Expected Routing Overhead for Location Service in MANETs Under Flat Geographic Routing

Richard J. La and Eunyoung Seo

***Abstract*—We study routing overhead due to location information collection and retrieval in mobile ad-hoc networks employing geographic routing with no hierarchy. We first provide a new framework for quantifying overhead due to control messages generated to exchange location information. Second, we compute the minimum number of bits required on average to describe the locations of a node, borrowing tools from information theory. This result is then used to demonstrate that the expected overhead is $\Omega(n^{1.5} \log(n))$, where n is the number of nodes, under both proactive and reactive geographic routing, with the assumptions that (i) nodes' mobility is independent and (ii) nodes adjust their transmission range to maintain network connectivity. Finally, we prove that the minimum expected overhead under the same assumptions is $\Theta(n \log(n))$.**

***Index Terms*—Data communications, mobile communication systems, network management, routing protocols.**

I. INTRODUCTION

A mobile ad-hoc network (MANET) is a collection of mobile nodes that construct and maintain a network without a centralized authority. Unlike in a more traditional wired network (e.g., the Internet), there are no dedicated routers or switches responsible for forwarding packets; instead, every node participates in relaying packets. In addition, since nodes are assumed mobile, one-hop connectivity between nodes and the network topology can change over time. Consequently, underlying routing protocols are asked to cope with potentially frequent changes in topology.

Recently there has been much research on understanding the network transport throughput, or simply transport throughput, of multi-hop wireless networks: In their seminal paper [12] Gupta and Kumar investigated the transport throughput of static multi-hop wireless networks and showed that the transport throughput increases, at best, as $\sqrt{n}$ with an increasing number of nodes n, i.e., $O(\sqrt{n})$.[1] This finding implies that per-node throughput decreases to zero as $n \to \infty$. Grossglauer and Tse [10] exploited the mobility of nodes and demonstrated that, if unbounded delays can be tolerated, under some technical conditions per-node throughput of $\Theta(1)$ can be achieved. To bridge the gap in the transport throughput between static networks and mobile networks, Sharma et al. [30] examined the trade-off between the transport throughput and delays that must be tolerated in order to achieve certain level of transport throughput. Other related work can be found in [6], [7], [19], [21].

This work was supported by the National Science Foundation under Grant ANI 02-05330, the U.S. Army Research Laboratory, and the Laboratory for Telecommunications Sciences (LTS).

Authors are with the Department of Electrical and Computer Engineering and the Institute for Systems Research at the University of Maryland, College Park. E-mail: hyongla@umd.edu, eyseo@umd.edu

[1]The notation we use throughout the paper is explained in subsection III-C.

A. Background

In most of these studies, however, authors do not explicitly address the issue of routing overhead. To be more precise, they do not explain how necessary routing information is obtained and how much network resource (e.g., transport throughput) is required to obtain needed routing information in order to achieve claimed transport throughput. Therefore, in order to better understand the scalability of MANETs with an increasing number of nodes and to find out how to dimension them properly (e.g., bandwidth), one should examine how routing overhead scales in MANETs, in particular, in comparison to network transport throughput. A good understanding of routing overhead may also allow us to correctly identify critical bottlenecks and to deal with them more effectively.

To the best of our knowledge, the first serious attempt at an analytical study of protocol overhead was carried out by Gallager in [8]. There are also several recent analytical studies on routing overhead in MANETs, some of which we summarize here: Zhou and Abouzeid [35], [36] applied the tools from information theory to examine the overhead due to the changes in network topology under two-tier hierarchical routing. Their key idea is to model the time-varying network topology as a stochastic process and to evaluate the overhead required to describe the local network topology in subregions to cluster heads and to distribute the global ownership information to all cluster heads. Then, they studied the scaling laws of the memory requirement and routing overhead under three different physical scalings of the network.

In another study [3] Bisnik and Abouzeid formulated the problem of characterizing the minimum routing overhead as a rate-distortion problem. They considered geographic routing with location servers that have *known* locations and store location information of other mobile nodes, and investigated the information rate required to satisfy a prescribed squared-error distortion constraint. Viennot et al. [34] examined control overhead under both proactive and reactive routing, and suggested that control overhead is proportional to the square of the number of nodes in the network.

B. Motivation

In this paper we take another step towards understanding routing overhead in MANETs: We assume that nodes employ flat geographic or position-based routing *without* designated location servers that maintain the location information of mobile nodes. In other words, unlike the settings studied in [35], [36], there is no (two-tier) hierarchy in routing or specialized location servers whose locations are *known* to mobile nodes. Location servers with fixed locations may be vulnerable to attacks. However, we will show that we can use existing nodes in the network to form *virtual* location servers to provide similar location service to nodes.

The goal of our study is twofold: First, we aim to provide a new framework for studying routing overhead, especially for geographic routing, which can capture the differences that arise from the specific schemes employed to disseminate and acquire location information. To this end we develop a new framework, borrowing tools from information theory to compute the minimum average number of bits required to describe *approximated* locations of mobile nodes. Secondly, based on the proposed framework, we explore how routing overhead scales with the network size under different routing schemes. In particular, we focus on the routing overhead *only* due to *dissemination* and *acquisition* of location information, i.e,. location service.

Although information theoretic approach has been used to study routing overhead in the past (e.g., [3], [35], [36]), our approach is quite different from those employed by existing studies. First, we focus on the scenario of practical interest where the network is connected with a high probability. To be more precise, we assume that the transmission range of the nodes is selected so that the network is connected with probability approaching one as the number of nodes grows. This issue of network connectivity has been studied extensively by various researchers (e.g., [11], [15], [24]), and we summarize the results relevant to our study in subsection II-B. With this assumption in place, we compute the minimum average number of bits needed to describe the approximated locations of the nodes to support geographic routing.

Second, rather than simply computing the "information rate" required to describe the changes in network topology or node locations subject to a constraint on distortion in location information as done in [3], [35], we identify suitable *quantization levels* for approximating the node locations for the purpose of geographic routing. This is motivated by the following viewpoint: When a network is entrusted to provide timely exchange of information between nodes, inaccurate location information of nodes should not prevent the network from carrying out its task, *regardless of the locations of the nodes*. Therefore, the location information used for routing ought to be accurate enough to allow successful delivery of packets to their respective destinations.

Another important difference from previous studies, in particular the work by Bisnik and Abouzeid [3], is that, as proposed by Gupta and Kumar [12], in order to be more consistent with the measure of transport throughput and required resource consumption, we adopt the unit of bits$\times$meters per unit time for measuring overhead and explicitly take into account the distances traveled by control messages when computing the overhead. As we will demonstrate, this is *necessary* for capturing a disparity in resource expenditure by control messages under different geographic routing schemes.

C. Summary of main results

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

i. We show that, under the assumption that nodes employ the critical transmission range (CTR) for network connectivity, the *minimum* expected number of bits required on average to describe the approximated locations of a node for successful routing of packets based on provided location information is asymptotically $\log(n)$ as the number of nodes n grows [Section IV].[2]

ii. Making use of the first finding, we examine the expected routing overhead due to location service under both proactive and reactive *geographic* routing[3] where the mobility of the nodes is independent, and demonstrate that the expected overhead under these schemes is $\Omega(n^{1.5} \cdot \log(n))$ [Section V].

iii. Finally, we prove that, under the same mutually independent mobility of the nodes, the *minimum* expected overhead required for location service is $\Theta(n \cdot \log(n))$ [Section VI].

Let us provide some intuition behind our findings. There are three main sources of overhead under consideration: First, a linear term n in the scaling law of expected routing overhead in both $\Omega(\underline{\underline{n}} \cdot \sqrt{n} \cdot \log(n))$ and $\Theta(\underline{\underline{n}} \cdot \log(n))$ comes from the assumption that there are n nodes (moving according to n mutually independent mobility processes).

Secondly, $\log(n)$ term reflects the average number of bits carried by control messages containing both the identity (ID) and the location information of a node; if source-destination pairs are selected randomly, at least $\log(n)$ bits (and at most $\log(n)+1$ bits) are needed to identify a node. In addition, our first finding states that the expected number of bits required on average to describe the approximated locations of a node for geographic routing is asymptotically $\log(n)$. Therefore, this tells us that the expected number of bits carried by a location information message is on average $\Theta(\log(n))$, giving us $\Theta(n \cdot \log(n))$ for the minimum expected routing overhead.

Thirdly, the additional $\sqrt{n}$ term in $\Omega(n \cdot \underline{\underline{\sqrt{n}}} \cdot \log(n))$ is caused by costly flooding of control messages under both proactive and reactive geographic routing schemes. It is important to note that capturing this key source of the discrepancy in the scaling laws of routing overhead, hence resource requirements, demands accounting for the distances traveled by control messages. In other words, simply computing the information rates needed to describe the changes in topology or node locations would *not* disclose the disparity caused by the details of employed routing schemes.

While we assumed that the transmission ranges of the nodes are set to the CTRs to facilitate our analysis, we show that this assumption is not necessary; our main findings still hold under a mild technical condition even when the transmission ranges of the nodes are of different *orders* (Sections V and VI). Furthermore, we demonstrate that they are still true under a quasi unit disk or cost based network connectivity model [20], [29] (subsection VII-D).

Although our findings are based on simple models for mathematical tractability, they already shed initial light on major sources of overhead due to location service. They also tell us how various assumptions, including those on nodes' mobility and the selection of source-destination pairs, affect

[2]Throughout this paper $\log(\cdot) = \log_2(\cdot)$, i.e., logarithm to base 2.

[3]A geographic routing scheme is said to be *proactive* if every node tries to maintain consistent, up-to-date *location information* of all other known nodes. Likewise, a geographic routing scheme is called *reactive* if the location information is provided only upon request.

the overhead. Thus, our study hints at how the overhead may change as some of these assumptions are relaxed or adjusted to model more realistic scenarios, inviting further studies.

D. Organization

The rest of the paper is organized as follows: Section II describes the problem we are interested in studying and provides a short summary of the results on network connectivity. Section III explains the mobility models, assumptions we introduce on mobility and the parametric scenario used to study the scaling law of expected routing overhead due to location service under different routing schemes. The minimum expected number of bits required on average to describe the approximated locations of a node is derived in Section IV, followed by a discussion on how expected routing overhead scales under proactive and reactive geographic routing schemes in Section V. We study the minimum expected routing overhead and describe a scheme that achieves the same scaling order as the minimum expected routing overhead in Section VI. A discussion on our findings is provided in Section VII.

