Modeling Path Length in Wireless Ad-hoc Network

Quan Jun Chen , Salil S. Kanhere , Mahbub Hassan
School of Computer Science and Engineering,
The University of New South Wales, Sydney, Australia,
Email: {quanc, salilk, mahbub}@cse.unsw.edu.au

UNSW-CSE-TR-0617

August 2006

Abstract

Path length(i.e: the number of hops), the fundamental metric of multi-hop wireless network, has a determinative effect on the performance of wireless network, such as throughput, end-to-end delay and energy consumption. In this paper, we propose a stochastic process based mathematical model to analyze the shortest path length. The model is based on the observation that greedy forwarding in geographic routing can approximately find the shortest path in reasonably dense network. We present formula for the probability mass function of path length, given the distance between source and destination. In addition, we also propose a simple but efficient formula to estimate the mean path length. Our analytical results are well justified by a rich set of simulations, in which both random and realistic mobility scenarios have been investigated.

I. Introduction

The length of routing path refers to the number of transmission hops from source to destination. In multi-hop wireless network, the effect of path length on the performance has been explicitly discussed in many previous well-known works [1] [2] [3] [4] [5] [6]. Overall, path length has a determinative effect on the following performance metrics in wireless network:

- Energy Consumption: Energy saving is an important issue in some genres of wireless network, e.g. sensor network and personal network. With increasing path length, more intermediate nodes are involved in wireless transmitting, which inevitably consume more energy.
- Packet Collision: As nodes share a common wireless channel, packet collision is an inherent characteristic of wireless network. When routing path is longer, the collision chance of a packet increases.
- End-to-end Delay: Each extra wireless transmission of data packet introduces some delay, i.e. propagation delay, queuing delay and collision back off delay. Thus longer routing path have longer end-to-end delay consequently.
- Throughput: In the wireless network, such as Mobile Ad-hoc Network (MANET), where topology changes dynamically, longer routing path have higher path break probability. This results in higher packet loss.
- Routing Overhead: As discussed in [7], higher path break probability also leads to higher routing overhead, especially in reactive routing protocols, e.g. DSR [8], AODV [9].

Although path length issues have been studied in random graph theory, existing literature does not sufficiently cover wireless network scenario. This is mainly because of differences between random graphs and wireless network topologies. For example, the probability of edge occurrences in random graph are independent, while in case of wireless network it depends on relative distances between the nodes [10]. Several previous works [11] [12] have tried to analyze path length in wireless ad-hoc network. However, they either make some unrealistic assumptions, e.g. square radio range, or they only give a bound on path length. The sound analysis of path length in wireless ad-hoc network is still absent in the literature.

This paper targets at analyzing the shortest path length given the distance of communication pair. By leveraging the truth that greedy forwarding in geographic routing can almost find the shortest path in reasonably dense network [1] [13], we propose a stochastic process based mathematical model to analyze the shortest path length. The underlying principle of greedy forwarding involves selecting the next routing hop from amongst a node's neighbors, which is geographically closest to the destination. This localized

forwarding simplifies the original shortest path length problem without compromising the accuracy in reasonably dense network.

To justify this model, we conduct a comprehensive set of simulations covering scenarios with both random and realistic mobility. The comparison results prove our model works very well for the random mobility model, which can satisfy the assumptions of our model. Further, even for the realistic mobility model extracted from real-word vehicle networks, we find our analytical results are still valid.

The main contributions of this work have three folds.

- 1) We propose a stochastic process based model to analyze the probability distribution of the shortest path length, given Euclidean distance L between source and destination in reasonably dense network (i.e: The average number of neighbors is greater than 10).
- 2) We show that the average distance progress per hop converges when $L \gg R$ (R is radio range), and we build the formula for the convergency value λ, which only depends on average number of immediate neighbors.
- 3) Based on the results of 2), we propose an simple formula $\overline{H(L)} = \frac{L}{\lambda \cdot R}$ to estimate the mean shortest path length. We also show this estimation can improve the accuracy of widely used equation $\frac{L}{R}$ by up to 37%, depends on network density.

The rest of the paper is organized as follows. Section II provides an overview of our mathematical model and elaborates the assumptions in our analysis. Section III derives the probability mass function of path length and the mean value, given the distance between the source and destination. The complexity of this computation is considerably high. Hence, in Section IV we provide an efficient $O(1)$ estimation of the path length. Section V presents extensive simulations to validate our theoretical results. We also evaluate the relevance of our analytical results in realistic scenarios such as vehicular networks. Finally, Section VI concludes the paper.

II. Overview of Methodology

Given the complex nature of the analysis, we make the following simplifying assumptions in our analysis:

- Two Dimensional Terrain: We assume that the terrain under consideration is two dimensional. Our work can be easily extended to three dimensional topologies.
- Uniformly Distributed Node Layout: We assume that nodes are uniformly distributed in the target area. However, this does not entail using a particular mobility model for modeling the node move-

ment. Any mobility model that leads to a uniform distribution of the nodes is acceptable. Previous analytical exercises in ad hoc networks [6], [14], have made similar simplifying assumptions.

- Complete Knowledge of Local Topology: We assume that nodes always have an up-to-date view of their local topology, i.e. each node is aware of the locations of its immediate neighbors. The nodes can employ a neighbor discovery protocol for this purpose. Consequently, each intermediate node can always find the optimal next hop.
- No Boundary: In a typical ad hoc deployment, nodes located near the network boundary have fewer neighbors that nodes located elsewhere. To avoid this distinction, we ignore the existence of the boundary. Consequently, the probability distribution function for the number of neighbors at each node is identical [15].
- Dense Deployment: It is well-known that if the ad-hoc network is reasonably dense, then the forwarding path established by employing greedy forwarding between any source and destination node is the same as the shortest-hop path between these two nodes [1], [13]. In our analysis, we assume that the node density is sufficiently dense to allow us to use greedy forwarding for determining the shortest path between nodes. Since greedy forwarding relies on localized decisions, our analysis is greatly simplified as each node is not required to have knowledge of the entire topology. A consequence of this assumption is that the occurrence of a local minima, i.e. a situation where a node cannot find a suitable relay node that is closer to the destination than itself, is highly unlikely.
- Existence of Nodes in the Shadow Region: In greedy forwarding, the next hop chosen by a node is the one that is closest to the destination from amongst its one-hop neighbors. In other words, none of the other neighbors are closer to the destination than the chosen next hop node. For example, as shown in Fig. 1 the next hop node chosen by node X is node N. Consequently, there are no nodes in the shaded area, referred to as the *Shadow Region* in Fig. 1. However, our earlier assumption of uniformly distributed nodes requires that the neighbors of node N are equally likely to be placed anywhere within N's radio range including the shadow region. This implies that from the point of view of node N, it may be possible for nodes to exist within the shadow region. However, as discussed in Sec. III-C, the area of the shadow region is considerably smaller than the area within a node's radio. Further, as demonstrated in Sec. V, the impact of these contradictory assumptions on the analytical results of our model are negligible.

