we evaluate a lower bound on the minimum rate at which a node must transmit its location information to the location server so that the expected error in location information is less than a given value. For the beacon updates, we evaluate a lower bound on the minimum rate at which a node must transmit beacon packets so that the probability that its neighbors maintain a correct neighborhood information is greater than a given value. We first evaluate the bounds for one-dimensional networks and then extend the results to two-dimensional networks. We also characterize the deficit in capacity caused by the protocol overheads.

Developing an information theoretic framework for evaluating the minimum protocol overhead incurred by several classes of network protocols, such as proactive, reactive, and hierarchical routing protocols, is an important application of the framework proposed here. Such a universal analytical framework for characterizing protocol overheads would be useful when applied to compare various paradigms and useful in determining which protocol is suitable for a given scenario. Development of efficient network protocols for mobile ad hoc networks, using the information theoretic results as a guideline, would also be a part of future research.

References

[1] R. Gallager, "Basic limit on protocol information in data communication networks," *IEEE Transactions on Information Theory*, vol. 22, no. 4, pp. 385–398, July 1976.

[2] B. Karp and H. T. Kung, "GPSR: Greedy perimeter stateless routing for wireless networks." in *MOBICOM '00*, 2000, pp. 243–254.

[3] P. Bose, P. Morin, I. Stojmenovic, and J. Urrutia, "Routing with guaranteed delivery in ad hoc wireless networks," *ACM Wireless Networks*, vol. 7, no. 6, pp. 609–616, 2001.

[4] R. C. Shah, A. Wolisz, and J. M. Rabaey, "On the performance of geographical routing in the presence of localization errors," in *Proc. of ICC'05*, 2005, pp. 2979–2985.

[5] P. Gupta and P. R. Kumar, "Capacity of wireless networks," *IEEE Trans. on Information Theory*, pp. 388–404, March 2000. [Online]. Available: citeseer.ist.psu.edu/gupta99capacity.html

[6] A. Ephremides and B. E. Hajek, "Information theory and communication networks: An unconsummated union." *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2416–2434, 1998.

[7] C. Rose and R. Yates, "Location uncertainity in mobile networks: A theoretical framework." *IEEE Comm. Magazine*, vol. 35, no. 2, pp. 94–101, 1997.

[8] A. Bhattacharya and S. K. Das, "LeZi-update: An information-theoretic approach to track mobile users in PCS networks," in *MobiCom '99*, 1999, pp. 1–12.

[9] B. An and S. Papavassiliou, "An entropy-based model for supporting and evaluating route stability in mobile ad hoc wireless networks." *IEEE Communications Letters*, vol. 6, no. 8, pp. 328–330, 2002.

[10] Q.-M. Tran, A. Dadej, and S. Parreau, "Characterizing mobility in ad hoc networks: A generalized approach," in *Proc. of ASWN'05*, 2005, pp. 65–70.

[11] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, and J. Jetcheva, "A performance comparison of multi-hop wireless ad hoc network routing protocols," in *MobiCom '98: Proceedings of the 4th annual ACM/IEEE international conference on Mobile computing and networking*. New York, NY, USA: ACM Press, 1998, pp. 85–97.

[12] S. R. Das, C. E. Perkins, and E. M. Belding-Royer, "Performance comparison of two on-demand routing protocols for ad hoc networks." in *Proceedings of IEEE INFOCOM '00*, 2000, pp. 3–12.

[13] P. Jacquet and L. Viennot, "Overhead in mobile adhoc network protocols," INRIA Rocquencourt, Paris, France, Technical Report, 2000. [Online]. Available: citeseer.ist.psu.edu/jacquet00overhead.html

[14] P. Johansson, T. Larsson, N. Hedman, B. Mielczarek, and M. Degermark, "Scenario-based performance analysis of routing protocols for mobile ad-hoc networks," in *MobiCom '99: Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking*. New York, NY, USA: ACM Press, 1999, pp. 195–206.

[15] V. Park and M. Corson, "A performance comparison of the temporally-ordered routing algorithm and ideal link-state routing," in *Proceedings of IEEE International Symposium on Systems and Communications (ISSC'98)*. "IEEE Computer Society Press", "June" "1998". [Online]. Available: "citeseer.ist.psu.edu/park98performance.html"

[16] C. A. Santivanez, A. B. McDonald, I. Stavrakakis, and R. Ramanathan, "On the scalability of ad hoc routing protocols," in *Proceedings of IEEE INFOCOM '02*, 2002, pp. 1688–1697.

[17] L. Viennot and P. Jacquet, "Analyzing control traffic overhead in mobile ad-hoc network protocols," INRIA Rocquencourt, Paris, France, Tech. Rep., 2002. [Online]. Available: citeseer.ist.psu.edu/viennot02analyzing.html

[18] H. Wu and A. Abouzeid, "Cluster-based routing overhead in networks with unreliable nodes," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC 2004)*, 2004, pp. 2557–2562.