II. Setup

Throughout the paper we use a discrete-time model and assume that time is divided into contiguous timeslots $t \in \mathbb{Z}_+ := \{0, 1, 2, \ldots\}$, where the duration of a timeslot is taken to be a unit time. Although the mobility of a node is continuous in real life, we approximate it using a discrete-time stochastic process and assume that the location of a node is fixed during a timeslot. This may be a reasonable assumption when a node is (quasi-)stationary much of the time and spends a relatively small fraction of time in transition between locations or if the duration of timeslot is small enough so that, with high probability, the location of a node does not change significantly over the duration of a single timeslot.[4] A similar assumption is often introduced in the literature (e.g., [10], [35], [36]).

In a multi-hop wireless network, one-hop connectivity between nodes is likely to be maintained through exchange of control messages (e.g., HELLO messages) at the data link layer. For our analysis we model the one-hop network connectivity using a random geometric graph (RGG) [23]: Each node i is aware of and can communicate with all other nodes within its *communication* or *transmission* range γ (according to the Euclidean distance), which we call *immediate neighbors*, or simply *neighbors*, of node i. We say that there is a bi-directional link, or simply a link, between two neighbors.

The RGG model has been used extensively in the literature as an approximate model to one-hop connectivity of wireless networks (e.g., [11], [14], [15], [16], [28]). The transmission range γ in the RGG model is assumed to be determined by the transmit power employed by the nodes, channel propagation and the signal-to-noise ratio corresponding to a bit error rate constraint [11]. Under a channel loss model often used in the literature, the received power P_{rcv} is related to the transmit power P_{tx} and the distance d by

$$P_{rcv} = P_{tx} \cdot G_{tx} \cdot G_{rcv} \cdot L \cdot d^{-\alpha}, \quad (1)$$

where G_{tx} and G_{rcv} are the transmitter and receiver antenna gain, respectively, L accounts for system loss and other factors that may depend on the wavelength, and α is the path loss exponent [27]. If one requires that the received power $P_{rcv} \geq P_{\min}$ for some threshold $P_{\min}$, we must have

$$d \leq \left(\frac{P_{tx} \cdot G_{tx} \cdot G_{rcv} \cdot L}{P_{\min}} \right)^{1/\alpha} \quad (2)$$

and $P_{tx} \propto d^{\alpha}$. While our analysis is carried out under the RGG model, we will discuss how our results can be extended to different network connectivity models such as quasi unit disk model [20] and cost based model [29] in Section VII.

Throughout the paper we assume that every node knows its immediate neighbors. In addition, when a packet reaches an immediate neighbor of its destination, the neighbor can deliver it to the destination in one-hop without any other information.

A. Geographic routing and overhead for location service

We assume that nodes are equipped with Global Positioning System (GPS) devices and know their positions, which are assumed accurate throughout. Each node is aware of exact locations of its immediate neighbors.[5] This can be done either by exchanging the GPS location information between one-hop neighbors (for example, by piggybacking it in HELLO messages) or by observing the received signal strength and angle in which signals arrive.

Nodes employ *geographic* (or *position-based*) routing; they route packets using location information of the destinations [32], [33]. It has been suggested [17], [22] that geographic routing leads to better performance in large multi-hop wireless networks than other routing schemes that do not exploit location information (e.g., destination-sequenced distance vector (DSDV) routing [25] or dynamic source routing (DSR) [18]). A main reason for the performance gain is that, while routing schemes such as DSDV require *global* topological information that can change frequently, geographic routing allows nodes to make *local* decisions based on the locations of their immediate neighbors and the destination, without having to learn end-to-end route information.

Obviously, for proper operation of geographic routing, the location information of the destination contained in packets must be accurate enough so that nodes can route them to their destinations using the destination ID and location information. However, more accurate location information requires more bits, hence, larger overhead. We are interested in the case where the provided location information of destinations is accurate enough so that multi-hop packet routing can be performed using the location information *without* having to flood the neighborhoods of destinations with packets, while minimizing the number of bits required to describe location information.

Our study aims at (i) developing a new framework for quantifying routing overhead in MANETs employing geographic routing and (ii) examining how the routing overhead

[4]However, with small probability, the location of a node may change significantly from one timeslot to next.

[5]In practice, for proper operation of geographic routing the location information of neighbors needs to be accurate *relative to* the transmission range of the nodes. However, for simplicity of exposition we assume that nodes know the exact locations of their neighbors.

(measured in the unit of bits×meters per unit time proposed in [12]) required to *disseminate* and *acquire* location information of the nodes, scales with the number of nodes. We do not, however, concern ourselves with the delays experienced by messages. More precisely, we assume: (i) nodes can deliver their location information at timeslot $t \in \{1, 2, \ldots\} =: \mathbb{N}$, to any other nodes within the same timeslot (assuming network connectivity discussed in the following subsection); and (ii) assuming that nodes know where to access it, they can retrieve the location information of other nodes during the same timeslot. This implicitly assumes that the network has sufficient bandwidth to handle all overhead, including routing overhead, and to transport data in a timely manner. In practice, however, the delays incurred during dissemination and/or acquisition of location information can be non-negligible and cause inconsistency or staleness of location information.

Exchange of control messages to discover neighbors and to maintain links with them introduces additional overhead at the *data link* layer. However, we do not consider this overhead at the data link layer, including the overhead due to exchange of location information with immediate neighbors, because it does not depend on the adopted routing scheme. We refer interested readers to a study by Bisnik and Abouzeid in [3].

B. Network connectivity and critical transmission range

A primary function of a communication network is to enable exchange of information between nodes. When information is time-sensitive or cannot tolerate large delays, timely delivery of information demands that the underlying network be connected. In other words, there must exist an end-to-end path from a source to a destination (with a high probability) when such a path is desired. This is the scenario of interest we consider in this paper.

Recently there has been much work on connectivity of a multi-hop wireless network (e.g., [11], [14], [15], [24], [28]). We refer interested readers to a monograph by Penrose [23]. In particular, Penrose [24] (and later by Santi [28]) proved the following result we will borrow: Suppose that $n, n \geq 1$, nodes are placed independently of each other, according to a common spatial density function f with connected and compact support $\mathbb{D}$ and smooth boundary $\partial\mathbb{D}$. Let γ be a common transmission range of the nodes. The network is said to be connected if, for *every* pair of nodes (i, j), we can find a sequence of links providing an end-to-end route between the two nodes.

Theorem 1 ([24], [28]): Define $f_\star := \inf_{\mathbf{x}\in\mathbb{D}} f(\mathbf{x})$ and assume $f_\star > 0$. The *minimum* common transmission range required for connectivity, denoted by $\gamma^\star(n)$, satisfies

$$\lim_{n\to\infty} \frac{n\ \pi\ \gamma^\star(n)^2}{\log(n)} = \frac{1}{f_\star} \qquad \text{with probability 1.} \tag{3}$$

A similar result in the case of a uniform spatial distribution of nodes is obtained by Gupta and Kumar [11].

III. Mobility model and parametric scenario

This section first describes the node mobility processes we consider, and then explains the parametric scenario we adopt to study how the *expected* routing overhead for location service increases with the network size. We define all the random variables (rvs) and stochastic processes of interest on some common probability space $(\Omega, \mathcal{F}, \mathbf{P})$.

A. Mobility model

Nodes move on a domain $[0, \overline{D}]^2 =: \mathbb{D}$.[6] As mentioned earlier, we approximate the mobility of the nodes using discrete-time processes; the *mobility process* or *trajectory* of a node i is given by a discrete-time stochastic process $\mathbb{L}_i = \{L_i(t); t \in \mathbb{Z}_+\}$, where $L_i(t) = (L_{i,x}(t), L_{i,y}(t)) \in \mathbb{D}$ specifies the *location* or *position* of the node at time t, using the Cartesian coordinate system. We assume that, at each timeslot $t \in \mathbb{N}$, the transition from $L_i(t-1)$ to $L_i(t)$ takes place at the *beginning* of the timeslot.

The steady-state spatial distribution of the nodes is assumed to yield a *continuous* density function $f : \mathbb{D} \to \mathbb{R}_+ := [0, \infty)$. For each $t \in \mathbb{Z}_+$, f^t denotes the joint density function of $(L_i(0), \ldots, L_i(t))$. We assume that there exist constants ξ_1 and ξ_2, $0 < \xi_1 \leq \xi_2 < \infty$, such that, for all $t \in \mathbb{Z}_+$ and for all $\underline{\ell}_t \in \mathbb{D}^{t+1}$,

$$0 < \xi_1^{t+1} \leq f^t(\underline{\ell}_t) \leq \xi_2^{t+1} < \infty\ , \tag{4}$$

i.e., for every finite t, the joint density function f^t is non-vanishing and is also upper bounded by ξ_2^{t+1} over $\mathbb{D}^{t+1}$. This implies that node's locations do not concentrate in some parts of the domain $\mathbb{D}$ over time. For example, a two-dimensional Brownian motion with reflection, starting with an appropriate initial condition and sampled periodically, satisfies this assumption. Removal of the assumption in (4) has a rather serious consequence on network connectivity (see [13] for an example). Its impact on expected routing overhead is discussed in more detail in Section VII.

B. Parametric scenario

In order to study how the expected overhead scales with the number of nodes in the network, we consider the following parametric scenario with increasing n: For each fixed $n \in \mathbb{N}$, there are n nodes moving on the domain $\mathbb{D}$, and we denote the set of nodes by $N^{(n)} = \{1, 2, \ldots, n\}$.[7] We assume homogeneous mobility of the nodes. The mobility process of node $i \in N^{(n)}$, given by $\mathbb{L}_i^{(n)} := \{L_i^{(n)}(t); t \in \mathbb{Z}_+\}$, is assumed *stationary* and *ergodic*. Moreover, the mobility processes $\mathbb{L}_i^{(n)}, n \in N^{(n)}$, are *mutually independent*.

1. Connection requests: For each $i \in N^{(n)}$ and $t \in \mathbb{N}$, let $A_i^{(n)}(t)$ denote the number of requests arriving at the *other* nodes for a connection *to* node i at timeslot t. Without loss of generality, we assume $\{A_i^{(n)}(t); t \in \mathbb{N}\} =: \mathbb{A}_i^{(n)}$, are independent and identically distributed (i.i.d.) Bernoulli rvs with parameter $p^{(n)} > 0$. This implies that at most one other node will generate a connection request to node i, which is called the *source* of the connection, in each timeslot. We assume that the source is equally likely to be any of the remaining $n-1$ nodes, independently of the past and the sources of other connection requests.

[6] We assume a square region for convenience. However, similar results hold with any arbitrary compact, convex domain.

[7] This is often called a *dense* network in the literature.

Each connection request arriving at its source needs the location information of its destination for geographic routing. We assume that connection requests arrive at their sources at the beginning of each timeslot $t \in \mathbb{N}$ *after* nodes move to their new locations $L_i^{(n)}(t), i \in N^{(n)}$. The connection request arrival processes $\mathbb{A}_i^{(n)}, i \in N^{(n)}$, are mutually independent and also independent of mobility processes $\mathbb{L}_i^{(n)}, i \in N^{(n)}$.