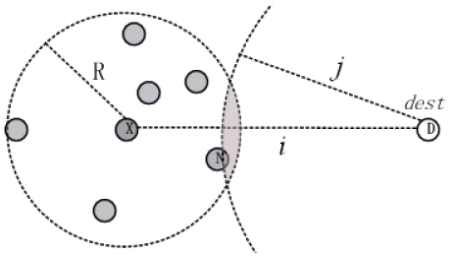

Fig. 1. Example of state transition (from state i to state j)

The above assumptions, some of which are somewhat unrealistic, are necessary in making the analysis tractable. However, in our simulation study, we relax several of these assumptions (e.g: uniform distribution of the nodes) to create more realistic scenarios and compare the resulting outcomes to those from our analysis.

Assuming that the distance between the source and destination is known, our analysis seeks to develop a model for analyzing the length of the shortest path between these nodes, where the path is measured in terms of the number of hops. We use a discrete Markov chain to model the hop-by-hop progress along the shortest path from the source to the destination, where the state is defined as the Euclidean distance (measured in some consistent metric unit) between the current forwarding node and the destination. Ideally, this distance should be modeled as a continuous random variable. However, to simplify our model, we use discrete state space to approximately represent the continuous distance values. We elaborate on our model using the example illustrated in Fig. 1, Assume that node X is the source and node D is the destination. Since node X is at a distance i from the destination, the initial state for this path is i. The next hop node chosen by node X using greedy forwarding is node N, which is at a distance of j from the destination. This results in a state transition from i to j. To restrict the number of possible states, we quantize the distances resulting in a state space of $(0, \varepsilon, 2\varepsilon, ..., n\varepsilon, ...)$, where the parameter ε is the interval of the state space (i.e. the quantization coefficient). The smaller the interval ε, the closer our model approximates reality. However, a small interval increases the computation time required for the evaluation. We discuss this tradeoff in our simulation study in Sec. V. Table II presents the notations used in our analysis

III. Analysis of the probability distribution of path length and the Mean value

Our analysis is composed of the following steps. The first step involves determining the state transition probabilities for the Markov chain used to model the path. Based on the transition probabilities, we

TABLE I

NOTATIONS USED IN OUR ANALYSIS

Symbol	Denotation
N	The total number of nodes in network
A	The area of network
R	The radio range
ε	The interval of state space
$\overline{m}$	The average number of one hop neighbors
N_{neighb}	The number of one hop neighbors
L	The distance between the source and destination
H	The shortest path length (measured as number of hops
$\overline{H(L)}$	The mean shortest path length given the distance L
$\overline{G(L)}$	The average progress per hop given the distance L
λ	The value that the average progress per hop converges to

compute the probability distribution of the path length. Finally, the probability distribution function is used to derive the mean value of the path length (measured in hops).

A. Evaluating the State Transition Probability

In the defined state space, state i denotes that the distance from current intermediate node to destination is i units (meters). The transition from state i to j, implies that the next hop neighbor is at a distance of j from the destination, as illustrated in Fig. 1. Intuitively, the transition probability depends on the node density, since the more dense the network, the greater the chance that the the next hop will make significant progress towards the destination. We first compute the probability distribution of the number of neighbors. Then we derive the conditional transition probability, given the number of neighbors. Finally,by applying the law of total probability we can get the final transition probability.

Theorem 1: The probability of a node having m neighbors is

$$P(N_{neigh} = m) = \quad \binom{N-1}{m}(\frac{\pi R^2}{A})^m(1-\frac{\pi R^2}{A})^{N-1-m} \\ , 0 \leq m \leq N-1 \tag{1}$$

and the average number of neighbors per node is

$$\overline{m} = E(N_{neigh}) = (N-1)\frac{\pi R^2}{A} \tag{2}$$

Proof: This can be formulated as a binomial probability distribution. Since the nodes are uniformly distributed in the network, given any node X, the probability that any other node is within the radio range of node X is given by,

$$p = \frac{\text{The area of radio coverage}}{\text{The area of network}} = \frac{\pi R^2}{A} \tag{3}$$

Since m nodes are within the radio coverage of X, the rest of the nodes, i.e., $(N-1-m)$, are outside it's radio range. Using the binomial probability distribution, we have,

$$P(N_{neigh} = m) = \binom{N-1}{m} p^m (1-p)^{N-1-m} \tag{4}$$

Combining Eq. (3) and Eq. (4), one can readily derive Eq. (1). Further, the expected value of a binomially distributed variable is given by,

$$\overline{m} = E(N_{neigh}) = (N-1)p = (N-1)\frac{\pi R^2}{A} \tag{5}$$

■

Theorem 2: Assume that the system is currently in state i, with node X being the current relay node, with a radio range of R, as depicted in Fig. 1. Given that node X has m neighbors, the conditional transition probability from state i to state j is given by,