[19] N. Zhou and A. Abouzeid, "Routing in ad-hoc networks: A theoretical framework with practical implications," in *Proceedings of INFOCOM'05*, 2005, pp. 1240–1251.

[20] N. Zhou, H. Wu, and A. Abouzeid, "The impact of traffic patterns on the overhead of reactive routing protocols," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 3, pp. 547–560, March 2005.

[21] N. Zhou, H. Wu, and A. A. Abouzeid, "Reactive routing overhead in networks with unreliable nodes," in *MobiCom '03: Proceedings of the 9th annual international conference on Mobile computing and networking*. New York, NY, USA: ACM Press, 2003, pp. 147–160.

[22] T. Berger, "Information rates of wiener processes," *IEEE Transactions on Information Theory*, vol. 16, no. 2, pp. 134–139, March 1970.

[23] M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, pp. 477–486, 2002.

[24] G. Sharma, R. Mazumdar, and N. Shroff, "Delay and capacity trade-offs in mobile ad hoc networks: A global perspective," in *Proc. of IEEE INFOCOM'06*, 2006.

[25] P. Santi, "The critical transmitting range for connectivity in mobile ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 4, no. 3, pp. 310–317, 2005.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & sons, 1991.

[27] N. Bisnik, "Stochastic and information theoretic models for design and performance evaluation of mobile ad hoc and sensor networks," Ph.D. Dissertation, Department of ECSE, RPI, Troy, NY. Available at http://www.ecse.rpi.edu/homepages/abouzeid/bisnik_thesis.pdf.

[28] F. Xue, L.-L. Xie, and P. R. Kumar, "The transport capacity of wireless networks over fading channels." *IEEE Trans. on Information Theory*, vol. 51, no. 3, pp. 834–847, 2005.

[29] I. Stojmenovic and P. Pena, "A scalable quorum based location update scheme for routing in ad hoc wireless networks," Technical Report, SITE, University of Ottawa, TR-99-09, 1999. [Online]. Available: citeseer.ist.psu.edu/stojmenovic99scalable.html

[30] J. Li, J. Jannotti, D. S. J. D. Couto, D. R. Karger, and R. Morris, "A scalable location service for geographic ad hoc routing," in *MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking*. New York, NY, USA: ACM Press, 2000, pp. 120–130.

[31] Y. Xue, B. Li, and K. Nahrstedt, "A scalable location management scheme in mobile ad-hoc networks," in *LCN '01: Proceedings of the 26th Annual IEEE Conference on Local Computer Networks*. Washington, DC, USA: IEEE Computer Society, 2001, p. 102.

[32] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Communications & Mobile Computing (WCMC): Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications*, vol. 2, no. 5, pp. 483–502, 2002. [Online]. Available: http://citeseer.ist.psu.edu/camp02survey.html

[33] A. Jardosh, E. M. Belding-Royer, K. C. Almeroth, and S. Suri, "Towards realistic mobility models for mobile ad hoc networks," in *MobiCom '03: Proceedings of the 9th annual international conference on Mobile computing and networking*. New York, NY, USA: ACM Press, 2003, pp. 217–229.

Nabhendra Bisnik received the B.Tech. in electrical engineering from Indian Institute of Technology, Kanpur in 2003, and the Ph.D. degree in electrical engineering from Rensselaer Polytechnic Institute, Troy, NY in 2007. He is currently member of technical staff, Intel, Portland, Oregon. His research interests include modeling and analysis of wireless ad hoc networks, phenomena coverage problems in wireless sensor networks and resource discovery in dynamic networks.

Alhussein A. Abouzeid received the B.S. degree with honors from Cairo University, Cairo, Egypt, in 1993, and the M.S. and Ph.D. degrees in electrical engineering from the University of Washington, Seattle, in 1999 and 2001, respectively. From 1993 to 1994, he was with the Information Technology Institute, Information and Decision Support Center, The Cabinet of Egypt, where he received a degree in information technology. From 1994 to 1997, he was a Project Manager at Alcatel Telecom. He held visiting appointments with the aerospace division of AlliedSignal (currently Honeywell), Redmond, WA, and Hughes Research Laboratories, Malibu, CA, in 1999 and 2000, respectively. In 2001 he joined Rensselaer Polytechnic Institute (RPI), Troy, NY, where he is currently an associate professor in the Electrical, Computer and Systems Engineering Department. From 2008 to 2010 he was on leave from RPI where he served as a Program Director in the Computer and Network Systems Division, the Computer and Information Science and Engineering Directorate, the National Science Foundation, Arlington, VA. Dr. Abouzeid is a member of ACM and IEEE, and has served on several technical program and executive committees of various conferences. He is also a member of the Editorial Board of IEEE Transactions on Wireless Communications, and Computer Networks (Elsevier). He is a recipient of the Faculty Early Career Development Award (CAREER) from the U.S. National Science Foundation in 2006. His research focuses on protocols for dynamic networks.