Since we are interested in studying how the routing overhead grows with the number of nodes, we assume that the average number of connection requests to each node per timeslot is fixed, i.e., $\mathbf{E}\left[A_i^{(n)}(t)\right] = p^{(n)} = p > 0$ for all $i \in N^{(n)}$ and all $n \in \mathbb{N}$. Because the source of a connection request to a node is equally likely to be any of the remaining $n-1$ nodes, it is clear that, for each fixed $n \in \mathbb{N}$, the number of connection requests that arrive at a node (as the source) in a timeslot is a binomial($n-1$, $\frac{p}{n-1}$) rv.

2. Transmission range: We are interested in the case where the nodes adjust their common transmission range to maintain network connectivity as discussed in subsection II-B. Therefore, the transmission range of the nodes should be at least the CTR $\gamma^\star(n) = c^\star\sqrt{\log(n)/n}$ with $c^\star = 1/\sqrt{\pi\, f_\star}$ [28]. In their seminal paper on transport throughput [12], Gupta and Kumar showed that, in order to minimize interference to other simultaneous transmissions and to maximize transport throughput in a multi-hop wireless network, nodes should employ the *smallest* transmission range while maintaining network connectivity (i.e., the CTR $\gamma^\star(n)$).

In the subsequent sections we follow this finding by Gupta and Kumar [12] and assume that nodes employ a common transmission range of $\gamma^\star(n)$ to maximize transport throughput and keep the network connected with a high probability.[8]

Assumption 1: For each fixed $n \in \mathbb{N}$, the transmission range of the nodes is given by $\gamma^\star(n)$.

We will discuss how different choices of transmission ranges affect our findings in Sections IV through VI.

C. Notation

In this subsection we describe the notation we will use throughout the paper.

N1. A function $a(n)$ is $O(b(n))$ if there exist $0 < c_1 < \infty$ and $n_1^\star < \infty$ such that, for all $n \geq n_1^\star$, we have $a(n) \leq c_1 \cdot b(n)$.
N2. A function $a(n)$ is $\Omega(b(n))$ if there exist $c_2 > 0$ and $n_2^\star < \infty$ such that, for all $n \geq n_2^\star$, we have $c_2 \cdot b(n) \leq a(n)$.
N3. A function $a(n)$ is $\omega(b(n))$ if for every $c > 0$, there exists $n^\star(c)$ such that, for all $n \geq n^\star(c)$, $c \cdot b(n) < a(n)$.
N4. A function $a(n)$ is $\Theta(b(n))$ if there exist $0 < c_3 < c_4 < \infty$ and $n_3^\star < \infty$ such that for all $n \geq n_3^\star$, we have $c_3 \cdot b(n) \leq a(n) \leq c_4 \cdot b(n)$. Note that $a(n) = \Theta(b(n))$ if and only if $a(n) = O(b(n))$ and $a(n) = \Omega(b(n))$.
N5. A function $a(n) \sim b(n)$ if $\lim_{n\to\infty}(a(n)/b(n)) = 1$.

[8]To ensure network connectivity with high probability for finite n, the transmission range should be set to $\beta^\star \cdot \gamma^\star(n)$, where $\beta^\star > 1$. However, for notational simplicity we omit $\beta^\star$ in the analysis. The omission of this constant $\beta^\star$ does not change our results.

IV. Description of Node Locations

First, note that the location $L_i^{(n)}(t) \in \mathbb{D}$ of node i at time t is a two-dimensional *continuous* random vector for all $t \in \mathbb{Z}_+$. Therefore, they *cannot* be described exactly with a *finite* number of bits in general. Moreover, for the purpose of routing packets using location information, exact locations are not necessary and sufficiently accurate *approximations* of locations suffice. Hence, we are interested in finding out how accurate the location information contained in packets must be so as to allow successful routing of packets based on the provided location information.

The number of bits needed for approximated location information carried by packets for geographic routing is governed by the aforementioned required accuracy and the way location information is encoded. The first determines the *quantization level* to be selected for approximation. Bisnik and Abouzeid [3] utilized the rate distortion theory to compute the necessary information rate subject to a squared-error distortion constraint. This approach, however, may require that different quantization levels be used in different regions, depending on the spatial distribution, and allows for the possibility that the location information of nodes in an area of low spatial density is not accurate enough for successful delivery of packets.

We argue that a communication network should be able to deliver packets irrespective of nodes' locations. This is especially true when the spatial distribution of the nodes is not correlated with their communication needs. In this case, non-uniform approximation of location information demanded by rate distortion theory, which does not consider the communication needs, may compromise the communication with nodes in low spatial density areas and, hence, may be unsuitable.

In this section we investigate the minimum expected number of bits required per timeslot to specify approximated locations of a node to enable geographic routing. For the reason explained above, we assume that the selected quantization level for approximating node locations does *not* depend on their locations. Furthermore, as stated in subsection III-B, we focus our study only on the case of practical interest where the network is connected with probability approaching one, by setting the common transmission range of the nodes to the CTR $\gamma^\star(n) = c^\star\sqrt{\log(n)/n}$.

Before stating our result, let us first briefly describe the class of packet routing schemes we consider. Packets carry both the destination ID and *approximated* location information. The encoding and decoding rules for approximated locations are assumed common knowledge.
1. The source of a packet encodes the location of its destination, which is approximated with a selected quantization level, using the common encoding rule and places the encoded location information in the packet.
2. A relay node that receives a packet first checks if the destination is an immediate neighbor. If so, it delivers the packet to the destination. If not, it decodes the approximated location of the destination using the common decoding rule. It then selects an immediate neighbor that is closest to the decoded approximated location as the next hop. Recall that the nodes are assumed to know the precise locations of their immediate neighbors.

It has been observed that as the network becomes dense, a greedy approach that either minimizes the distance or maximizes the forward progress to the destination works well [22]. However, when a greedy approach fails, other schemes, such as Greedy-Face-Greedy (GFG) routing scheme [4], can be used to guarantee the delivery.

The following lemma states that the minimum expected number of bits needed on average to describe the approximated locations of a node for geographic routing approaches $\log(n)$ asymptotically as $n \to \infty$. This finding will be used to study how the expected overhead scales under proactive or reactive geographic routing (Section V) and to derive the scaling law of minimum expected overhead (Section VI).

Lemma 1: The minimum expected number of bits required per timeslot to describe approximated locations of a node under Assumption 1, denoted by $m_{\text{loc}}(n)$, satisfies $m_{\text{loc}}(n) \sim \log(n)$.

Proof: We find lower and upper bounds for $m_{\text{loc}}(n)$ and show that both bounds are asymptotically $\log(n)$.

1. Lower bound: In order to find a lower bound for $m_{\text{loc}}(n)$, consider the following: Suppose that a quantization level of $4\gamma^\star(n)$ is selected for approximating locations and the domain $\mathbb{D}$ is divided into cells of length $4\gamma^\star(n)$, where $\gamma^\star(n) = c^\star\sqrt{\log(n)/n}$ is the CTR introduced in subsection II-B, as shown in Fig. 1.

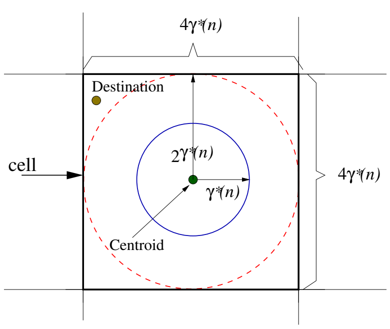

Fig. 1. Partition of $\mathbb{D}$ into cells of length $4\gamma^\star(n)$ on both sides.

Without loss of generality, we assume that the approximated locations of the nodes in a cell with the assumed quantization level are given by the centroid of the cell. This means that a relay node forwarding a packet to the destination shown in the figure will use the location of the centroid as the approximated location of the destination (after decoding the location using the common decoding rule). If none of relay nodes is an immediate neighbor of the destination (which is more likely than not), the packet will eventually enter the inner circle centered at the centroid with radius $\gamma^\star(n)$. Once this happens, the packet cannot be delivered to any node outside the outer (dotted) circle with radius $2\gamma^\star(n)$; the nodes inside the inner circle do not know the precise locations of the nodes outside the outer circle because they are not immediate neighbors. This implies that, without knowing a more precise location of the destination, the entire cell will need to be flooded with the packet before it can reach its destination. This tells us that the quantization level of $4\gamma^\star(n)$ is *not* accurate enough to prevent flooding of the packet.

Let us compute the expected number of bits required per timeslot to describe the locations using this *insufficient* quantization level of $4\gamma^\star(n)$. Under the stated assumptions on stationarity of the mobility processes and spatial density in (4) in subsection III-A, the differential entropy rate of the mobility process [5, p.416]

$$h_\star := \lim_{T\to\infty} \frac{h(L_i^{(n)}(0), L_i^{(n)}(1), \ldots, L_i^{(n)}(T-1))}{T} \quad (5)$$

exists and is bounded below (resp. above) by $-\log(\xi_2)$ (resp. $-\log(\xi_1) < \infty$).

For each $\Delta > 0$, let $L_{i,\Delta}^{(n)}(t)$ be an approximation of $L_i^{(n)}(t)$ with a quantization level Δ; $L_{i,\Delta}^{(n)}(t) = ((k_1+\frac{1}{2})\Delta, (k_2+\frac{1}{2})\Delta)$ if $L_i^{(n)}(t) \in [k_1\cdot\Delta,\ (k_1+1)\Delta) \times [k_2\cdot\Delta,\ (k_2+1)\Delta)$. Denote the *approximated* mobility processes by $\mathbb{L}_{i,\Delta}^{(n)} = \{L_{i,\Delta}^{(n)}(t); t \in \mathbb{Z}_+\}$. From the inherited stationarity of the approximated mobility processes, the entropy rate of $\mathbb{L}_{i,\Delta}^{(n)}$

$$H_\Delta^{(n)} := \lim_{T\to\infty} \frac{H(L_{i,\Delta}^{(n)}(0), L_{i,\Delta}^{(n)}(1), \ldots, L_{i,\Delta}^{(n)}(T-1))}{T} \quad (6)$$

exists [5, Thms 4.2.1 and 4.2.2, p.75]. In addition, the following equality holds [5, Thm 8.3.1, p.248]: For all $T \ge 1$,

$$\lim_{\Delta\downarrow 0}\left(\frac{H(L_{i,\Delta}^{(n)}(0), L_{i,\Delta}^{(n)}(1), \ldots, L_{i,\Delta}^{(n)}(T-1))}{T} + 2\log(\Delta)\right) = \frac{h(L_i^{(n)}(0), L_i^{(n)}(1), \ldots, L_i^{(n)}(T-1))}{T} \ . \quad (7)$$