$$P_{i,j|m>0} = P(s_i \to s_j | N_{neigh} = m > 0) = \begin{cases} 1 \quad \text{if } i \le R \text{ and } j = 0, \\ 0 \quad \text{if } i \le R \text{ and } j > 0, \\ \sum_{k=1}^{m} \binom{m}{k} (\frac{A_{i,j+\varepsilon} - A_{i,j}}{\pi R^2})^k (1 - \frac{A_{i,j+\varepsilon}}{\pi R^2})^{m-k} \\ \qquad \text{if } i > R \text{ and } i - R \le j < i, \\ \sum_{k=1}^{m} \binom{m}{k} (\frac{A_{i,j} - A_{i,j-\varepsilon}}{\pi R^2})^k (1 - \frac{A_{i,j}}{\pi R^2})^{m-k} \\ \qquad \text{if } i > R \text{ and } i < j \le i + R, \\ 0 \quad \text{if } i > R \text{ and } (i = j \text{ or } j < i - R \text{ or } j > i + R), \end{cases} \tag{6}$$

$$P_{i,j|m=0} = P(s_i \to s_j | N_{neigh} = m = 0) = \begin{cases} 1 \quad \text{if i=j,} \\ 0 \quad \text{otherwise,} \end{cases}$$

$$where,$$

$$A_{i,j} = R^2 \arccos \frac{i^2+R^2-j^2}{2iR} + j^2 \arccos \frac{i^2+j^2-R^2}{2ij} - \frac{\sqrt{(R+i+j)(R+i-j)(R-i+j)(i+j-R)}}{2} \tag{7}$$

Proof: To prove this theorem, we start with a simple case, where node X has no neighbors, i.e. $m = 0$. Consequently, there is no next hop node implying no state transition. Hence,

$$P_{i,j|m=0} = P(s_i \rightarrow s_j | N_{neigh} = m) = \begin{cases} 1 & \text{if } i = j, \\ 0 & otherwise. \end{cases} \tag{8}$$

When the number of neighbors is greater than zero, we sub-divide the problem into two parts, depending on the relationship between the distance i and the radio range R. When $i \leq R$, the destination node is within the one-hop neighborhood of the current node X. Hence, as the next hop is the destination, the state i must transition to state 0. Consequently, we have,

$$\begin{aligned} P_{i,j|m>0} &= P(s_i \rightarrow s_j | N_{neigh} = m) \\ &= \begin{cases} 1 & \text{if } i \leq R \text{ and } j = 0, \\ 0 & \text{if } i \leq R \text{ and } j > 0. \end{cases} \end{aligned} \tag{9}$$

Now let us consider the situation where $i > R$. Since the next hop node must be within the radio range of node X, the probability that state i will transition to a state j, which lies between $[i-R, i+R]$, is zero. In addition, the probability that the state does not change is zero as well, since node X has m neighbors. Thus, we have,

$$P_{i,j|m>0} = P(s_i \rightarrow s_j | N_{neigh} = m) = 0 \quad \text{,if } i > R \text{ and } (i = j \text{ or } j < i - R \text{ or } j > i + R) \tag{10}$$

Now, we discuss the more complicated and plausible case where, $i + R \leq j \leq i - R$. This can further be broken down into the following two sub-cases: (i) $i - R \leq j < i$ and (ii) $i < j \leq i + R$. For the sake of brevity, we will only discuss the former case. The latter can be proved in a similar manner.

The probability of state transition from i to j is the probability that at least one neighbor of node X lies on the perimeter of the curve of radius j centered at the destination (see Fig. 1) with the rest of X's neighbors located in the region to the left of this curve. Since we assume a discrete state space, with ε as the interval of the state space, we can approximate the curve as a ring of thickness ε, as illustrated in

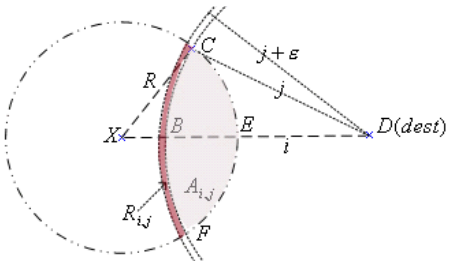

Fig. 2. Illustration used to prove Theorem 2

Fig. 2. Let $R_{i,j}$ represent the area of this curve that intersects with the radio range of node X (narrow dark region in Fig. 2). Let $A_{i,j}$ represent the area of the shaded region in Fig. 2, which is the intersecting region between the radio coverage of node X and a circle of radius j centered at the destination D. The area of the ring $R_{i,j}$ can be computed as follows:

$$R_{i,j} = A_{i,j+\varepsilon} - A_{i,j} \tag{11}$$

As shown in Fig. 2, the area $A_{i,j}$ can be computed as,

$$A_{i,j} = 2(A_{\widehat{CXE}} - A_{CXD} + A_{\widehat{CDB}}) \tag{12}$$

, where $A_{\widehat{CXE}}$ is the area of the sector CXE; A_{CXD} is the area of triangle CXD and $A_{\widehat{CDB}}$ is the area of sector CDB. By applying the law of cosines and Heron's formula, we have,

$$\begin{cases} A_{\widehat{CXE}} = \frac{R^2}{2}\angle CXD = \frac{R^2}{2}\arccos\frac{i^2+R^2-j^2}{2iR} \\ A_{\widehat{CDB}} = \frac{R^2}{2}\angle CDX = \frac{j^2}{2}\arccos\frac{i^2+j^2-R^2}{2iR} \\ A_{CXD} = \frac{\sqrt{(R+i+j)(R+i-j)(R-i+j)(i+j-R)}}{4} \end{cases} \tag{13}$$

Combining Eq. (12) and Eq. (13), we have,

$$\begin{aligned} A_{i,j} = {} & R^2\arccos\frac{i^2+R^2-j^2}{2iR} + j^2\arccos\frac{i^2+j^2-R^2}{2ij} \\ & - \frac{\sqrt{(R+i+j)(R+i-j)(R-i+j)(i+j-R)}}{2} \end{aligned} \tag{14}$$

Let $P_{j|i}$ denote the conditional probability that a neighbor is at a distance j from the destination, given that the current state is i. Similarly, let $P_{(farther\ than\ j)|i}$ denote the conditional probability that a neighbor is at a distance greater than j from the destination given that the current state is i. Since the nodes are

uniformly distributed, these conditional probabilities can be computed as follows,

$$\begin{aligned} P_{j|i} &\approx P(\text{a neighbor is present in the ring } R_{i,j}) \\ &= \frac{\text{the area of ring } R_{i,j}}{\text{the area of radio coverage}} = \frac{R_{i,j}}{\pi R^2} \\ &= \frac{A_{i,j+\varepsilon} - A_{i,j}}{\pi R^2} \end{aligned} \tag{15}$$

$$\begin{aligned} &P_{(farther\ than\ j)|i} \\ &\approx P(\text{a neighbor is in the region to the left of ring } R_{i,j}) \\ &= 1 - \frac{A_{i,j+\varepsilon}}{\pi R^2} \end{aligned} \tag{16}$$