Equations (5) through (7) imply that, for every $\nu > 0$, there exist $\Delta^*(\nu) > 0$ and $T^*(\nu) < \infty$ such that, for all $\Delta \le \Delta^*(\nu)$ and $T \ge T^*(\nu)$, we have

$$h_\star - 2\log(\Delta) - \nu \le \frac{H(L_{i,\Delta}^{(n)}(0), \ldots, L_{i,\Delta}^{(n)}(T-1))}{T} \le h_\star - 2\log(\Delta) + \nu \ . \quad (8)$$

Substituting $\Delta(n) := 4\gamma^\star(n)$ in place of Δ yields

$$\begin{aligned} &h_\star - 2\log(\Delta(n)) \pm \nu \\ &= h_\star - 2\log(4\gamma^\star(n)) \pm \nu = h_\star - 2\log\left(4\ c^\star\sqrt{\frac{\log(n)}{n}}\right) \pm \nu \\ &= \log(n) - \log(\log(n)) + (h_\star \pm \nu - 4 - 2\log(c^\star)) \ . \end{aligned} \quad (9)$$

Since $h_\star \pm \nu - 4 - 2\log(c^\star)$ are fixed, it is clear from (9) that $h_\star - 2\log(\Delta(n)) \pm \nu \ \sim\ \log(n)$. Together with (8), this proves that, for *all* sufficiently large T,

$$\frac{H(L_{i,\Delta(n)}^{(n)}(0), \ldots, L_{i,\Delta(n)}^{(n)}(T-1))}{T} \sim \log(n) \ . \quad (10)$$

The left hand side of (10) is equal to the minimum expected number of bits we need per timeslot to *jointly* code the locations, $L_{i,\Delta(n)}^{(n)}(0), \ldots, L_{i,\Delta(n)}^{(n)}(T-1)$,[9] using an insufficient quantization level $\Delta(n)$. Hence, it serves as a lower bound to the number of bits we need, and (10) tells us that this lower

[9] Joint coding of the locations of node i requires that, for each $t \in \mathbb{Z}_+$, the sequence of the locations $\{L_{i,\Delta(n)}^{(n)}(0), \ldots, L_{i,\Delta(n)}^{(n)}(t)\}$ be coded together, using a different coding scheme. As a result, such joint coding of node's locations will be difficult to implement in practice.

bound increases (asymptotically) as $\log(n)$.

2. Upper bound: We can obtain an upper bound for $m_{\text{loc}}(n)$ following essentially the same argument used to find the lower bound: Recall that, in order to route a packet to a node i, it suffices to deliver the packet to any immediate neighbor within the transmission range $\gamma^\star(n)$ of node i. As in the previous case of lower bound, suppose that the domain $\mathbb{D}$ is divided into cells of length $\varsigma^{(n)}$, where $\varsigma^{(n)} := \sqrt{2}\gamma^\star(n)/3$. The approximated location of a node in a cell is given by the centroid of the cell. This is shown in Fig. 2.

A packet is relayed using the location of the centroid of the cell in which its destination lies. If none of relay nodes the packet traverses before it enters the cell is an immediate neighbor of its destination, it will eventually be relayed to a node in the same cell as the destination.[10] It is clear from Fig. 2 that, once a packet reaches any node in the same cell as the destination, the node will be able to deliver the packet directly to the destination because the distance between any two nodes in the same cell is bounded by $2\gamma^\star(n)/3$. Therefore, approximating locations with a quantization level of $\varsigma^{(n)}$ is *sufficient* to ensure successful delivery of packets using the approximated location information.

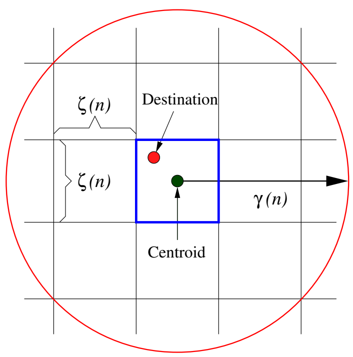

Fig. 2. Partition of $\mathbb{D}$ into cells with area of $\varsigma(n)^2$. ($\gamma^\star(n) = 3\ \varsigma(n)/\sqrt{2}$)

We proceed to compute the average number of bits needed per timeslot to approximate the locations using the quantization level $\varsigma^{(n)}$. From [5, Thm 8.3.1, p.248] and assumed stationarity of the mobility processes, we have

$$\lim_{\Delta \downarrow 0} H(L_{i,\Delta}^{(n)}(t)) + 2\log(\Delta) = h(L_i^{(n)}(t)) \qquad \text{for all } t \in \mathbb{Z}_+ \ .$$

Thus, for every $\nu > 0$, we can find $\Delta^\dagger(\nu) > 0$ such that, for all $\Delta \leq \Delta^\dagger(\nu)$,

$$h(L_i^{(n)}(t)) - 2\log(\Delta) - \nu \leq H(L_{i,\Delta}^{(n)}(t)) \leq h(L_i^{(n)}(t)) - 2\log(\Delta) + \nu \ . \tag{11}$$

Following the same steps in (9), after a little algebra

$$\begin{aligned} &h(L_i^{(n)}(t)) - 2\log(\varsigma^{(n)}) \pm \nu \\ &= h(L_i^{(n)}(t)) - 2\log\left(\frac{\sqrt{2}}{3}c^\star\sqrt{\frac{\log(n)}{n}}\right) \pm \nu \\ &= \log(n) - \log(\log(n)) \\ &\quad +(h(L_i^{(n)}(t)) \pm \nu - 2\log(\sqrt{2}/3) - 2\log(c^\star)) \\ &\sim \log(n) \ . \end{aligned} \tag{12}$$

Therefore, from (11) and (12) we find

$$H(L_{i,\varsigma(n)}^{(n)}(0)) \ \sim \ \log(n) \ . \tag{13}$$

Equation (13) suggests that, even when the locations of a node are coded *separately* at each timeslot, from the assumed ergodicity of mobility processes, the minimum *average* number of bits needed per timeslot to approximate node i's locations with a sufficient quantization level $\varsigma(n)$ is (asymptotically) $\log(n)$. Thus, from the lower and upper bounds in (10) and (13), respectively, one can conclude that the minimum expected number of bits needed per timeslot to describe the locations of a node satisfies $m_{\text{loc}}(n) \sim \log(n)$. ∎

The above proof of Lemma 1 reveals the following interesting observation: In the calculation of $m_{\text{loc}}(n)$, node i's mobility determines the differential entropy of $L_i^{(n)}(t), t \in \mathbb{Z}_+$, and the differential entropy rate $h_\star$ of the mobility process $\mathbb{L}_i^{(n)}$. When the network size is small, the number of bits required to describe node i's locations is mostly governed by these differential entropy and entropy rate that depend on the details of the mobility processes. However, as the number of nodes n grows, in a dense network[11] $m_{\text{loc}}(n)$ is predominantly shaped by the required quantization level for describing the locations of nodes, which is in turn dictated by the CTR needed for network connectivity. As a result, the details of nodes' mobility become less important in a large, dense network, as long as the differential entropy of the locations of nodes and the differential entropy rate of the mobility processes, $h_\star$, are bounded, which is satisfied under the assumption in (4).

A similar result to Lemma 1 can be obtained for the cases where nodes are allowed to use different transmission ranges under the following assumption.

Assumption 2: Suppose that the nodes employ heterogeneous transmission ranges and that there exist constants $c_1, c_2 \in (0, \infty)$ and $0 < a_2 \leq a_1 < \infty$ such that the transmission ranges of the nodes can be lower and upper bounded by $c_1 \cdot n^{-a_1}$ and $c_2 \cdot n^{-a_2}$, respectively, for all sufficiently large $n \in \mathbb{N}$.

Corollary 1: The minimum expected number of bits required per timeslot to describe approximated locations of a node under Assumption 2, denoted by $m_{\text{loc}}^\star(n)$, is $\Theta(\log(n))$.

The proof of the corollary is essentially the same as that of Lemma 1; we can show that the quantization level of $4 \cdot c_2 \cdot n^{-a_2}$ is not accurate enough, whereas $\sqrt{2} \cdot c_1 \cdot n^{-a_1}/3$ is a sufficient quantization level. These quantization levels give us asymptotic lower and upper bounds of $2 \cdot a_2 \cdot \log(n)$ and $2 \cdot a_1 \cdot \log(n)$, respectively, for $m_{\text{loc}}^\star(n)$. As we will see, this important observation allows us to relax Assumption 1 without voiding our findings in the following sections (Theorems 2 through 4).

[10]Here we assume that there is a node in the cell with a high probability. We will revisit this issue in Section VI and show that the probability that there is no node in the cell goes to zero as $n \to \infty$.

[11]A similar result can be obtained for extended networks with increasing domains.

V. Routing overhead under proactive and reactive geographic routing

In this section we examine how the expected routing overhead scales when proactive or reactive geographic routing is employed and address the issue of how to measure the total distance traveled by control messages. Recall from Section I that a geographic routing scheme is called a *proactive geographic* routing scheme if each node attempts to maintain consistent, up-to-date location information for *every* known destination in the network by flooding the network with *location update messages*. Similarly, a geographic routing scheme is said to be a *reactive geographic* routing scheme if location information is provided only when it is requested. When no location information of a desired destination is available at a source, the location information is discovered by flooding the network with a *location request message* until another node, possibly the destination itself, replies to the request with location information. We point out that these proactive or reactive *geographic* routing schemes are different from the traditional proactive or reactive routing algorithms that use topological information.

A. Routing overhead under proactive geographic routing

Suppose that location information of a node is forwarded to and stored at all other nodes within distance $\epsilon > 0$. If $\epsilon \geq \sqrt{2} \cdot \overline{D}$, the location information of every node is forwarded to *all* nodes in the network. This is because the distance between any two points in $\mathbb{D}$ is upper bounded by $\sqrt{2} \cdot \overline{D}$, i.e., $\sup_{\mathbf{x},\mathbf{y} \in \mathbb{D}} ||\mathbf{x} - \mathbf{y}|| = \sqrt{2} \cdot \overline{D}$, where $||\mathbf{x} - \mathbf{y}||$ is the Euclidean distance between $\mathbf{x}$ and $\mathbf{y}$. First, it is clear that, under our assumptions in subsection III-B, at least $\log(n)$ (and at most $\log(n) + 1$) bits are required to identify the source of a message.

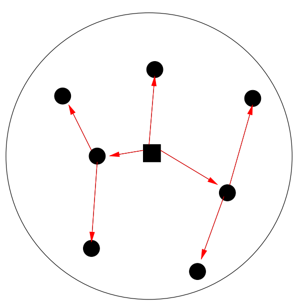

Fig. 3. Total distance traveled by a location message.