Now, the transition probability from state i to j is the probability that at least one neighbor of node X lies inside ring $R_{i,j}$, while the rest of its neighbors are to the left of ring $R_{i,j}$. Using the binomial probability theorem and using the conditional probabilities from Eq. (15) and Eq. (16), we finally have,

$$\begin{aligned} P_{i,j|m} &= P(s_i \rightarrow s_j | N_{neigh} = m) \\ &= \sum_{k=1}^{m} \binom{m}{k} (P_{j|i})^k \ (P_{\text{(farther than j)}|j})^{m-k} \\ &= \sum_{k=1}^{m} \binom{m}{k} (\frac{A_{i,j+\varepsilon} - A_{i,j}}{\pi R^2})^k (1 - \frac{A_{i,j+\varepsilon}}{\pi R^2})^{m-k} \end{aligned} \tag{17}$$

One can use a similar approach to compute the probability for the case when, $i < j \leq i + R$. We have omitted this due to lack of space and present the final result below,

$$\begin{aligned} P_{i,j|m} &= P(s_i \rightarrow s_j | N_{neigh} = m) \\ &= \sum_{k=1}^{m} \binom{m}{k} (\frac{A_{i,j} - A_{i,j-\varepsilon}}{\pi R^2})^k (1 - \frac{A_{i,j}}{\pi R^2})^{m-k} \end{aligned} \tag{18}$$

By combining Eqs. (8), (9), (10), (14), (17) and (18), we finally prove Theorem 2. ■

We can now readily compute the state transition probability as follows,

Theorem 3: The transition probability from state i to j is

$$P_{i,j} = \sum_{m=0}^{N-1} P(s_i \rightarrow s_j | N_{neigh} = m) P(N_{neigh} = m) \tag{19}$$

Proof: We have already derived $P(s_i \rightarrow s_j | N_{neigh} = m)$ and $P(N_{neigh} = m)$ in Theorems 1 and 2, respectively. The above theorem follows by applying the law of total probability. ■

B. Probability Distribution Function of the Path Length

Recall that, the state variable in our Markov model represents the distance between the current node and the destination. Based on the transition probability computed in Section III-A and using the approach of recursive computation, we obtain the probability distribution function of the path length as follows,

Theorem 4: Let H denote the number of hops. The probability distribution of the path length, given that the path originates at state i is given by,

$$P(H=h|s=i) = \begin{cases} 1 \quad \text{if } 0<i\leq R \text{ and } h=1, \\ 0 \quad \text{if } 0<i\leq R \text{ and } h>1, \\ 0 \quad \text{if } i>R \text{ and } h<\lceil \frac{i}{R} \rceil, \\ \frac{\sum_{j=i-R|\varepsilon}^{i-\varepsilon} P(H=h-1|s=j)P_{i,j}}{1-P_{i,i}} + \\ \quad \frac{\sum_{j=i+\varepsilon|\varepsilon}^{i+R} P(H=h-1|s=j)P_{i,j}}{1-P_{i,i}} \quad \text{otherwise,} \end{cases} \tag{20}$$

Proof: Assume that the node which corresponds to state i is node X as in Fig. 1. Now if $i \leq R$, then the destination is just one-hop away from node X and hence the path length is one. Thus,

$$P(H=h|s=i) = \begin{cases} 1 \quad \text{if } 0<i\leq R \text{ and } h=1, \\ 0 \quad \text{if } 0<i\leq R \text{ and } h>1. \end{cases} \tag{21}$$

On the contrary, when $i > R$, the minimum number of hops must be $\lceil \frac{i}{R} \rceil$, since each hop can at most progress by a distance R towards the destination. Thus, the probability that the path length is less than $\lceil \frac{i}{R} \rceil$ is zero, i.e.,

$$P(H=h|s=i) = 0 \quad \text{if } i>R \text{ and } h<\lceil \tfrac{i}{R} \rceil, \tag{22}$$

For the other cases, applying the law of total probability, we have,

$$P(H=h|s=i) = \sum_{j=i-R|\varepsilon}^{i+R|\varepsilon} P(H=h-1|s=j)P_{i,j} \tag{23}$$

This is illustrated in Fig. 3. Recall that each subsequent step in the state space is separated by ε. Further, as discussed in Theorem 2, the possible next hop states originating at i are constrained between $i+R$

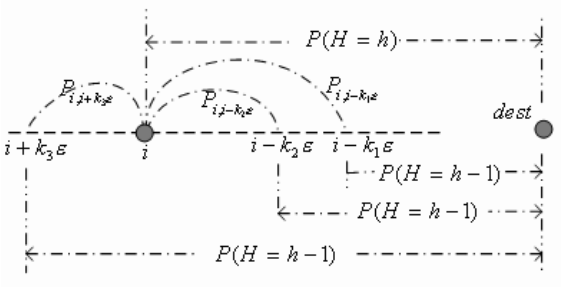

Fig. 3. Computation of the probability distribution of the path length

and $i - R$. Now, Eq.(23) can be expanded as follows,

$$P(H = h|s = i) = \sum_{j=i-R|\varepsilon}^{i-\varepsilon} P(H = h-1|s = j)P_{i,j} + \\ P(H = h|s = i)P_{i,i} + \sum_{j=i+\varepsilon|\varepsilon}^{i+R} P(H = h-1|s = j)P_{i,j} \tag{24}$$

Hence,

$$P(H = h|s = i) = \frac{\sum_{j=i-R|\varepsilon}^{i-\varepsilon} P(H=h-1|s=j)P_{i,j}}{1-P_{i,i}} + \\ \frac{\sum_{j=i+\varepsilon|\varepsilon}^{i+R} P(H=h-1|s=j)P_{i,j}}{1-P_{i,i}} \tag{25}$$

Combining Eqs. (21), (22) and (25), the theorem is proved. ■

Using Eq. (21) and (22), one can readily determine the probability of single hop paths. Subsequently, using Eq. (25) and the probability of single hop paths, the probability of two hops paths can be computed. Similarly, employing recursive computations, the probability of all h hop paths can be computed.