The total distance traveled by a location update message from a node, say node i, to all its neighbors within distance ϵ can be computed in different ways. In this paper, we take the viewpoint that once a neighbor receives the location information of node i, it can serve as a surrogate source of the location information for other nodes. This is shown in Figure 3. It is more consistent with the operation of a multi-hop wireless network where each relay node is responsible for delivering a packet to the next hop, and thus each transmitter-receiver pair can be viewed as a source-destination pair for the purpose of exchanging location information.

If we count only the first copy that arrives at each node, the total distance traveled by a location update message to all the nodes within distance ϵ is given by the total length of a *spanning tree* constructed by the propagation of the message, which connects all the nodes within ϵ. Obviously, this distance is lower bounded by the total length of a *minimum* spanning tree (MST). In fact, by the definition of an MST, the total length of an MST is the minimum among all (reasonable) measures of the total distance connecting all neighbors.

Theorem 2: The minimum expected overhead required per timeslot under Assumption 1 for *disseminating* location information in proactive geographic routing is $\Omega(n^{1.5} \log(n))$.

Proof: Let us first introduce a lemma that will be used in the proof of the theorem. Suppose $\{X_n; n \in \mathbb{N}\}$ is a sequence of rvs, where X_n is a binomial(n, p) rv with $0 < p < 1$.

Lemma 2: Define $Z_n^\alpha := \left(\frac{X_n}{n \cdot p}\right)^\alpha$, where $0 < \alpha \leq 1$. Then,

$$\lim_{n \to \infty} \mathbf{E}\left[Z_n^\alpha\right] = 1 \qquad \text{for all} \;\; 0 < \alpha \leq 1 \; .$$

Proof: Let $Y_n := Z_n^1 = \frac{X_n}{n \cdot p}$. The strong law of large numbers [9, p.326] tells us that Y_n converges to 1 in mean square, i.e., $\mathbf{E}\left[|Y_n - 1|^2\right] \to 0$ as $n \to \infty$. Since $|Z_n^\alpha - 1| \leq |Y_n - 1|$ for all $0 < a \leq 1$, we have

$$\mathbf{E}\left[|Z_n^\alpha - 1|^2\right] \leq \mathbf{E}\left[|Y_n - 1|^2\right] \to 0 \qquad \text{as} \;\; n \to \infty,$$

which implies $Z_n^\alpha \to 1$ in mean square (clearly, $\mathbf{E}\left[(Z_n^\alpha)^2\right] \leq 1 + \mathbf{E}\left[(Y_n)^2 \cdot \mathbb{1}\{Y_n > 1\}\right] < \infty$ for all $n \in \mathbb{N}$).

Recall that convergence in mean square implies convergence in mean [9, p.310]. Hence,

$$\mathbf{E}\left[|Z_n^\alpha - 1|\right] \to 0 \qquad \text{as} \;\; n \to \infty.$$

Theorem 3 [9, p.351] tells us that $Z_n^\alpha \to 1$ in mean *if and only if* $\mathbf{E}\left[Z_n^\alpha\right] \to 1$ as $n \to \infty$ (and, equivalently, $\{Z_n^\alpha; n \geq 1\}$ is uniformly integrable). This completes the proof of the lemma. ■

We now proceed with the proof of Theorem 2. Steele [31] showed the following result on the total length of an MST with an increasing number of nodes: Suppose that nodes are placed independently of each other in accordance with distribution μ with compact support $\mathbb{S} \subset \mathbb{R}^2$. Let $M(n)$ denote the total length of an MST connecting the first n nodes. Then, with probability 1,

$$\lim_{n \to \infty} \frac{M(n)}{\sqrt{n}} = e^\star \int_{\mathbf{x} \in \mathbb{S}} \sqrt{g(\mathbf{x})} \; d\mathbf{x} \tag{14}$$

for some constant $e^\star$, where g is the density of the absolutely continuous part of μ. In other words, the total length of an MST is asymptotically proportional to $\sqrt{n}$.

From the assumed mutual independence and stationarity of the mobility processes $\mathbb{L}_i^{(n)}, i \in N^{(n)}$, the number of nodes in an area $\tilde{D} \subset \mathbb{D}$ at timeslot t is a binomial rv with parameter $(n, p_{\tilde{D}})$, where $p_{\tilde{D}} = \int_{\tilde{D}} f(\mathbf{y}) \; d\mathbf{y}$. Let $d_\epsilon(\mathbf{x})$ denote the intersection of the mobility domain $\mathbb{D}$ and the disk centered at $\mathbf{x} \in \mathbb{D}$ with radius ϵ. Then, for every $\mathbf{x} \in \mathbb{D}$, Area$(d_\epsilon(\mathbf{x})) \geq \pi \; \epsilon^2/4$, hence $\xi_1 \; \pi \; \epsilon^2/4 \leq \int_{d_\epsilon(\mathbf{x})} f(\mathbf{y}) \; d\mathbf{y} \leq \xi_2 \; \pi \; \epsilon^2$.

This observation, combined with the result by Steele in (14) and Lemma 2 with $\alpha = 1/2$, suggests that the expected total length of an MST that connects all nodes in $d_\epsilon(\mathbf{x})$ is asymptotically proportional to $\sqrt{n}$ for all $\mathbf{x} \in \mathbb{D}$. Therefore, from the assumed ergodicity and mutual independence of the mobility processes, the average total distance traveled by a location update message of node i to its neighbors within a fixed distance ϵ is $\epsilon^2 \cdot \Omega(\sqrt{n})$.

Since (i) there are n nodes that move according to mutually independent mobility processes, (ii) each message requires at least $\log(n)$ bits to identify the source of the message, (iii) location information of the source needs asymptotically $\log(n)$ bits from Lemma 1, and (iv) the average total distance traveled by a location message is $\Omega(\sqrt{n})$, the expected routing overhead (measured in bits×meters per unit time) for disseminating location information to local neighborhoods per timeslot under proactive geographic routing is $\Omega(n \cdot \log(n) \cdot \sqrt{n}) = \Omega(n^{1.5} \log(n))$. ■

B. Routing overhead under reactive geographic routing

As stated earlier we assume that, under reactive geographic routing, if location information is not available at a source when a connection request arrives, it generates a location request message and floods the network. When a node with the requested location information receives the request message, it generates a *location reply message* with the location information. In this subsection, we study the expected overhead due to the location request messages and location reply messages under reactive geographic routing.

In practice there may be additional overhead due to location recovery when a destination moves while the connection is active and the source does not know the new location of the destination. However, we do not study the overhead due to the recovery of location information while connections are still active. We will discuss this issue in subsection VII-C.

In order to make progress, we introduce following simplifying assumptions:

A1. Only the destination for which a location request is generated responds with a reply message;
A2. Location request messages reach the nodes in the order of increasing distance from their sources.

Assumption A2 implies that if a node that generates a reply message is at distance $\overline{d}$ from the source, the request message reaches all the nodes within distance $\overline{d}$ from the source.

Under our assumption in subsection III-B that $A_i^{(n)}(t)$ are i.i.d. Bernoulli rvs, at most one request is generated for a connection to node i in each timeslot. Thus, no other node will have cached up-to-date location of node i. However, when more than one node can generate a connection request to node i in a timeslot, it is possible that some other nodes that acquired node i's location information may cache the location information, and a reply can be generated by another node with cached location information. In this case, we can replace Assumption A1 with the following alternate assumption, without modifying our findings below:

A1a. Suppose that the location of a source generating a location request message is $\underline{\ell} \in \mathbb{D}$. Then, the location of the *closest* node that generates a reply message depends only on $\underline{\ell}$ and has distribution $M(\cdot, \underline{\ell})$.

Assumption A1a means that the distance to the closest node sending a reply does *not* depend on the number of nodes in the network. This may be reasonable when the location information of each node is available only at a limited number of other nodes, in particular in a small neighborhood around the node. When only the destination is allowed to generate a reply message, Assumption A1a holds by virtue of mutual independence of the mobility processes. Here, we assume that Assumption A1 (instead of Assumption A1a) is in place.

Theorem 3: The minimum expected overhead required per timeslot under Assumption 1 for location request and reply messages in reactive geographic routing is $\Omega(n^{1.5} \log(n))$.

Proof: We examine the routing overhead that arises from location requests and replies separately. We first show that the expected overhead due to handling location requests is $\Omega(n^{1.5} \cdot \log(n))$, and then demonstrate that the expected overhead from location replies is $\Theta(n \cdot \log(n))$.

First, each location request message must have the ID of the destination, which requires at least $\log(n)$ bits. Second, analogously to the proactive geographic routing case, the total distance traveled by a location request message to all the nodes within the distance to the destination is lower bounded by the total length of an MST connecting the nodes. Under these assumptions, by conditioning on the distance to the destination and following the same argument used in the proof of Theorem 2, one can show that the expected total length of such an MST is $\Omega(\sqrt{n})$. Therefore, since location requests arrive at a rate of p at each node, the expected overhead for handling the request messages is $\Omega(n \cdot \log(n) \cdot \sqrt{n}) = \Omega(n^{1.5} \ \log(n))$.

Unlike location requests, location replies need not be flooded.[12] Also, because (i) the source of a request for the location information of node i is equally likely to be any of the other $n-1$ nodes, (ii) spatial density f does not vary with n, (iii) connection request processes $\mathbb{A}_i^{(n)}, i \in N^{(n)}$, are independent of the mobility processes, and (iv) the mobility processes $\mathbb{L}_j^{(n)}, j \in N^{(n)}$, are assumed *stationary* and *ergodic* and are also *mutually independent*, the *average* distance between the sources and the destinations (averaged over all timeslots and all source-destination pairs) is equal to the *expected* distance between a pair of randomly selected nodes. This expected distance is given by[13]

$$d_{\text{avg}} \equiv \int_{\mathbb{D}} \int_{\mathbb{D}} ||\mathbf{x} - \mathbf{y}|| \ f(\mathbf{x}) \ f(\mathbf{y}) \ d\mathbf{y} \ d\mathbf{x} > 0 \ . \qquad (15)$$

The inequality follows from the assumption $\inf_{\mathbf{x} \in \mathbb{D}} f(\mathbf{x}) \geq \xi_1 > 0$ in (4) with $t = 0$. Note that d_{avg} does *not* depend on the number of nodes n. Since reply messages must carry the ID of the source and the location information of the destination (and the source), the overhead due to reply messages is $d_{\text{avg}} \cdot \Theta(n \cdot \log(n))$. Therefore, the overall routing overhead under reactive geographic routing is $\Omega(n^{1.5} \cdot \log(n))$. ■

It is clear from the proof of Theorems 2 and 3 that the derived scaling laws for the expected overhead under

[12]Replies can be routed back either by using the location information of the sources attached to the request messages or by maintaining a cache at intermediate nodes which temporarily stores all request messages received over a sliding time window along with the first nodes that forwarded them.