C. Mean Path Length

Based on the result of Theorem 4, we can easily calculate the mean path length (measured in number of hops) for a pair of communication nodes, which are separated by distance L. Let $\overline{H(L)}$ represent the mean path length given L, which can be computed as follows,

$$\overline{H(L)} = \sum_{h=1}^{\infty} h \cdot P(H = h|s = L) \tag{26}$$

Recall that, $P(H = h|s = L)$ can be calculated using Theorem 4. now illustrate the mean path length for Cons $N = 117, A = 1500 * 1500m, R = 250m, \varepsilon = 5$. Fig. 4 plots the mean path length as a function of the distance L between the source and destination for a network of size $1500 * 1500m$ consisting of 117 uniformly distributed nodes with the radio range set to 250m. One can readily observe from Fig. 4 that

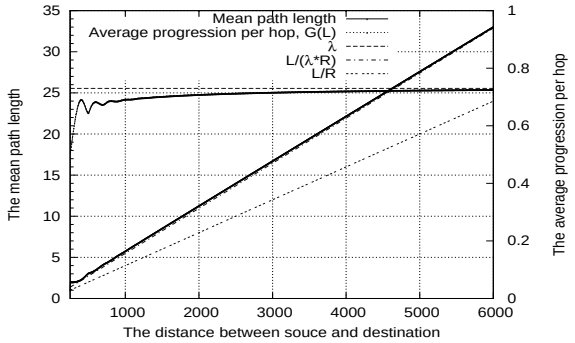

Fig. 4. Mean path length (actual & estimated) and average progression v/s distance

the mean path is a linear function of the distance between the source and the destination. This observation implies that the ratio of the source-destination distance and the mean length is constant. We will explore this feature in the subsequent section.

Note that, the above computation of the mean path length requires us to recursively compute the entire probability distribution function of the path length. We have proved that this computation involves a complexity of $O(N^{L/R})$. The details of this proof have been excluded due to space restrictions. It is evident that evaluating the mean path length for a sizable network can be an considerably computationally intensive task. Hence, in the next part of our analysis, we evaluate an $O(1)$ technique for estimating the mean path length. We also demonstrate that our estimate is quite accurate, especially when $L >> R$.

IV. Approximate Analysis of the Mean Path Length

As discussed in the previous section, the exact computation of the mean path length is highly complex. To simplify this analysis, in this section, we develop a novel technique to estimate the mean path length. In order to achieve this, we first introduce a new parameter, known as the average progression per hop. Given the distance between the source and destination is L and the radio range of each node is R, the average progression per hop, $G(L)$, is defined as follows,

$$G(L) = \frac{L}{R \cdot \overline{H(L)}}. \tag{27}$$

$G(L)$ is a measure of how far each hop can progress towards to destination, normalized by the radio range. The results from Fig. 4 indicate that the ratio of the destination between the source and destination and the mean path length of the hop is constant. $G(L)$ will provide a means to investigate this observation. Fig. 4 plots $G(L)$ as a function of L for the same example as that used for the mean path length in Sec. III-C. One can observe that $G(L)$ converges at a certain value when $L \gg R$. This implies that when

Fig. 5. As $L >> R$, the arc $\widehat{CF}$ can be approximated as line CF

$L >> R$, the average progress made along each hop approaches a constant value. The following theorem proves this observation. It is well-known that in ad-hoc networks, the throughput decreases substantially with each additional hop. Hence, most real deployments may not necessarily have long multi-hop paths and thus, $L >> R$ may not always hold true. However, to evaluate the convergence of $G(L)$ in our analysis, this assumption is necessary.

Theorem 5: The average progression per hop converges as $L >> R$ and the value λ that it converges to is given by,

$$\lambda = 1 - \int_{-1}^{1} (1 - \frac{\arccos(t) - t\sqrt{(1-t^2)}}{\pi})^{\overline{m}} dt \tag{28}$$

,where $\overline{m}$ is the average number of neighbors.

Proof: From Theorem 4, we know that the probability distribution of the path length given the initial state i depends on the state transition probability. Further from Theorem 2, we know that the transition probability is dependent on the area of $A_{i,j}$. Fig. 2 illustrates that $A_{i,j}$ is determined by the shape of curve $\widehat{CF}$. This area is computed using Eq. (14) and depends on both i and j. If the distance L between the source and destination is very large, as depicted in Fig. 5, the curve $\widehat{CF}$ can be approximated by a straight line. Consequently, this simplifies the computation of the area $A_{i,j}$, which now solely depends on the distance between node X and its next one-hop neighbor, $i - j$. Let x represent $i - j$. Now, we can calculate the area of $A_{i,j}$, as depicted in Fig. 6, as follows,

$$A_{i,j} = 2(A_{\widehat{CXE}} - A_{CXD}) = R^2 \arccos\frac{x}{R} - x\sqrt{(R^2 - x^2)} \tag{29}$$

The probability that a neighbor of X lies in the region (within the radio range R) to the left of CF is given by,

$$P(\text{a neighbor of } X \text{ lies to the left of } CF) = 1 - \frac{A_{i,j}}{\pi R^2} \tag{30}$$

In order that the progress made along the next hop from node X towards the destination (i.e. change of state from i to j) is less than x, all neighbors of X must reside in the region to the left of CF. Let T represent the progress made along this hop towards the destination. Combining Eq. (29) and (30), the