[13]When Assumption A1a is in place instead and node i is a node with cached location information of the requested destination, (15) is replaced by $d_{\text{avg}} \equiv \int_{\mathbb{D}} (\int_{\mathbb{D}} ||\mathbf{x} - \mathbf{y}|| \ m(\mathbf{y}, \mathbf{x}) \ d\mathbf{y}) \ f(\mathbf{x}) \ d\mathbf{x} > 0$, where $m(\cdot, \mathbf{x})$ is the derivative of $M(\cdot, \mathbf{x})$.

proactive and reactive geographic routing do *not* change when Assumption 1 is replaced by Assumption 2. This is because the average number of bits in control messages remains $\Theta(\log(n))$ from Corollary 1.

In the following section, we will show that, compared to the minimum expected routing overhead, both proactive and reactive geographic routing suffers a penalty of at least $\sqrt{n}$ for flooding the network with either the location information of nodes (proactive geographic routing) or location request messages (reactive geographic routing). This also hints that if we eliminate or reduce flooding of messages, we can alter the way routing overhead scales with the increasing network size.

VI. MINIMUM EXPECTED ROUTING OVERHEAD

In this section we examine how the *minimum* expected routing overhead from location service scales with the number of nodes under the assumptions stated in Section III: For each fixed $n \in \mathbb{N}$, let us denote the minimum expected overhead required per timeslot for *disseminating* and *acquiring* location information under Assumption 1 by $R_{\min}(n)$. We prove that $R_{\min}(n) = \Theta(n \cdot \log(n))$ in two steps: First, we show that $R_{\min}(n)$ increases at least as $\alpha \cdot n \cdot \log(n)$ for some constant α, i.e., $R_{\min}(n) = \Omega(n \cdot \log(n))$. Second, we demonstrate that, for all sufficiently large n, the minimum expected routing overhead is upper bounded by $\beta \cdot n \cdot \log(n)$ for another constant β, proving $R_{\min}(n) = O(n \cdot \log(n))$. These two findings yield our claim that $R_{\min}(n) = \Theta(n \cdot \log(n))$.

Lemma 3: The minimum expected overhead for location service per timeslot under Assumption 1, $R_{\min}(n)$, is $\Omega(n \cdot \log(n))$.

Proof: Let us first focus on a single connection request originating, say, at node $k \in N^{(n)}$, with node $i, i \neq k$, as the destination. First, *any* location message of node i must carry its ID and location. As mentioned earlier, a minimum of $\log(n)$ bits are needed to identify node i in the message, and from Lemma 1, the minimum expected number of bits required on average to describe the locations of node i asymptotically approaches $\log(n)$. Second, the *expected* distance the location message of node i must travel from node i to node k is given by d_{avg} in (15) and does not depend on n.

Summarizing these, (i) for the same reason provided before (15) in the proof of Theorem 3, the average distance the location messages have to travel from the destinations to the sources of connection requests equals $d_{\text{avg}} > 0$, and (ii) the expected number of bits in each location message required for both the ID and the location of a destination is $\Theta(\log(n))$. Since the average total number of connection requests in a timeslot equals $n \cdot p$, the minimum expected routing overhead required on average (in bits$\times$meters per unit time) for delivering location information from the destinations to the sources of connection requests is $\Theta(n \cdot \log(n))$. Obviously, the minimum expected routing overhead for location service cannot be smaller than the overhead required for transporting location information directly from the destinations to the sources. Hence, $R_{\min}(n) = \Omega(n \cdot \log(n))$. ∎

Lemma 4: The minimum expected overhead for location service per timeslot under Assumption 1, $R_{\min}(n)$, is $O(n \cdot \log(n))$.

Proof: In order to prove the lemma, it suffices to find a scheme under which the expected routing overhead per timeslot is upper bounded by $\beta \cdot n \cdot \log(n)$ for all sufficiently large n, for some finite constant $\beta > 0$. The scheme we describe here combines the features of both proactive and reactive geographic routing schemes in such a way we can avoid expensive *multi-hop* flooding of messages, by forming *virtual* location servers using the existing nodes. A similar idea of using existing nodes as location servers without knowing their identities was used by Li et al. [22].

A key idea is that we store the location information of each node i in a small region (relative to the transmission range) so that once a location request message for node i reaches *some* node in the region, the node, if it does not have the location information of node i, can find another node with the location information *without* having to flood a multi-hop neighborhood. In this sense, the set of nodes in the region, which varies with time, *collectively* serve as a virtual location server on behalf of node i. Hence, individual nodes participate not only in routing packets, but also in providing location service for other nodes.

To this end, we choose a quantization level of $\varsigma^{(n)} = \sqrt{2}\gamma^{\star}(n)/3$ for approximating node locations[14], divide the domain of mobility into *cells* of area $\mathcal{A}(n) = \varsigma^{(n)} \times \varsigma^{(n)}$, and store the location information of each node in a cell with a *known* coordinate.[15] The coordinate of the cell where the location information of node i resides, is computed using a hash function $h^{(n)} : N^{(n)} \rightarrow S_{h^{(n)}}$, where $S_{h^{(n)}}$ is the set of coordinates of the cells that hold location information. This allows us to skirt the problem of not having location servers with known or fixed locations. In addition, the hash functions can be designed to distribute the load of storing location information among the nodes. The hash functions $h^{(n)}$ are assumed common knowledge.

First, if we are to store location information in a cell, we must ensure that there is at least one node in the cell (with probability approaching 1 as $n \rightarrow \infty$) so that location information can be stored in the cell and be accessible to other nodes. It is obvious that $\mathcal{A}(n) = 2\ c^{\star 2}\ \log(n)/(9\ n) = \omega(1/n)$. From the assumption on the spatial distribution in (4) and mutual independence of the mobility processes, for the given cell size $\mathcal{A}(n)$ the probability that there is no node in a cell at timeslot $t \in \mathbb{Z}_+$ approaches zero as $n \rightarrow \infty$:

$$\mathbf{P}\left[\text{No node in a cell at timeslot } t\right] \leq (1 - \xi_1 \cdot \mathcal{A}(n))^n$$

For $\mathcal{A}(n) = 2\ c^{\star 2}\ \log(n)/(9\ n)$,

$$\begin{aligned}(1 - \xi_1 \cdot \mathcal{A}(n))^n &= \exp\left(n \cdot \log(1 - \xi_1 \cdot \mathcal{A}(n))\right) \\ &\rightarrow 0 \quad \text{as} \quad n \rightarrow \infty .\end{aligned}$$

In the rest of the proof we describe how the location information is disseminated and retrieved by nodes and compute the overhead due to these operations.

1. Dissemination of location information: As mentioned earlier, under our scheme, we convey and store the location

[14]Recall from the proof of Lemma 1 that $\varsigma^{(n)}$ is sufficient for enabling geographic routing.

[15]By storing location information in a cell, we mean storing it at one or more nodes in the cell.

information of a node i in cell $h^{(n)}(i)$ that serves as a virtual location server for node i. Intermediate nodes route a location message of node i using the location of the cell $h^{(n)}(i)$ computed using the common hash function.[16] Unlike a unicast data packet that is routed to a specified destination node, however, a location update message of node i does not include a specific destination ID in the message. This is because node i is unlikely to know in advance which nodes are in cell $h^{(n)}(i)$. Instead, since there is a node in cell $h^{(n)}(i)$ (with probability approaching 1), when a location update message reaches *some* node in cell $h^{(n)}(i)$, the node stores the location information of node i and terminates the message without relaying it further. Obviously, the distance traveled by a location message of node i to any node in cell $h^{(n)}(i)$ is upper bounded by $\sqrt{2} \cdot \overline{D}$.

A location update message of a node carries the node's ID and location information. Combining with our finding in (13) that the minimum average number of bits needed to describe the locations of a node with quantization level $\varsigma(n)$ is asymptotically $\log(n)$, we conclude that the routing overhead due to transporting the location information of the nodes to their respective cells that store their location information is $R_T(n) = \Theta(n \cdot \log(n))$.

2. Retrieval of location information: In order for a node j to access the location information of another node, say node i, node j first generates a location request with (i) its own ID and location information (with the same quantization level $\varsigma(n)$), and (ii) the ID of node i. The request message is then relayed by intermediate nodes using the location of the cell $h^{(n)}(i)$ computed from the ID of node i in the request message and the common hash function, until it reaches *some* node in cell $h^{(n)}(i)$.

When the request message arrives at a node in cell $h^{(n)}(i)$, one of following two events occurs: (i) If the node has the location information of node i, it generates a reply message, or (ii) if it does not, it broadcasts the request message to its neighbors in cell $h^{(n)}(i)$, all of which lie *within* its transmission range. In the latter event, since there is at least one node in the cell $h^{(n)}(i)$ with the location information of node i (with probability approaching 1), another node in cell $h^{(n)}(i)$ with the location information generates a reply message. Again, the reply message is heard by all other nodes in the cell because they are all within the transmission range, hence only a single reply message is generated.

In the case of second event, compared to the first, one additional broadcast transmission is required. However, there is no need to flood a *multi-hop* neighborhood in search of a node with location information (which is the case with reactive geographic routing). The total distance traveled by the broadcast request message over the last hop to all the nodes in cell $h^{(n)}(i)$ can be computed as follows: From the assumed mutual independence of the mobility processes, the number of nodes in cell $h^{(n)}(i)$ is a binomial$(n, \tilde{p})$ rv, where $\tilde{p}$ is the steady-state probability that a node is in cell $h^{(n)}(i)$. Recall that (i) from the assumption on spatial distribution in (4), $\tilde{p}$ is upper bounded by ξ_2 $\times$ area of a cell $(= \xi_2 \times 2\ c^{\star 2}\ \log(n)/(9 \cdot n))$ and (ii) the distance from the last relay node to any node in $h^{(n)}(i)$ that hears the message is bounded by the transmission range $\gamma^\star(n)$. Thus, the total expected distance from the last relay node to all nodes in $h^{(n)}(i)$ is upper bounded by

$$\begin{aligned} n \cdot \tilde{p} \cdot \gamma^\star(n) &\leq \frac{2\ \xi_2\ c^{\star 2}\ \log(n)}{9} \times c^\star \sqrt{\frac{\log(n)}{n}} \\ &= \frac{2\ \xi_2\ c^{\star 3}\ \log^{1.5}(n)}{9\ \sqrt{n}}\ . \end{aligned} \tag{16}$$

It is clear that (16) decreases to zero as $n \to \infty$. This tells us that the contribution from the last hop to the total expected distance traveled by a request message vanishes as $n \to \infty$, and that the total expected distance is $\Theta(1)$.

A reply message produced in response to a request message contains the IDs and location information of both nodes j and i. The reply message is then routed back to the source (i.e., node j), using the location information of node j copied from the request message. It is obvious that the expected distance traveled by a reply message is $\Theta(1)$.