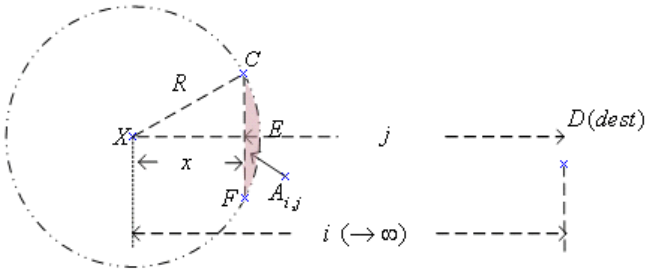

Fig. 6. As i tends to infinity, $A_{i,j}$ is approximated by the shaded region

probability that the progression of the next hop towards to destination is less than x is,

$$\begin{aligned} F_T(x) &= P(T < x) \\ &= P(\text{all neighbors are in the left side of } CF) \\ &= (1 - \tfrac{A_{i,j}}{\pi R^2})^{\overline{m}} \\ &= (1 - \tfrac{R^2 \arccos \frac{x}{R} - x\sqrt{(R^2 - x^2)}}{\pi R^2})^{\overline{m}} \end{aligned} \tag{31}$$

Consequently, the probability density function (pdf) of the progression T is given by,

$$f_T(x) = \frac{d(F_T(x))}{x} \tag{32}$$

Further, the average of the progression T is,

$$E(T) = \int_{-R}^{R} x f_T(x) dx \tag{33}$$

Recall that λ denotes the value that $G(L)$ converges to. Thus, we have,

$$\begin{aligned} \lambda = \quad & \tfrac{E(T)}{R} = \tfrac{1}{R} \textstyle\int_{-R}^{R} x f_T(x) dx \\ = \quad & \textstyle\int_{-R}^{R} \frac{x}{R} dF_T(x) \\ = \quad & [\tfrac{x}{R} F_T(x)]_{-R}^{R} - \textstyle\int_{-R}^{R} \frac{1}{R} F_T(x) dx \\ = \quad & 1 - \textstyle\int_{-R}^{R} \frac{1}{R} (1 - \frac{R^2 \arccos \frac{x}{R} - x\sqrt{(R^2 - x^2)}}{\pi R^2})^{\overline{m}} dx \\ \overset{x=Rt}{\Longrightarrow} \quad & 1 - \textstyle\int_{-1}^{1} (1 - \frac{\arccos(t) - t\sqrt{(1-t^2)}}{\pi})^{\overline{m}} dt \end{aligned} \tag{34}$$

■

Interestingly, the above equation reveals that λ only depends on the average number of neighbors (node density) and is independent of the radio range. Fig. 4 compares the value of λ as derived in the theorem above with that of $G(L)$ for different values of the distance between the source and destination. In this particular example, $\overline{m} = 10$, which results in $\lambda = 0.7298$. As is evident, from Fig. 4, Theorem 5 provides an accurate approximation of the convergene of $G(L)$, especially when $L \gg R$.

The above theorem and Eq. (27) leads to the following interesting corollary,

Corollary 1: The mean path length, $\overline{H(L)}$, can be approximated as follows,

$$\overline{H(L)} = \frac{L}{\lambda \cdot R} \tag{35}$$

Eq. (35) has an implementation complexity of $O(1)$ and hence can simplify the computation of the mean path length significantly as compared with the highly complex Eq. (26). Fig. 4 illustrates how the estimated mean path length closely approximates the actual mean path length for a network with the same parameters as the one in Section III-C. As observed the discrepancy between the approximation and the exact value is indistinguishable when $L >> R$.This approximation eliminates the need to recursively compute the probability distribution function using Theorem 4 and significantly reduces the complexity involved in determining the mean path length, without compromising the accuracy.

In the past, researchers who have theoretically analyzed the capacity and other aspects of ad-hoc networks [1], [2], have frequently used L/R as the lower bound for the mean path length. Comparing the estimated mean path length in Eq. (35) with L/R, we observe an improvement in accuracy as follows,

$$\frac{\frac{L}{\lambda \cdot R} - \frac{L}{R}}{\frac{L}{R}} = \frac{1}{\lambda} - 1 \tag{36}$$

Fig 4 compares L/R with our analytical results. The graph illustrates that the estimated mean path length is a superior approximation of the mean path length. For example, the improvement is 37% when the average number of neighbors are 10.

Lastly, we discuss the consequence of one of our assumptions, "Existence of Nodes in the Shadow Region" (see Sec. II). Recall that, in our analysis we have assumed the existence of nodes in the shadow region depicted in Fig 1, even though this contradicts with the greedy forwarding strategy. The *shadow region* in our assumption corresponds to the region $A_{i,j}$ in Fig. 6. The inconsistency introduced by these contradicting assumptions depends on the ratio of $A_{i,j}$ to the area within a node's radio coverage, which can be computed as follows,

$$\begin{aligned} \frac{A_{i,j}}{\pi R^2} &= \frac{R^2 \arccos\frac{x}{R} - x\sqrt{(R^2 - x^2)}}{\pi R^2} \\ &= \frac{\arccos(\lambda) - \lambda\sqrt{(1-\lambda^2)}}{\pi} \end{aligned} \tag{37}$$

Assuming a density of $\overline{m} = 10$, this ratio is a mere 8 %. This further reduces to 5% when $\overline{m} = 15$. In addition, the simulation results in Section V demonstrate that the impact of this shadow region on our analytical results is virtually non-existent.

V. Simulations

In this section, we present comprehensive simulations to validate our theoretical analysis. Our first goal is to corroborate the claims of Section III-C, which proposes an exact solution for computing the mean path length. Subsequently, we also wish to verify the accuracy of the estimated mean path length (Corollary 1). It is important to note that our analysis is independent of the node mobility and only requires the node distribution in the network to be uniform. In the first part of our simulations, where we validate our analysis, we use a mobility model called *Random Direction Mobility with Reflection*, which conforms to this uniform distribution assumption. Finally, we carry out simulations using mobility patterns that do not conform to the assumptions in our analysis, such as the popular Random Way Point and realistic movement traces of a vehicular network. The objective of this exercise is to compare our analytical results with those from more realistic scenarios and more importantly to ascertain if the analysis can serve as bounds in these situations.

A. Random Direction Mobility Model with Reflection

The probability mass function (pmf) of the path length given the distance between the source and destination nodes is computed in Theorem 4. This pmf is subsequently used to determine the mean path length as shown in Eq. (26). In this sub-section we compare the analytically derived pmf and mean path length with those from simulation experiments. We assume a network where the nodes are uniformly distributed in a square region of size $3000m * 3000m$, with the radio range of each node set to 250m. The average number of neighbors, $\overline{m}$ are equal to 15. The corresponding total number of nodes as derived from Eq. (2) are 688. We use the *Random Direction Model with Reflection(RDMR)* mobility model [16] in our simulation since it has been proven in [16] that this model results in a uniform distribution of nodes within the deployment area. We assume that the speed of each node varies uniformly from 0 to 20 m/s (average speed of 10 m/s), and the travel duration at each transition varies uniformly from 5 to 15 seconds. Recall that our analysis ignores the network boundary. Hence, in our comparative study, we only consider nodes within the central square region of size $2500m * 2500m$. In our simulations we use a state space interval, ε of 5m (i.e., the distances are quantized to discrete values, which are multiples of 5). We have evaluated the effect of choosing different values of ε, and have found that a value of 5

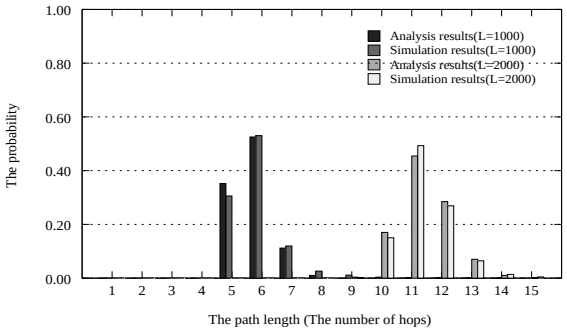

Fig. 7. Probability mass function of the path length, RMDR mobility model

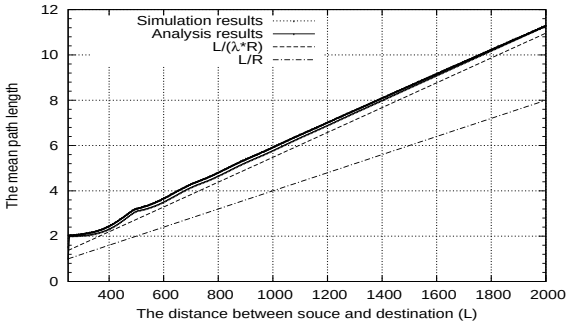

Fig. 8. Mean path length v/s L, RDMR mobility model

achieves a suitable balance between the accuracy and the computation duration. Due to space restrictions, we do not present those results here.

A simulation run lasts for 3000 seconds during which we take a snapshot of the network every five seconds. For each snapshot, we compute the shortest path that exists between all possible source destination pairs and the number of hops along the path. This leads to a very large sample space, since we have 600 snapshots in total, leading to a comprehensive set of 10^7 source-destination pairs. Note that, in our simulations we assume that each node is aware of the complete global topology, which enables the node to compute accurate shortest paths to all destinations. The resulting probability distribution and mean path length are compared with their analytical derived counterparts: pmf (Theorem 4), the exact and estimated mean path length (Eq. (26) and Corollary 1, respectively). We also include the popular approximation for the mean path length, L/R in our comparisons.

Fig. 7 illustrates the probability mass function of the path length, with $\overline{m} = 15$, for 2 values of L: 1000 and 2000 meters. Fig. 8 compares the mean path length obtained from our analysis results with the corresponding simulation results. From Figs. 8 and 7 it is evident that the simulation results are in

line with our theoretical results, thus validating our analysis exercise. Further, Fig. 8 indicates that our estimated mean path length provides a better approximation of the mean path length as opposed to using L/R.

B. Realistic Mobility Models

In the previous sub-section, all simulations parameters conformed to the assumptions used in our analysis. However, not all these assumptions will hold true for realistic mobile ad hoc networks. In particular, a real-world network would not usually consist of uniformly distributed nodes. In this sub-section, we wish to investigate if our theoretical results are relevant in practical scenarios. For this we first use the popular random way point mobility model [17] to model the node movements. In the second instance, we investigate a real-world vehicular ad hoc network, a popular application domain for MANETs. Our aim is to determine if the mean path length and its estimated value as derived in our analysis are pertinent for these real-world networks.

For the first case we choose a scenario that uses the random way point mobility to model the motion of the nodes. The rest of the simulation parameters (network size, number of nodes, etc) are identical to those used in Section V-A and so is the methodology (snapshots every 5 seconds, etc). Note that, the random way point model does not result in a uniform distribution of the nodes. Fig. 9 illustrates that our analytical and estimated results slightly over-estimate the mean path length as compared to the simulations. The reason for this is that in the random way point model, nodes tend to move towards the central area of the network, with the consequence that the central area is much denser as compared to the regions near the border [17]. Hence, on average, the progress made per hop towards the destination is larger as compared to a purely uniform distribution (as in our analysis), resulting in a shorter path length. However, Fig. 9 clearly demonstrates that our results are far more accurate than the frequently used measure of the mean path length, L/R.

The mobility model used in the second instance of our simulations is based on the actual movement of buses in the King County Metro bus system in Seattle, Washington [18]. We extract an area of size $4000m * 7000m$ corresponding to the the downtown area of Seattle. The duration of this trace spans 30 minutes. We assume that the radio range of each node is 1000m, which is consistent with that for DSRC[19] and the results from [20]. During this simulation run, we found that the average number of nodes in the selected region was 288. Using the above parameters and Eq. (2), we can calculate the average number of neighbors per node to be 32. We used this parameter in our analysis. Fig. 10 compares the analytical and estimated (using Corollary 1) mean path length values with those from the simulations