Since a location request message generated by node j contains the IDs of both nodes j and i and the location information of node j with quantization level $\varsigma(n)$, the required expected number of bits in a location request is on average $b_{\text{req}}(n) \sim 3\log(n)$ from (13). Similarly, the expected number of bits required in reply messages for the IDs and location information of both nodes is on average $b_{\text{res}}(n) \sim 4\log(n)$. Hence, the expected number of bits needed for handling a single location request under our scheme is on average $b(n) = b_{\text{req}}(n) + b_{\text{res}}(n) \sim 7\log(n)$. Recall from Section III that each node generates route request messages at a rate of p requests per timeslot. Together with earlier findings on the expected total distance traveled by request and reply messages and their sizes, we conclude that the expected routing overhead incurred per timeslot due to retrieval of location information under our scheme is $R_A(n) = \Theta(n \cdot \log(n))$.

The minimum expected routing overhead $R_{\min}(n)$ is obviously not greater than the expected routing overhead incurred by our scheme, which is $R^\star(n) \equiv R_T(n) + R_A(n)$. Since $R^\star(n) = \Theta(n \cdot \log(n))$, we have $R_{\min}(n) = O(n \cdot \log(n))$. ∎

We note that nodes actively disseminate their location information to parts of the network under both proactive geographic routing (subsection V-A) and our scheme in the proof of Lemma 4. However, there are some key differences between our scheme and both proactive and reactive geographic routing: First, proactive geographic routing floods and stores location information in the neighborhood around the nodes, whereas in our scheme the location information of a node is stored only in a small area (a cell) with a pre-assigned location that can be computed using its ID, *independently of its actual location*. Since nodes are mobile, unless sources are always close to selected destinations, promulgating location information to a small neighborhood around the nodes will be of limited use. Secondly, we limit the area to be flooded with a location request message to the same cell. In other words, only the cell in which the location information of a requested destination is stored, is flooded with the location

[16]The location of the cell is the same as the approximated location of a node in the cell, i.e., the centroid of the cell, as explained in Section IV.

request message. These simple features eliminate the need for unnecessary and expensive flooding of control messages in the network, resulting in lower overhead.

It is noteworthy that, in computing the routing overhead under both our scheme and proactive/reactive geographic routing, flooding of location information or request messages changes only the distances traveled by them, while the number of bits carried by them remains the same. Therefore, the disparity in expected routing overhead is caused only by larger distances traveled by control messages under proactive/reactive geographic routing (which demands higher resource expenditure by their transmissions). Therefore, this highlights the importance of modeling and accounting for the traveled distances; computing only the information rate required to model nodes' mobility and uncertainty in their locations (e.g., [3]) would *not* reveal this discrepancy in (the scaling law of) expected routing overhead under these schemes.

Theorem 4: The minimum expected overhead for location service per timeslot under Assumption 1, $R_{\min}(n)$, is $\Theta(n \cdot \log(n))$.

Proof: The theorem follows from Lemmas 3 and 4. ■

Corollary 2: The minimum expected overhead for location service per timeslot under Assumption 2 is $\Theta(n \cdot \log(n))$.

This corollary follows from the observation that the average number of bits in control messages is still $\Theta(\log(n))$ under Assumption 2 (as a consequence of Corollary 1) and a minor modification of the proof of Lemma 4.

VII. Discussion

Throughout this paper we assumed geographic routing and a non-vanishing spatial density of the nodes while adopting the RGG model for one-hop network connectivity. In this section we first compare the expected overhead of geographic routing schemes to that of topology-based routing schemes. Then, we examine the effects of a vanishing spatial density and location recovery procedures on routing overhead. Finally, we consider a family of network connectivity models, which contains the RGG model as a special case, and show that our results still hold under the new models.

A. Geographic routing vs. topology-based routing: In Section V we showed that the expected routing overhead under proactive or reactive geographic routing is $\Omega(n^{1.5} \cdot \log(n))$. Here, we briefly discuss the same under a proactive or reactive routing scheme that uses *topological* information of the network (i.e., network connectivity) for routing decisions: Each node maintains and uses the next-hop information, for example, along a minimum-hop path, for each known destination through exchange of (local) topology information. We call these routing schemes *topological* routing schemes.

Proactive topological routing: Suppose that, under a proactive topological routing scheme, each node advertises the IDs of its neighbors along with its own ID to all other nodes within distance $\epsilon > 0$, which we call an *advertisement*. The information on immediate neighbors is the minimal amount of information needed to reconstruct the network topology and is the same information reported by regular nodes in [35], [36]. Given the assumptions in Section III and the transmission range $\gamma^\star(n)$, the expected number of neighbors of a node is $\Theta(\log(n))$. Thus, the average number of bits required for an advertisement containing the list of neighbors is $\Theta(\log(n)^2)$ because each ID requires on the average $\log(n)$ bits. From the proof of Theorem 2, we know that the expected total distance traveled by each advertisement is $\Omega(\sqrt{n})$. Therefore, the overall expected overhead due to advertisements per timeslot is $\Omega(n \cdot \sqrt{n} \cdot \log(n)^2) = \Omega(n^{1.5} \cdot \log(n)^2)$.

Zhou and Abouzeid [36] studied similar routing overhead under two-tier hierarchical proactive routing where regular nodes report the detailed local topology information to their cluster heads that maintain global ownership information. When the number of subregions M (with one cluster head per subregion) is fixed, the overall routing overhead is $\Omega(n^2 \cdot \log(n))$ under all three different physical scalings of the network they considered (Table IV in [36]). In particular, under the second physical scaling in which the communication range of the nodes is adjusted so that the expected number of neighbors of a node is $\Theta(\log(n))$, the routing overhead is $\Theta(n^{2.5})$. Furthermore, even when the number of subregions M is allowed to depend on n, one can show that the overall routing overhead is $\Omega(n^{1.5})$ under all three different physical scalings in [36] (and $\Theta(n^2/\sqrt{\log(n)})$ under the second physical scaling).

Reactive topological routing: Assume that routing information is discovered by flooding the network with a *route request message* under a reactive topological routing scheme, and Assumptions A1 and A2 in subsection V-B hold (with 'location request' replaced by 'route request'). Then, the overhead stemming from flooding of route request messages is $\Omega(n^{1.5} \cdot \log(n))$ by a similar argument in the proof of Theorem 3. As mentioned in the same proof, replies need not be flooded. Instead, they can be routed back to the source by maintaining a cache at intermediate nodes and temporarily storing all request messages with the IDs of the sources and the first nodes that forwarded them. Then, following a similar reasoning, one can show that the overhead due to reply messages is $\Theta(n \cdot \log(n))$, giving the overall routing overhead of $\Omega(n^{1.5} \cdot \log(n))$.

We also note that introducing virtual servers with *route information* to nodes (analogous to the virtual location server in the proof of Lemma 4) will be problematic in topological routing schemes. This is because, unlike in geographic routing where the *same* location information for node i can be provided to any node that wishes to communicate with node i, the end-to-end route information to node i varies from one node to another, depending on the position of the node in the network topology relative to that of node i. Hence, it is not obvious how one can reduce the routing overhead brought about by costly flooding of control messages.

B. Vanishing spatial density: When the spatial density of the nodes is non-vanishing, it is easy to see from (3) that the CTR is proportional to $1/\sqrt{f_\star}$, where $f_\star$ is the infimum of the density function. Thus, the uniform distribution requires the smallest CTR among all distributions with non-vanishing density. The intuition behind this finding is that the transmission range must be increased to maintain a certain level of degree (i.e., the number of immediate neighbors) of the nodes in lower density regions.

In [16] Han and Makowski used a simple example to illustrate the effects of vanishing density on network connectivity: Suppose that n nodes are placed on a unit interval [0, 1] according to a distribution $F_p(x) = x^{1+p}$, $x \in [0, 1]$, where $p > 0$. The distribution yields a continuous density function $f_p(x) = (1+p)x^p$, $x \in [0, 1]$, which vanishes at $x = 0$. When n nodes are distributed according to this density function, the CTR is given by $n^{-\frac{1}{1+p}}$ (as opposed to $\frac{\log(n)}{n}$ in the case of uniform distribution [1]). This demonstrates not only that the presence of vanishing point(s) significantly alters the way the CTR decreases with an increasing number of nodes, but also that how it decreases depends on how the density approaches zero around the vanishing points. Moreover, the CTR $n^{-\frac{1}{1+p}}$ is of a larger order than the CTR of $\log(n)/n$ with the uniform distribution, i.e., $\frac{\log(n)}{n} = o(n^{-\frac{1}{1+p}})$.

The findings by Han and Makowski and by Penrose suggest that the order of the CTR may be the smallest when the spatial distribution is uniform. This means that the average number of bits required to describe the locations of a node is $O(\log(n))$. Therefore, since the number of bits needed to identify a node is lower bounded by $\log(n)$ when source-destination pairs are selected randomly, it is likely that the minimum expected routing overhead is $\Theta(n \cdot \log(n))$ under a large class of spatial distributions (or mobility models) of nodes.

C. Overhead due to location recovery in reactive geographic routing: As discussed in subsection V-B, suppose that a destination of a connection moves while it is still active. In this case, unless the destination informs the source of its new location, the location information at the source will be outdated and the source will need to acquire the new location of the destination through a recovery process. If we assume that the recovery is performed by flooding a control message similar to the original location request message, then the additional overhead due to recovery will be comparable to the overhead incurred during the original location discovery process (through location request and reply messages). Thus, if we assume that connections need, on average, K recovery processes (per connection) while they are active, the expected routing overhead will scale by a factor of K, and the scaling law of the expected routing overhead will remain $\Omega(n^{1.5} \ \log(n))$. In practice, however, the frequency of location recovery will depend on the details of an adopted routing scheme.

D. Different network connectivity models and choices of transmission ranges: While we modeled the network connectivity using an RGG model so far, our results can be generalized to other connectivity models: Given $n \in \mathbb{N}$ nodes in the network, let $\overline{\gamma}(n)$ be a *target* transmission range selected by the nodes. There exist constants $0 < \sigma_1 \leq 1 \leq \sigma_2 < \infty$ so that, given $\overline{\gamma}(n)$, (i) nodes i and j have a link if their distance $d(i,j) \leq \sigma_1 \cdot \overline{\gamma}(n)$, and (ii) they do not have a link if $d(i,j) > \sigma_2 \cdot \overline{\gamma}(n)$. When $\sigma_1 \cdot \overline{\gamma}(n) < d(i,j) \leq \sigma_2 \cdot \overline{\gamma}(n)$, however, we do not specify whether or not there exists a link between nodes i and j. Different rules, such as a probabilistic rule, can be applied to this case. The RGG model is a special case with $\sigma_1 = \sigma_2$. The interpretation of this family of models is that once nodes select a *target* transmission range, they should be able to communicate directly with other nodes that are well within the target range, whereas other nodes that are (much) farther away than the target range would not be directly reachable. Connectivity between nodes roughly target range away from each other, however, may depend on other factors, and we do not provide a specific rule for this case.

From Theorem 1 and the above rules, the minimum target transmission range required for network connectivity satisfies $\gamma^\star(n)/\sigma_2 \leq \overline{\gamma}(n) \leq \gamma^\star(n)/\sigma_1$. If this condition is met, following the proof of Lemma 1, one can show that the necessary quantization level for approximating location information is $\Theta(\gamma^\star(n))$ and Lemma 1 still holds: When the target range $\overline{\gamma}(n) = \Theta(\gamma^\star(n))$, the necessary quantization level changes at most by a constant factor from the RGG case. Thus, as mentioned in Sections IV and VI, this does not affect the findings in Lemma 1 and, hence, Theorems 2 through 4.

Under a quasi unit disk graph (QUDG) model [20], presumably with fixed transmit power, there are two thresholds – $0 \leq \gamma_1 \leq \gamma_2 < \infty$, where $\gamma_1 = \tau \cdot \gamma_2$ for some $\tau \in [0, 1]$. (i) If the distance $d(i,j)$ between nodes i and j is at most γ_1, there is a link between i and j; (ii) if $d(i,j) > \gamma_2$, no link exists between them; and (iii) if $\gamma_1 < d(i,j) \leq \gamma_2$, there may or may not exist a link between them. It is obvious that, under suitable scaling of γ_1 and γ_2 (through transmit power control) as a function of n while maintaining network connectivity, the QUDG model is similar to the above model. Hence, our results are true under the QUDG model when $\tau > 0$.

Under a cost-based model (e.g. [29]), there is a cost function $c : \mathbb{R}_+ \to \mathbb{R}_+$ such that, (i) the cost at distance d is given by $c(d) \in [\varphi_1 \cdot d, \ \varphi_2 \cdot d]$, where $0 < \varphi_1 \leq \varphi_2 < \infty$, and (ii) nodes i and j have a link if and only if $c(d(i,j)) \leq c_{th}$ for some threshold c_{th}. The cost function c is not assumed monotonic in distance. It is clear that, given a threshold c_{th}, we can find an upper and lower bound on the maximum distance between two nodes that would permit a link between the nodes. Therefore, by selecting appropriate thresholds $c_{th}(n)$ as a function of the number of nodes n that would ensure network connectivity and following a similar reasoning as above, we can show that our results hold under this model as well.

VIII. Conclusion

We investigated the expected routing overhead in MANETs employing geographic routing, with an emphasis on the overhead rising from dissemination and retrieval of location information. We focused on a scenario where packets can be routed to their intended destinations using only the ID and location information of the destinations without being flooded. We showed that when nodes move independently while employing a common transmission range to ensure network connectivity (with a high probability), a minimum of $\log(n)$ bits are needed on average to describe the approximated locations of each node, where n is the number of the nodes. Making use of this finding, we first proved that the expected routing overhead is $\Omega(n^{1.5} \ \log(n))$ under both proactive and reactive geographic routing. Then, we demonstrated that the minimum expected routing overhead scales as $\Theta(n \ \log(n))$.

Our study also revealed a main source of inefficiency of proactive and reactive geographic routing as well as the major contributors to routing overhead in MANETs.

It is clear from our findings that, when nodes' mobility is independent, the locations of destinations are mutually independent and introduce the linear term n in the scaling law (as pointed out in the introduction). When nodes' mobility is correlated, however, the expected routing overhead may grow slower; the exact manner in which it will grow is likely to depend on many factors, including the details of correlation structure imposed on nodes' mobility as well as the selection of source-destination pairs.

REFERENCES

[1] M. J. B. Appel and R. P. Russo, "The connectivity of a graph on uniform points on $[0,1]^d$," Statistics & Probability Letters, 60:351-357, 2002.
[2] S. Basagni, I. Chlamtac, and V. R. Syrotiuk, "A distance routing effect algorithm for mobility (DREAM)," in Proceedings of ACM MobiCom, pp. 76-84, Dallas (TX), 1998.
[3] N. Bisnik and A. A. Abouzeid, "Capacity deficit in mobile wireless ad hoc networks due to geographic routing overheads," in Proceedings of 2007 IEEE Infocom, pp. 517-525, May 2007.
[4] P. Bose, P. Morin, and I. Stojmenovic, "Routing with guaranteed delivery in ad hoc wireless networks," *Wireless Networks*, 7(6):609-616, Nov. 2001.
[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Wiley Interscience, John Wiley & Sons, Inc., Hoboken (NJ), 2006.
[6] O. Dousse and P. Thiran, "Connectivity vs. capacity in dense ad hoc networks," in Proceedings of 2004 IEEE Infocom, pp. 476-486, Mar. 2004.
[7] M. Franceschetti, O. Dousse, D. Tse, and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory," *IEEE Trans. on Information Theory*, 53(3):1009-1018, Mar. 2007.
[8] R. Gallager, "Basic limits on protocol information in data communication networks," *IEEE Trans. on Information Theory*, 22(4):385-398, Jul. 1976.
[9] G. Grimmett and D. Stirzaker, *Probability and Random Processes*, third ed., Oxford University Press, 2001.
[10] M. Grossglauser and D. Tse "Mobility Increases the Capacity of Ad Hoc Wireless Networks," *IEEE/ACM Transactions on Networking*, 10(4):477-486, Aug. 2002.
[11] P. Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks," *Stochastic Analysis, Control, and Optimization and Applications*, pp. 547-566, 1998.
[12] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. on Information Theory*, 46(2):388-404, Mar. 2000.
[13] G. Han, "Connectivity analysis of wireless ad-hoc networks," *Ph.D dissertation*, Department of Electrical and Computer Engineering, University of Maryland, College Park, 2007.
[14] G. Han and A. M. Makowski, "Very sharp transitions in one-dimensional MANETs," in Proceedings of 2006 IEEE International Conference on Communications, vol. 1, pp. 217-222, Jun. 2006.
[15] G. Han and A. M. Makowski, "A very strong zero-one law for connectivity in one-dimensional geometric random graphs," *IEEE Communications Letters*, 11(1):55-57, Jan. 2007.
[16] G. Han and A. M. Makowski, "On the critical communication range under node placement with vanishing densities," in Proceedings of 2007 IEEE International Symposium on Information Theory, pp. 831-835, Jun. 2007.
[17] R. Jain, A. Puri, and R. Sengupta, "Geographical routing using partial information for wireless ad hoc networks," *IEEE Personal Communications*, 8(1):48-57, Feb. 2001.
[18] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," *Mobile Computing*, pp. 153-181, 1996.
[19] A. Jovicic, P. Viswanath, and S. Kulkarni, "Upper bounds to transport capacity of wireless networks," *IEEE Trans. on Information Theory*, 50(11):2555-2565, Nov. 2004.
[20] F. Kuhn, R. Wattenhofer, and A. Zollinger, "Ad-hoc networks beyond unit disk graphs," *Wireless Networks*, 14(5):715-729, Jul. 2007.
[21] O. Leveque and E. Telatar, "Information-theoretic upper bounds on the capacity of large extended ad hoc wireless networks," *IEEE Trans. on Information Theory*, 51(3):858-865, Mar. 2005.
[22] J. Li, J. Jannotti, D. S. J. De Couto, D. R. Karger, and R. Morris, "A scalable location service for geographic ad hoc routing," in Proceedings of the 6th annual International Conference on Mobile Computing and Networking, pp. 120-130, Boston (MA), 2000.
[23] M. D. Penrose, *Random Geometric Graphs*, Oxford Studies in Probability, Oxford University Press, 2003.
[24] M. D. Penrose, "A strong law for the longest edge of the minimal spanning tree," *The Annals of Probability*, 27(1):246-260, 1999.
[25] C. Perkins and P. Bhagwat, "Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers," *Computer Communications Review*, 24(4):234-244, Oct. 1994.
[26] C. Perkins and E. M. Royer, "Ad hoc on-demand distance vector routing," in Proceedings of the second annual IEEE workshop on mobile computing systems and applications, pp. 90-100, Feb. 1999.
[27] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd Ed., Prentice Hall PTR, 2002.
[28] P. Santi, "The critical transmitting range for connectivity in mobile ad hoc networks," *IEEE Trans. on Mobile Computing*, 4(3):310-317, May/June 2005.
[29] C. Scheideler, A. W. Richa, and P. Santi, "An O(log n) dominating set protocol for wireless ad-hoc networks under the physical interference model," in Proceedings of the 9th ACM International Symposium on Mobile Ad-hoc Networking and Computing, pp. 91-100, 2008.
[30] G. Sharma, R. Mazumdar, and N. Shroff "Delay and capacity tradeoffs in mobile ad hoc networks: a global perspective," *IEEE Trans. on Networking*, 15(3):981-992, Oct. 2007
[31] J. M. Steele, "Growth rates of Euclidean minimal spanning trees with power weighted edges," *The Annals of Probability*, 16(4):1767-1787, 1988.
[32] I. Stojmenovic, "Position-based routing in ad hoc networks," *IEEE Communications Magazine*, 40(7):128-134, Jul. 2002.
[33] H. Takagi and L. Kleinrock, "Optimal transmission ranges for randomly distributed packet radio terminals," *IEEE Trans. on Communications*, 32(3):246-257, Mar. 1984
[34] L. Viennot, P. Jacquet, and T. H. Clausen, "Analyzing control traffic overhead versus mobility and data traffic activity in mobile ad-hoc network protocols," *Wireless Networks*, 10:447-455, 2004.
[35] N. Zhou and A. A. Abouzeid, "Information-theoretic bounds for mobile ad-hoc networks routing protocols," *Lecture Notes in Computer Science*, Vol. 2662/2003,pp. 651-661, Springer Berlin, Oct. 2003.
[36] N. Zhou and A. A. Abouzeid, "Routing in ad-hoc networks: a theoretical framework with practical implications," in Proceedings of 2005 IEEE Infocom, pp. 517-525, Mar. 2005.

PLACE PHOTO HERE

Richard J. La received his B.S.E.E. from the University of Maryland, College Park in 1994 and M.S. and Ph.D. degrees in Electrical Engineering from the University of California, Berkeley in 1997 and 2000, respectively. From 2000 to 2001 he was with the Mathematics of Communication Networks group at Motorola Inc,. Since 2001 he has been on the faculty of the Department of Electrical and Computer Engineering at the University of Maryland, where he is currently an Associate Professor.

PLACE PHOTO HERE

Eunyoung Seo received her B.S. and M.S. degrees in Electrical Engineering and Computer Science from Seoul National University, Seoul, Korea, in 2005 and 2007, respectively. She is currently working towards the Ph.D. degree at the University of Maryland, College Park. Her research interests include network overhead and communication theory.