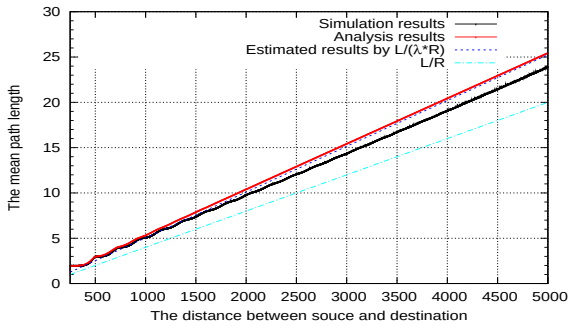

Fig. 9. Mean path length for the random way point model

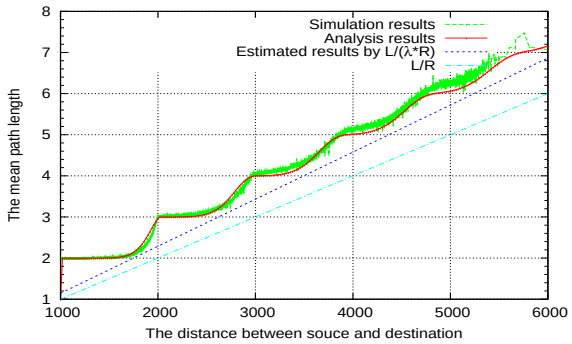

Fig. 10. Mean path length for the vehicular network

for this realistic vehicular network. It is evident that the analysis results are well matched with those from the simulations. As observed with other mobility models, the analytically derived estimated mean path length is a significantly closer approximation as compared with L/R.

In conclusion, it is obvious from these simulations, that our theoretical results are highly relevant to practical cases and can be used for accurately estimating the mean path length in a realistic deployment.

VI. Conclusion

Motivated by the importance of path length(i.e: the number of hops) in multi-hop wireless network, this paper proposes a stochastic process based model to analyze the length of shortest path in virtue of simplicity introduced by greedy forwarding. We present formula for the probability mass function of path length given distance L between source and destination. In order to reduce the computation complexity, we also propose a simple but efficient formula to approximate the mean path length. The rich set of simulation results, generated from different mobility models, validate our analysis model. We show our results can improve the accuracy of the widely used equation L/R by up to 37%. In the future, we are going to apply the analysis results to the performance analysis of other metrics, such as capacity,

end-to-end delay, and routing overhead. In addition, we will investigate how the knowledge of path length can assist routing protocol or localization so as to achieve better performance.

REFERENCES

[1] J. Li, C. Blake, D.S.J. De Couto, H.I. Lee, and R. Morris. Capacity of ad hoc wireless networks. In *Proceedings of ACM MOBICOM*, Rome, Italy, July 2001.

[2] P. Gupta and P.R. Kumar. The capacity of wireless networks. *IEEE Transactions on Information Theory*, 46(2):388–404, March 2000.

[3] G.K. Holland and N.K. Vaidya. Analysis of tcp performance over mobile ad hoc networks. *Wireless Networks*, 8(2-3):275 – 288, March 2002.

[4] J.H. Chang and L. Tassiulas. Energy conserving routing in wireless ad-hoc networks. In *Proceedings of IEEE INFOCOM*, Tel Aviv, Israel, 2000.

[5] R. Draves, J. Padhye, and B. Zill. Routing in multi-radio, multi-hop wireless mesh networks. In *Proceedings of ACM MOBICOM*, Philadelphia, PA, USA, September 2004.

[6] M. Grossglauser and DNC Tse. Mobility increases the capacity of ad hoc wireless networks. *IEEE/ACM Transactions on Networking*, 10(4):477–486, August 2002.

[7] F. Bai, N. Sadagopan, B. Krishnamachari, and A. Helmy. Modeling path duration distributions in manets and their impact on reactive routing protocols. *IEEE Journal on Selected Areas in Communications*, 22(7):1357– 1373, September 2004.

[8] David B. Johnson, David A. Maltz, and Josh Broch. DSR: the dynamic source routing protocol for multihop wireless ad hoc networks. pages 139–172, 2001.

[9] C. E. Perkins and E. M. Royer. Ad-hoc on-demand distance vector routing (AODV). *WMCSA*, 00:90–100, 1999.

[10] P. Santi and D.M. Blough. The critical transmitting range for connectivity in sparse wireless ad hoc networks. *IEEE Transactions on Mobile Computing*, 2(1):25–39, Jan-March 2003.

[11] D. Lebedev and J. M. Steyaert. Path lengths in ad hoc networks. pages 207–211. 2004 international workshop on wireless ad hoc network, May 2004.

[12] S. De, A. Caruso, T. Chaira, and S. Chessa. Bounds on hop distance in greedy routing approach in wireless ad hoc networks. *International Journal on Wireless and Mobile Computing (accepted)*.

[13] B. Karp and H. T. Kung. GPSR: Greedy perimeter stateless routing forwireless networks. In *Proceedings of 6th Annual International Conference on Mo-bile Computing and Networking (MobiCom 2000)*, pages 243–254, Boston, MA, USA, 2000.

[14] I. D. Aron and S. Gupta. Analytical comparison of local and end-to-end error recovery in reactive routing protocols for mobile ad hoc networks. In *Proceedings of ACM Workshop Modeling, Analysis, and Simulation of Wireless and Mobile Systems (MSWiM)*, 2000.

[15] C. Bettstetter and O. Krause. On border effects in modeling and simulation of wireless ad hoc networks. In *Proceedings of IEEE MWCN*, Recife, Brazil, 2001.

[16] P. Nain, D. Towsley, B. Liu, and Z. Liu. Properties of random direction models. In *Proceedings of IEEE Infocom 2005*, Miami, FL, March 2005.

[17] C. Bettstetter, H. Hartenstein, and X. Prez-Costa. Stochastic properties of the random waypoint mobility model. *Wireless Networks*, 10(5):555–567, September 2004.

[18] J. G. Jetcheva, Y. C. Hu, S. Palchaudhuri, A. K. Saha, and D. B. Johnson. Design and evaluation of a metropolitan area multitier wireless ad hoc network architecture. In *in Proceedings of Fifth IEEE Workshop on Mobile Computing Systems and Applications*, 2003.

[19] Dedicated short range communications (DSRC). http://grouper.ieee.org/groups/scc32/dsrc/index.html.

[20] G. Setiwan, S. Iskander, S. S. Kanhere, Q. J. Chen, and K. C. Lan. Feasibility study of using mobile gateways for providing internet connectivity in public transportation vehiclesand. In *Proceedings of the ACM International Wireless Communications and Mobile Computing Conference (IWCMC)*, Vancouver, Canada, 2006.